A UNITED STATES DEPARTMENT OF COMMERCE PUBLICATION

NBS SPECIAL PUBLICATION 300

VOLUME 10

Precision Measurement and Calibration

Image Optics

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards

Precision Measurement and Calibration

Volumes available or to be published in the NBS Special Publication 300 series

SP300, Vol. 1, Statistical Concepts and Procedures, H. H. Ku, Editor, Feb. 1969, 436 pages (SD Catalog No. C 13.10:300/v.1)* Price $5.50

SP300, Vol. 2, Temperature, J. F. Swindells, Editor, Aug. 1968, 520 pages (SD Catalog No. C 13.10:300/v.2)* Price $4.75

SP300, Vol. 3, Electricity–Low Frequency, F. L. Hermach and R. F. Dziuba, Editors, Dec. 1968, 498 pages (SD Catalog No. C 13.10:300/v.3)* Price $4.50

SP300, Vol. 4, Electricity–Radio Frequency, A. J. Estin, Editor, June 1970, 450 pages (SD Catalog No. C 13.10:300/v.4)* Price $5.50

SP300, Vol. 5, Frequency and Time, B. E. Blair and A. H. Morgan, Editors, June 1972, 565 pages (C 13.10:300/v.5)* Price $6.00

SP300, Vol. 6, Heat, D. C. Ginnings, Editor, Feb. 1970, 400 pages (SD Catalog No. C 13.10:300/v.6)* Price $5.00

SP300, Vol. 7, Radiometry and Photometry, H. K. Hammond, III, and H. L. Mason, Editors, Nov. 1971, 690 pages (SD Catalog No. C 13.10:300/v.7)* Price $7.00

SP300, Vol. 8, Mechanics, R. L. Bloss and Mary J. Orloski, Editors, Jan. 1972, 590 pages (SD Catalog No. C 13.10:300/v.8)* Price $6.25

SP300, Vol. 9, Colorimetry, I. Nimeroff, Editor, June 1972, 460 pages (SD Catalog No. C 13.10:300/v.9)* Price $5.50

SP300, Vol. 10, Image Optics, Calvin S. McCamy, Editor, Nov. 1973, 953 pages (SD Catalog No. C 13.10:300/v.10)* Price $11.10

---

*Available from the Superintendent of Documents, Government Printing Office, Washington, D.C. 20402. Order by the Supt. of Documents Catalog number as given for each volume.

UNITED STATES DEPARTMENT OF COMMERCE • Peter G. Peterson, *Secretary*

NATIONAL BUREAU OF STANDARDS • Lawrence M. Kushner, *Acting Director*

Precision Measurement and Calibration

Selected NBS Papers on

IMAGE OPTICS

Calvin S. McCamy, Editor

Institute for Basic Standards
National Bureau of Standards
Washington, D.C. 20234

A compilation of previously published papers by the staff of the National Bureau of Standards, including selected abstracts by NBS and non-NBS authors. Issued in several volumes, see page IV.

NBS Special Publication 300—Volume 10

Nat. Bur. Stand. (U.S.), Spec. Publ. 300, V. 10, 953 pages (Sept. 1972)

CODEN: XNBSAV

Issued (Nov. 1973)

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402
(Order by SD Catalog No. C13.10:300/V.10).

Price $11.10

Stock Number 0303-01179

Abstract

This volume is one of an extended series which brings together the previously published papers, monographs, abstracts, and bibliographies by NBS authors dealing with precision measurement of specific physical quantities and the calibration of the related metrology equipment. The contents have been selected as being useful to the standards laboratories of the United States in tracing to NBS standards the accuracies of measurement needed for research work, factory production, or field evaluation.

Volume 10 deals with image optics, including photography. It contains 62 reprints assembled in 4 sections: (1) Refractometry and Optical Homogeneity, (2) Interferometry in Image Optics, (3) Optical Design and Image Evaluation, (4) Photographic Science. Each section is introduced by an interpretive foreword.

Key words: Camera calibration; image evaluation; image optics; image stability; interferometry; lens testing; light filters; light sources; photography.

Library of Congress Catalog Number: 68–60042

Foreword

In the 1950's the tremendous increase in industrial activity, particularly in the missile and satellite fields, led to an unprecedented demand for precision measurement, which, in turn, brought about the establishment of hundreds of new standards laboratories. To aid these laboratories in transmitting the accuracies of the national standards to the shops of industry, NBS in 1959 gathered together and reprinted a number of technical papers by members of its staff describing methods of precision measurement and the design and calibration of standards and instruments. These reprints, representing papers written over a period of several decades, were published as NBS Handbook 77, Precision Measurement and Calibration in three volumes: Electricity and Electronics; Heat and Mechanics; Optics, Metrology, and Radiation.

Some of the papers in Handbook 77 are still useful, but new theoretical knowledge, improved materials and increasingly complex experimental techniques have so advanced the art and science of measurement that a new compilation has become necessary. The present volume is a part of a new reprint collection, designated NBS Special Publication 300, which has been planned to fill this need. Besides previously published papers by the NBS staff, the collection includes selected abstracts by both NBS and non-NBS authors. It is hoped that SP 300 will serve both as a textbook and as a reference source for the many scientists and engineers who are concerned with accurate measurement.

RICHARD W. ROBERTS, *Director*

Preface

The general plan for this compilation has been reviewed by the Information Committee of the National Conference of Standards Laboratories. The plan calls for Special Publication 300 to be published in 10 volumes as presented on the inside of the front cover.

The division of subject matter has been chosen to assure knowledgeable selection of context rather than to attain uniform size. It is believed, however, that the larger volumes of approximately 500 pages, will still be small enough for convenient handling in the laboratory. The compilation for this volume consists primarily of original papers by NBS authors which have been reprinted by photoreproduction.

Each volume has a subject index and author index, and within each volume, contents are grouped by subtopics to facilitate browsing. Many entries follow the recent Bureau practice of assigning several key words or phrases to each document; these may be collated with titles in the index. Pagination is continuous within this volume, the page numbers in the original publications also being retained and combined with the volume page numbers. The index notation 10–1 refers to volume 10, page 1 of this volume. A convenient list of SI (Système International) physical units and a conversion table are to be found inside the back cover.

The publications listed herein for which a price is indicated are available from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (foreign postage, one-fourth additional). Many documents in the various NBS non-periodical series are also available from the National Technical Information Service, Springfield, Va. 22151. Reprints from the NBS Journal of Research or the non-NBS journals may sometimes be obtained directly from the author.

Editor's Note

Image optics is a broad field, encompassing many specialties. Though only a few specialties are represented by these reprints, there is considerable diversity. It is hardly to be expected that all of the papers will be of interest to workers in any one specialty. However, requests for reprints and other indications of need for these publications suggest that each will serve a need in some sector of the field. Two of the selections appear in print for the first time in many years, although they have been very much in demand. The first of these describes the procedures for making the Davis-Gibson filters for producing daylight in the laboratory. These filters have been referenced in a number of national and international standards in the fields of photometry, colorimetry, and photography, but the publication has been unavailable for many years. The second is the proceedings of a symposium on optical image evaluation held at the Bureau of Standards in 1951. This symposium was held at just the right time to bring forth the evidence of a new wave of development in image evaluation. The papers presented are still often referenced but the volume has been out of print so long that it is not easily found. As evidence of the "rare book" status of these two items, we may note that even at the National Bureau of Standards it has been difficult to find two expendable copies to be cut apart for reproduction in this collection.

The fact that most photographs are made by amateurs, for their enjoyment, tends to obscure the connection between photography and precision measurement. Photography has been an indispensable tool in spectroscopy, astronomy, and the applications of photogrammetry to the mapping of the earth and the moon. This last application has been the objective of much of the work in image optics at NBS, as is evident in a number of the papers reprinted here. When photography is employed in precision measurement, as in astronomy, there is often a need to maintain the images over long periods of time. It was for this reason that the publications on the preservation of photographs were included, although the original interest in this topic was primarily centered in the microfilming of archival records.

C. S. McCamy, *Editor*

Contents

Contents

Contents

Contents

Contents

Contents

Contents

4. Photographic Science

Contents

1. Refractometry and Optical Homogeneity

Foreword

Tilton's paper on "Standard Conditions for Precise Prism Refractometry" is a basic reference for anyone undertaking precision measurement of refractive index. Most of the remaining papers on refractive index are included because of the refinements of methodology presented, rather than the experimental results. Rosberry's papers deal with the measurement and practical significance of inhomogeneity of refractive index in optical materials.

U. S. Department of Commerce National Bureau of Standards

RESEARCH PAPER RP776

Part of Journal of Research of the National Bureau of Standards, Volume 14, April 1935

STANDARD CONDITIONS FOR PRECISE PRISM REFRACTOMETRY

By Leroy W. Tilton

ABSTRACT

Air as a standard reference medium for precise refractive-index measurements is discussed with respect to the precision necessary in the control and measurement of its temperature, pressure, humidity, and carbon-dioxide content. Particular attention is given to the selection of provisional values for α, the optical temperature coefficient for air; and the preparation of accurate correction tables for the reduction of refractive-index observations to standard conditions of reference is described and exemplified.

The approximate range of temperature effects on absolute refractive index is indicated; also the range of pressure effects on liquids. A new (empirical) relation between index and density is derived from Pockels' optical data on elastically deformed glass and a basis for quantitative treatment of permissible stress-birefringence is given. For glasses, the character of annealing and the permissible degree of striation are considered; and for several media the requisite constancy in wave length of light source is determined.

Tolerance equations are given for all requirements that are quantitatively discussed and it is concluded that both precision and accuracy to within ± 2 or 3×10^{-6} can be attained in determining refractive index by the classical method of minimum deviation.

CONTENTS

518-743 O - 73 - 2

I. INTRODUCTION

In the practice of precise refractometry a desirable degree of approximation to the ideal physical and chemical conditions of measurement is seldom realized. Although the necessity for control of some of the physical conditions is more or less well known, only a comparatively small amount of the published index data is accompanied by adequate references to the conditions under which the measurements were actually made or to the conditions which are considered as standard.[1]

The tolerances for variations in air temperature, pressure, and humidity have not been explicitly and precisely evaluated, and the effects of temperature and pressure on the absolute indices of various media are not adequately known. Moreover, the question of the effect of heterogeneities in the sample has not hitherto been discussed in connection with refractive-index measurements. On the other hand, it appears that no attempt has been made to justify or to disprove the commonly accepted opinion that even a moderate amount of birefringence in a medium may make precise index measurements useless.

Certain refractive-index measurements of high precision, that are being made in the refractometric laboratory at the National Bureau of Standards, have made it necessary to determine the precision with which all of the working conditions must be controlled or, in some instances, quantitatively observed. It has seemed advisable to differentiate clearly between the effects of such conditions on the reference medium and on the sample to be measured and, wherever possible, to specify on a quantitative basis the requirements and tolerances relating to these working conditions; also, to set forth in detail a system of accurate corrections to standard reference conditions. However, when considering such matters as heterogeneity of the sample and the permissible degree of its birefringence, the treatment is necessarily limited to simple approximations.

For convenience of reference, the definitions of various symbols that are used and explained in this paper are summarized here as follows:

$\alpha \equiv$ temperature coefficient of expansion of air at constant pressure; also (negative) temperature coefficient of refractivity;

$\beta \equiv$ pressure coefficient of volume change at constant temperature (often called compressibility), that is

$\beta = -\frac{1}{V_0}\left(\frac{dV}{dP}\right)$; also, for air, the pressure coefficient of refractivity;

$C \equiv$ total (volume) proportion of CO_2 in atmosphere;

$C', C'', C''' \equiv$ constants of proportionality in various equations expressing index-density relationship;

$c \equiv$ excess of CO_2 above the normal proportion, that is, $c = C - 0.00035$;

$d \equiv$ density;

[1] Brief historical references have been made in a former paper, BS J.Research 2, 912-915 (1929) RP64.

$E \equiv$ Young's modulus of elasticity;
$\epsilon = 170 \times 10^{-6}/(\mu_o - 1)$;
$\gamma = 11 \times 10^{-6}/[760(\mu_o - 1)]$;
$K \equiv$ modulus of volume elasticity or "bulk modulus", that is, $K = 1/\beta$;
$\lambda \equiv$ wave length of light;
$m \equiv$ mass;
$\mu \equiv$ index of refraction of air, the general subscripts t, p, v, and c referring to the temperature, pressure, humidity, and carbon-dioxide content, respectively, while the subscript 0 indicates standard conditions of 0° C and 760 mm pressure and D specifies a wave length of 5893 A;
$n \equiv$ relative (to air) index of refraction of a medium;
$\bar{n} \equiv$ absolute (or vacuum) index of refraction of a medium;
$P \equiv$ pressure of air;
$p \equiv$ excess in mm of Hg over 1 atmosphere of air pressure, that is, $p = (P - 760)$;
(p'/v') and $(q'/v') \equiv$ the Pockels experimental constants for certain glasses, v' being the velocity of light in the unstressed glass;
$s \equiv$ proportion by volume of striae in glass;
$\sigma \equiv$ Poisson's ratio;
T and $t \equiv$ temperatures on absolute and centigrade scales, respectively;
$T_{\Delta t_a} \equiv$ tolerance in air temperature control and measurement that corresponds to an error of $\pm 1 \times 10^{-6}$ in (relative) refractive index;
$T_{\Delta p_a} \equiv$ tolerance in air-pressure control and measurement that corresponds to an error of $\pm 1 \times 10^{-6}$ in (relative) refractive index;
$T_{\Delta v} \equiv$ tolerance in absolute humidity that corresponds to an error of $\pm 1 \times 10^{-6}$ in (relative) refractive index;
$T_{\Delta c} \equiv$ tolerance in abnormal proportion (by volume) of CO_2 in the atmosphere that corresponds to an error of $\pm 1 \times 10^{-6}$ in (relative) refractive index;
$T_{\Delta P} \equiv$ tolerance in hydrostatic-pressure variations corresponding to $\pm 1 \times 10^{-6}$ in refractive index of liquid and solid media;
$T_{\Delta n_b} \equiv$ tolerance in absolute stress-birefringence corresponding to an error of $\pm 1 \times 10^{-6}$ in refractive index (for the transverse ray if $n < 1.8+$);
$T_{\Delta \lambda} \equiv$ tolerance in wave length of source corresponding to an error of $\pm 1 \times 10^{-6}$ in refractive index;
$V \equiv$ volume; and,
$v \equiv$ pressure of water vapor in the atmosphere in mm of Hg.

II. AIR AS A STANDARD MEDIUM OF REFERENCE

The great majority of all measurements involving light velocities are made while using the earth's atmosphere as the practical standard medium of immersion or reference. Since the velocity of light in air is a function of chemical composition and also of the temperature and pressure of air, it is necessary to specify CO_2, moisture content, and other factors with appropriate precision in order that the measurements may have definite value.

The composition of CO_2-free dry air has been discussed in a former paper.[2] Although this composition appears constant to a high degree when air is chemically analyzed, it has long been estimated that variations in the normal density of air are as large as 5 or 10 parts in 10,000 and for this reason air is no longer considered a suitable reference medium for accurate specific gravities of gases. Statistical analysis of existing data indicates, likewise, that the refractive index of air at standard conditions is not constant but varies over a range perhaps as large as $\pm 12\times10^{-7}$ from a general average value of 1.0002925 for $\lambda=5893$ A. Consequently all relative indices of refraction may be subject to an inherent uncertainty which can be estimated as

$$\Delta n=\pm 12n\times10^{-7} \tag{1}$$

unless accompanied by reliable measurements on air itself.

In reducing relative indices to absolute values when direct measurements on air are lacking, it is probable that this uncertainty is statistically diminished by using for μ, the index of air, a value properly computed [3] from dispersion data and from an equation that gives the sodium-lines index of air for 0° C and 760 mm pressure as

$$(\mu_D-1)\times10^7=\mu_{S=0}-kS \tag{2}$$

where S is the relative annual sunspot number (for the year in which the relative indices were measured) as given in the Wolf-Wolfer series of sunspot observations. For use in this equation the parameters $\mu_{S=0}$ and k should be determined from the best available data on the refractive index of air for the period in question. Provisional values obtained by the writer for the whole interval 1857 to 1934 are, respectively, 2932.2 and 0.148; whereas for the limited interval 1912 to 1934, during which a number of very important observations have been made, the corresponding values of the parameters are 2927.5 and 0.079. With either of these sets of values for the parameters, however, the probable errors of estimates of index interpolated by use of equation 2 are large, namely ±5 and $\pm3\times10^{-7}$, respectively, and it is readily seen that for certain years within the interval 1912 to 1934 two estimates can actually differ by 5×10^{-7}. Extrapolations will probably be even less reliable and consequently it appears (see equation 1) that uncertainties of $\pm$ 1 or 2 units of the sixth decimal place of index can not be entirely excluded when refractivities of optically dense media are measured in CO_2-free dry air. From evidence at present available it does not seem likely, however, that such differences can occur over periods of time measured in days or weeks but rather that such possibilities should be remembered when comparing absolute determinations separated by or extending over a period of years.

[2] L. W. Tilton, BS J. Research 13, 111 (1934) RP695.
[3] See equation 22 of this paper.

Obviously, it is seldom convenient to use as a reference medium pure air having a standard temperature and pressure, and free from CO_2. Usually it is advisable to allow these conditions to vary continuously and to keep adequate records during index determinations. In investigating the relationships involved, a unit error of 1×10^{-6} will be considered as was done in former papers [4] discussing the geometrical aspects of prism refractometry.

1. TOLERANCES IN MEASUREMENT AND CONTROL OF REFERENCE MEDIUM

The most obviously important effects of varying air conditions on measured relative indices are those arising from the changes in air density through changes in pressure, P, and in air temperature, T (absolute). The relationship between the index of air and its density, d, is usually assumed to be that constancy expressed by Gladstone and Dale's law as $(\mu-1)=Cd$. Then, in accordance with the ideal gas law, $PV=RT$, it is customary to write for any given wave length

$$(\mu-1)_{TP}=(\mu_0-1)\left(\frac{273}{760}\right)\left(\frac{P}{T}\right) \tag{3}$$

where μ_0 is a value for standard conditions of 0° C and 1 atmosphere pressure, and P is measured in mm of Hg; or, defining p as the excess in mm of Hg over 1 atmosphere of pressure, $p=(P-760)$ and

$$(\mu-1)_{tp}=(\mu_0-1)\frac{(1+\beta p)}{(1+\alpha t)} \tag{4}$$

where t is the temperature on the centigrade scale, β is the pressure coefficient of volume change of a gas at constant temperature (namely 1/760 for a 1 mm change at a pressure of one atmosphere) and α is the temperature coefficient of expansion of air at constant pressure. The value of α would be $0.00366=1/273$ for a change of 1° C, at a temperature of 0° C, if air behaved like a perfect gas, but 0.00367 if the results of experiments on the expansion of air are considered. The only published references to computed reductions or corrections for indices measured in air at other than standard conditions have been based on these or on similar assumptions.

Since at constant temperature the proportionality between $(\mu-1)$ and d has been found to hold without serious discrepancy over a pressure range of many atmospheres, it is safe to assume that the pressure coefficient of refractivity is identical with β for such small variations in atmospheric pressure as those which occur during index measurements and also, for variations which exist between various places on the earth's surface. Concerning α, however, the similar assumption does not seem equally reliable. According to generally accepted theories of dispersion, a change in the temperature of a medium affects the resonance frequency as well as the mass per unit volume. As a result there is a change or "shift" of the effective wave length of the absorption band, and for each given wave length a different secondary temperature influence on index is superposed on that which is directly ascribed to the change in density. Thus it is

[4] L. W. Tilton, BS J. Research 6, 59 (1931) RP262; 11, 25 (1933) RP575.

not necessarily true that values of $\partial \mu/\partial t$ can be accurately obtained by differentiating equation 4 with respect to t and solving for

$$\frac{\partial \mu}{\partial t} = -\frac{\alpha(\mu-1)_{tp}}{1+\alpha t} \tag{5}$$

with a constant value of α determined from the expansion of air or even from index data for a single wave length. Instead of such a constant "optical temperature coefficient",[5] the measurements on air by W. F. Meggers and C. G. Peters [6] yield an α varying with wave length from 0.00367_4 at 7500 A to 0.00373_8 at 3500 A.

Table 1 summarizes the results of direct determinations of this constant for air. The Gylden value was deduced from astronomical observations. Mascart's value should not be considered because of a systematic inconsistency in results obtainable from his data.[7] Scheel's

TABLE 1.—*Optical temperature coefficient for air*

(Visible range of frequencies)

Observer	Reference	Temperature interval on Centigrade scale	$\alpha \times 10^4$
Gylden	Mém.acad.St.Petersburg 10, no. 1 (1866); 12, no. 4 (1868).	Seasonal range	369
L. Lorentz	K.Danske Videnskab.Selskabs Skrifter [5] 10, 485 (1875).	1 to 17°, 100°	367
Von Lang	Pogg.Ann. 153, 463 (1874). (This α is given as computed by Benoit).	2 to 95°	303
Mascart	Compt.Rend. 78, 617-679 (1874); Ann.Sci.École Normale Supérieure [2] 6, 9-78 (1877).	5 to 38°	382
Benoit	Travaux et Mémoires Bur. Int. Poids et Mesures 6, 102 (1888).	1 to 80°	367
Walker	Trans.Roy.Soc. (London) A201, 454 (1903)	10 to 100°	360
Scheel	Verhandl.deut.physik.Ges. 9, 24-36 (1907)	−192 to 0°	373
Barus	Carnegie Inst.Wash.Pub.no. 249-I, page 137 (1916)	22 to 100°	361
Meggers and Peters	See page 735 Bul BS 14, (1918) S327	0 to 30°	368.8
Pérard	J.phys.radium [6] 6, 223 (1925)	0 to 100°	371.4
Cheney	Phys.Rev. [2] 29, 292-298 (1927)	25 to 300°	367

α is very high, but his temperature interval is much lower than that used by the others. He selected $\alpha=0.00367_4$ for reducing his own index of air from 16 to 0° C. Cheney's temperature interval is much higher than that used by the others. A close analysis of his published numerical data shows that a low value of $\alpha=0.00364$ fits as well or better than the conventional value which he considered verified. Von Lang's α is too low for serious consideration and no great accuracy was claimed for the Barus data.[8]

Considering the amount of data taken, particularly the number of temperatures used, the results by Benoit and by Pérard are probably entitled to special consideration. They used the same method (Fizeau dilatometer) and to some extent the same apparatus. If their values are averaged one has $\alpha=0.00369$ which agrees closely with the average value of $\alpha=0.00368_8$ that is obtained from the Meggers

[5] From inspection of equation 5 one finds that α can be regarded as the (negative) temperature coefficient of index of refraction of air per unit of "excess refractivity", as determined for $t=0°$ C and $p=0$.

[6] Bul. BS 14, 697 (1918) S327.

[7] If from the experiments which he conducted when investigating temperature effects one discards those for temperatures below the middle group, which averages 14.4° C for t, the writer finds that $\alpha=0.00395$ is indicated. On the other hand, when the observations at temperatures above those of the middle group are discarded an $\alpha=0.00356$ is obtained.

[8] It is of interest to note that Barus' qualitative conclusion concerning the decrease of α for high temperatures is consistent with the trend which has been mentioned for some of the data of table 1, particularly the high value which Scheel found in the very low temperature interval that he used, and the low value which the writer would recommend from Cheney's data.

and Peters data for those particular wave lengths [9] which Pérard used. Thus it seems certain that the optical temperature coefficient is appreciably larger than the coefficient of expansion which has been so generally used. Since the Meggers and Peters values of α are based on data for such a large number of wave lengths, and since their average for the visible region is in excellent agreement with the average α from the Benoit and the Pérard data, the Meggers and Peters values are therefore recommended for all intercomparisons between precise relative-index data taken at markedly different temperatures.

The wave-length variation in α is of importance chiefly in the ultraviolet region; but even for the visible region, especially for the shorter wave lengths, it is in some cases preferable to use the variation in α when reducing six- and seven-decimal-place indices to certain reference conditions. In particular, when relative indices are used as a basis for computing temperature variations in the absolute index of a medium, it is desirable to consider the most accurate data available on air.

(a) TEMPERATURE OF AIR

Whether α is regarded as constant or as varying with wave length, a formal expression for the error in relative index caused by air-temperature variations is readily written after considering the definitive relation

$$n \equiv \frac{\bar{n}}{\mu} \tag{6}$$

which exists between the relative index, n, of a medium and its absolute index, $\bar{n}$. Since the latter is necessarily independent of air temperatures and μ^2 is approximately unity

$$\frac{dn}{dt} = -\bar{n}\frac{d\mu}{dt} \tag{7}$$

and consequently, using equation 5, it is found that changes in air temperature result in index variations of like sign according to the equation

$$\Delta n = \frac{\bar{n}\alpha(\mu-1)_{tp}}{(1+\alpha t)}\Delta t_a \tag{8}$$

from which precise index corrections may be computed for the whole wave-length interval 2218 to 9000 A by using the Meggers and Peters values [10] of α, which they approximately expressed by the equation

$$\alpha = 0.00367 + \frac{3\times10^{-6}}{\lambda^3} \tag{9}$$

where λ is to be used in microns.[11]

For the purpose of establishing practical tolerance limits in temperature control and measurements, $\bar{n}$ is essentially equivalent to n, and α for the visible region is approximately the ratio 1/270. Moreover, it happens that the refractivity of air at temperatures between 0 and 50° C varies only from about 24 to 30×10^{-5} over the range of

[9] Pérard used only five wave lengths, all within the visible region, and the data do not seem sufficiently concordant either to confirm conclusively or to disprove the variation of α with wave length.

[10] See page 735 Bul. BS 14 (1918) S327.

[11] For precise refractometry the angstrom is generally favored by the writer as the appropriate unit of wave length, but the micron is often more convenient in the writing and using of dispersion equations.

the visible spectrum, and ordinary atmospheric-pressure changes of $p=\pm 30$ mm do not materially increase this variation. Therefore, with an accuracy of approximately $\pm$ 10 percent, the tolerance corresponding to $\pm 1\times10^{-6}$ in refractive-index measurement is

$$T_{\Delta t_a}=\pm n^{-1}(1+\alpha t) \text{ degrees C} \tag{10}$$

or, for work on media having indices of 1.9 and 1.3, respectively, the air-temperature changes of $\pm 0.53°(1+\alpha t)$ and $\pm 0.77°(1+\alpha t)$ are equivalent to errors of $\pm 1\times10^{-6}$ in the measured relative indices. Thus it is evident that a neglect of seasonal changes in air temperatures may easily affect the fifth decimal place of relative-index work on any medium, even though its temperature coefficient of absolute index is negligibly small.

(b) PRESSURE OF AIR

For the purpose of arriving at similar correction and tolerance equations for air pressures, one differentiates equations 4 and 6 with respect to p and finds that an increase in air pressure results in a decreased value of the measured relative index of other media according to the equation

$$\Delta n=-\frac{\bar{n}\beta(\mu-1)_{tp}}{(1+\beta p)}\Delta p \tag{11}$$

from which precise corrections to measured indices can be satisfactorily computed.

Again approximating, as was done with equation 8, equation 11 reduces to the practical working tolerance in air pressures

$$T_{\Delta p_a}=\pm 2.8n^{-1}(1+\beta p)\text{mm} \tag{12}$$

which is written for an error of $\pm 1\times10^{-6}$ in refractive index and, obviously, for the usual small values of p only one term is needed. Thus for media having indices of 1.9 and 1.3, respectively, the air-pressure changes of 1.5 and 2.2 mm are equivalent to variations of unity in the sixth decimal place of measured relative index of refraction. For many localities, therefore, it is apparent that the fluctuations in barometric pressure within a single working day should not be neglected in precise measurements of relative index, and it is certain that the air-pressure variations that occur over somewhat longer periods affect the fifth decimal of such indices.

(c) HUMIDITY OF AIR

The proportion of water vapor in air is such a variable factor, depending on locality and weather conditions, that it is a difficult matter to decide upon any definite proportion as a satisfactory normal moisture content of the atmosphere. Consequently it is probably better to treat moisture as an impurity and refer all refractive-index measurements to dry air as a standard reference medium.

L. Lorenz [12] gives, for sodium light, a value of $\mu_D=1.0002500$ as the index of refraction of water vapor at a concentration of 1 mol in 22.4 liters. Comparing this with his index of dry air for equivalent conditions, he gives a correction of $\Delta\mu_v=-0.000041\ v/760$ to be applied to his results on dry air when it is desired to obtain indices for moist

[12] Ann. Physik [3] 11, 91 (1880).

air containing water vapor of v mm pressure. The Cuthbertsons [13] and others [14] have confirmed the magnitude of this correction and, for the region 4779 to 6708 A, have shown that the dispersion of water vapor is closely comparable with that of air. It seems, in fact, from experimental evidence, that the coefficient of this Lorenz correction, namely, 41×10^{-6}, should not vary more than $\pm 1 \times 10^{-6}$ throughout the visible spectrum.

Since the pressure of water vapor in the air seldom exceeds a small fraction of an atmosphere, it seems permissible to assume the equivalence of optical temperature coefficients for water vapor [15] and for air, and to use the Lorenz humidity correction for all ordinary temperatures by writing for the whole visible range of frequencies

$$(\mu_t - 1)_v = (\mu_t - 1)_{v=0}(1 - \gamma v) \tag{13}$$

where $\gamma = 185 \times 10^{-6}$ is used to replace $41 \times 10^{-6}/[760\ (\mu_0 - 1)]$. This equation is written for a pressure of 760 mm, but is applicable at all ordinary atmospheric pressures. Differentiating equation 13 with respect to v gives

$$\frac{d\mu}{dv} = -\gamma(\mu_t - 1) \tag{14}$$

and, referring to a similar differentiation of equation 6, it is evident that an increase in the moisture content of the air results in an increased relative index of a medium which is being measured therein, the error equation being

$$\Delta n = +\bar{n}\gamma(\mu_t - 1)\Delta v \tag{15}$$

for use in precise computations of refractive index of the medium which is measured in air.

For an error of unity in the sixth decimal of index, and with sufficient accuracy for many purposes, equation 15 yields

$$T_{\Delta v} = \pm\ 20n^{-1}\text{mm} \tag{16}$$

which shows that the unit tolerances in absolute humidity vary from approximately $\pm$ 10 to $\pm$ 15 mm of vapor pressure. These figures apply to measurements of relative index on media having indices of 1.9 and 1.3, respectively. Vapor pressures of 10 and 15 mm correspond at 50° C to 10 and 15 percent in relative humidity and at 30° C to 30 and 45 percent, respectively. Thus it is evident that humidity corrections are not negligible in sixth-decimal-place refractometry, especially when working in air at temperatures somewhat higher than those of ordinary room conditions or whenever there is danger of humidities approaching the saturation point (because, for example, of leaks in water-jacketed housings used for temperature control).

[13] C. and M. Cuthbertson, Trans. Roy. Soc. (London) A213, 16 (1913).
[14] J. Wüst and H. Reindel, Z. physik. Chem. B24, 176 (1934). See, also, P. Hölemann and H. Goldschmidt, page 204 of same volume.
[15] From the Hölemann and Goldschmidt data (see footnote 14) on water vapor, considered together with published densities of steam, one may deduce a tentative value $\alpha = 0.0035$ to compare with 0.0037 for air.

(d) CARBON-DIOXIDE CONTENT OF AIR

After water vapor the next atmospheric constituent of recognized importance, so far as its variability is concerned, is carbon dioxide. Although the range in its fluctuations is known to be much smaller than in the case of water vapor, there is a prevalent opinion that when using air as a reference medium for measurements of highest precision one should eliminate either the CO_2 itself or the effects of its variations. While there are on record some presumably reliable analyses of air showing a carbon-dioxide content as high as 0.3 percent by volume, such samples were taken under exceptional circumstances, such as in very crowded theaters of an obsolete and most unfavorable type, or in rooms with no ventilation, and they have no interest in the present connection. Boothby and Sandiford [16] reported that outdoor air sampled under favorable conditions varies in its CO_2 content by only a few thousandths of one percent from a normal value of 0.036 percent. Similarly, F. G. Benedict [17] found that outdoor air does not vary in this respect more than a few thousandths from an average value of 0.031 percent, but Hann [18] states that the percentage of CO_2 varies from 0.033 to 0.043 in the air of cities. Consequently, in view of these small variations, it seems permissible and somewhat preferable to adopt some value, say 0.00035, as the normal volume proportion of CO_2 in air rather than to favor the more stringent requirement that a standard reference medium must be entirely free from carbon dioxide.

Precise index measurements are, of course, seldom made in outdoor air, but in any well ventilated modern laboratory it is unlikely that the carbon-dioxide content ever exceeds two or three times the normal value. In air from the Boston and New York subways, sampled just after the rush hours, Benedict found only twice the normal CO_2 content. Moreover he found that an increase in CO_2 is accompanied by an equal volume decrease in O_2, and consequently it is the difference in the refractivities of carbon dioxide and of oxygen which must be considered in making corrections to indices measured in air having an abnormal CO_2 content.

The averages of a number of published [19] values of the refractive indices of CO_2 and O_2 are 1.000450 and 1.000271, respectively, for sodium light, a temperature of 0° C and a pressure of 760 mm. Hence a correction of

$$\Delta\mu_C = +0.000179(C - 0.00035) \tag{17}$$

should be applied to the measured index of normal air (containing the proportion 0.00035 of CO_2) at 0° C and 760 mm pressure when it is desired to obtain the index of air containing the total proportion C of carbon dioxide.

A comparison of the dispersions of oxygen and of carbon dioxide shows that the correction given by equation 17 varies only 1 percent or less for the visible-spectrum region. Proceeding as with the water-vapor correction (part c of this section) and writing $c = C - 0.00035$ to represent excess CO_2, one finds for all ordinary temperatures

$$(\mu_t - 1)_C = (\mu_t - 1)_{c=0}(1 + \epsilon c) \tag{18}$$

[16] Am. J. Physiol. 55, 295 (1921).
[17] The Composition of the Atmosphere, Carnegie Inst., Wash., Pub. no. 166., pages 110–114 (1912).
[18] J. von Hann's Lehrbuch der Meteorologie, page 5 (1915).
[19] See Landolt-Börnstein Tabellen.

where $\epsilon=0.612$ replaces $179\times10^{-6}/(\mu_0-1)$. Differentiating equation 18 with respect to c gives

$$\frac{d\mu}{dc}=+\epsilon(\mu_i-1) \tag{19}$$

whence, again using equation 6, it is apparent that the presence in laboratory air of CO_2 in abnormally high proportions results according to the equation

$$\Delta n=-\bar{n}\epsilon(\mu_i-1)\Delta c \tag{20}$$

in decreased values for such relative indices as are measured in the laboratory.

For unit error in the sixth decimal place of refractive index the corresponding tolerance in volume proportion of abnormal CO_2 content of air is

$$T_{\Delta c}=\pm\ 0.006n^{-1} \tag{21}$$

with sufficient accuracy for all purposes. Evidently no consideration whatever need be given to CO_2 content of air in sixth-decimal-place refractometry. Even for the extreme case when a medium of index 1.9 is being measured, the excess proportion of CO_2 must reach the high value of 0.003, or nine times the proportion normally present in fresh air, in order to produce an error of -1×10^{-6} in relative index. For seventh-decimal-place index work, however, it may be ascertained from these results that the normal CO_2 content of fresh air causes a lowering of approximately 1×10^{-7} in the measured relative indices of very dense media as compared with the use of a standard reference air containing no carbon dioxide. Consequently, it may be safely inferred that, in general, indices of laboratory air, or relative indices measured in such air, are not likely to have any precision whatever in the eighth decimal place unless simultaneous CO_2 determinations are made.

2. CORRECTIONS TO STANDARD REFERENCE CONDITIONS

If the working conditions affecting the reference medium are not maintained constant at the desired standard values or controlled within the tolerance limits as specified above, then it becomes necessary to observe the existing conditions and make the proper corrections to the results actually obtained. When it is necessary to correct sixth-decimal-place indices to standard conditions, or to convert them to absolute indices, the use of slide-rule corrections is often lacking in adequate precision and correction tables for these purposes are advisable. Their preparation in detail is a profitable preliminary whenever numerous data are to be reduced for a limited number of spectral lines. For highest precision such tables are especially desirable because certain refinements in precision of correction need be considered explicitly only during their original preparation.

The writer recommends the use of logarithms in the computation of indices, and the evaluation of all corrections in terms of log n. One of the first requirements is a double-entry table of barometer corrections with barometer temperature and barometer reading as arguments. With the temperature reductions [20] to be listed in this table the barometer calibration corrections (for a given instrument)

[20] Temperature reductions for use with a Fortin-type barometer and brass scale are given in tables 46 and 47 of the 4th revised edition, Smithsonian Meteorological Tables.

are combined and then all entries are modified to reduce pressures from local [21] to standard gravity, $g=980.665$ dynes.

Another requirement is a series of double-entry tables (or their equivalent), with air pressure and temperature as arguments, giving the positive corrections which, when applied to log n, will yield log $\bar{n}$, of which the antilog is absolute index. The various members of such a series of tables relate to the different spectral lines that are to be frequently used for index measurement and these member tables will differ principally because of the dispersion of air. (There is also a slight additional difference because of the nonconstancy of α, the optical temperature coefficient of index. See discussion in part 1 of section II.) As an example of a member table of such a series, table 2 is abstracted (by decimation in each argument) from a usefully detailed tabulation for the hydrogen line, $\lambda=4861$ A.

TABLE 2.—*Data for converting from relative to absolute index* ($\lambda=4861$ A) *by the equation* $\log \bar{n}=\log n+\log \mu$

P=air pressure (in mm of mercury)	Values of $10^6\times\log \mu_P$ for dry air at various temperatures (with normal CO_2 content)				
	0° C ($\Delta P=0.06$)	10° C ($\Delta P=0.13$)	20° C ($\Delta P=0.24$)	30° C ($\Delta P=0.44$)	40° C ($\Delta P=0.77$)
730	12295	11857	11449	11069	10713
740	12464	12020	11606	11220	10859
750	12632	12182	11763	11372	11006
760	12800	12344	11920	11524	11153
770	12969	12507	12077	11675	11300
780	13137	12669	12234	11827	11446

NOTES

1. For each 10 percent of relative humidity subtract tabulated values of ΔP from air pressure before entering this table.
2. Tentatively it may be advisable to reduce log μ by 5 or 10 parts in 10,000 for each increase of 20 in relative annual sunspot number above average $S=46$.
3. If it is desired to use these data for wave lengths other than 4861 A, modify air temperature according to table 4 before entering this table.

In preparing such correction tables in this laboratory, the temperature and dispersion data as given by Meggers and Peters [22] have been combined with the average value of refractive index, $\mu=1.0002926$ (for $\lambda=5893$ A, $t=0°$ C and $P=760$ mm), which is obtained by considering all published data [23] and adding a very slight correction for normal CO_2 content. Since the ratio of log μ to $(\mu-1)$ is very nearly constant over the whole range in μ that is involved in temperature and pressure reductions and corrections, the actual computation of air indices is unnecessary except for 0° C and 760 mm pressure. The values of log μ for all ordinary temperatures and pressures may be directly computed (see equation 4 and citations of footnote 22) from the equation

$$\log \mu_{tp}=\frac{1+\beta p}{1+\alpha t}\log\left[1+\left(2884.3+\frac{13.412}{\lambda^2}+\frac{0.3777}{\lambda^4}\right)\times 10^{-7}\right] \quad (22)$$

where α varies with wave length as approximately expressed by equation 9, and λ should be written in microns.

[21] Gravity reductions are discussed on pages XXXV and XXXVI of 4th revised edition, Smithsonian Meteorological Tables.
[22] See pages 722 and 735, Bul. BS 14 (1918) S327.
[23] See page 115, BS J. Research 13 (1934) RP695.

If, in table 2, the values for a standard pressure are subtracted from those for other pressures in the same column, a set of air-pressure corrections to the logarithms of observed indices is obtained. Table 3 gives such corrections for limited ranges in pressure and in temperature variations. Such a table may, of course, be computed directly from the basic values of the logarithms for the indices at standard pressure, p_0=760 mm. For this purpose the equation

$$\Delta \log \mu_{tp} = \frac{\log \mu_{tp0}}{760} \Delta p \qquad (23)$$

is obtained from equation 22.

TABLE 3.—*Data for referring relative indices to dry air at standard pressure by the equation* $\log n_{760} = \log n_p + \Delta \log n$.

Use tabulated values with negative sign when $P-760$ is negative]

($P-760$ mm) ≡ excess atmospheric pressure	Values of $10^8 \times \Delta \log n$ for various air temperatures		
	10° C ($\Delta P=0.13$)	20° C ($\Delta P=0.24$)	30° C ($\Delta P=0.44$)
20	325	314	303
19	309	298	288
18	292	282	273
17	276	267	258
16	260	251	243
15	244	235	227
14	227	220	212
13	211	204	197
12	195	188	182
11	179	173	167
10	162	157	152
9	146	141	136
8	130	125	121
7	114	110	106
6	97	94	91
5	81	78	76
4	65	63	61
3	49	47	45
2	32	31	30
1	16	16	15
0	0	0	0

NOTE.—For each 10 percent of relative humidity subtract tabulated values of ΔP from air pressure before entering this table.

Although the direct use of table 2 is strictly limited to the correction of indices for the F line of hydrogen, table 3 is applicable over a very wide range in the spectrum. In fact, for pressure corrections as large as 30 mm the use of table 3 may be extended over the whole visible range of frequencies without errors exceeding $\pm 5 \times 10^{-8}$ in the logarithmic corrections so determined.

As they are written, tables 2 and 3 both relate to dry air and must be modified because of the presence of water vapor. From equations 11 and 15 it is evident that any given increase in moisture can be represented as an equivalent decrease in pressure, and thus it is possible to find a simple system of modifying the actual corrected pressures to fictitious or pseudo pressures which, when used as actual pressures, permit tables 2 and 3 to serve for moist as well as dry air. Hence the ΔP equivalents which appear in the column headings of tables 2 and 3.

If for any reason it is objectionable to prepare a complete series of log μ tables, similar to the single example given as table 2, one may use a relationship between $\Delta\mu/\Delta\lambda$ and $\Delta\mu/\Delta t$ for extending to other wave lengths the use of a table written primarily for one particular

TABLE 4.—*Auxiliary data for extending use of table 2 to various wave lengths*

$\lambda \equiv$ wave length (in angstroms)	Values of Δt for modification of air temperatures		
	0° C	20° C	40° C
4000	−3.18	−3.30	−3.43
4100	−2.70	−2.80	−2.91
4200	−2.25	−2.33	−2.42
4300	−1.83	−1.90	−1.98
4400	−1.45	−1.50	−1.56
4500	−1.10	−1.14	−1.18
4600	−0.76	−0.79	−0.82
4700	−0.45	−0.47	−0.49
4800	−0.16	−0.17	−0.18
4900	+0.11	+0.11	+0.12
5000	+0.36	+0.38	+0.40
5200	+0.81	+0.84	+0.87
5400	+1.21	+1.26	+1.31
5600	+1.57	+1.63	+1.69
5800	+1.88	+1.96	+2.04
6000	+2.17	+2.26	+2.35
6200	+2.42	+2.52	+2.63
6400	+2.65	+2.76	+2.88
6600	+2.86	+2.98	+3.10
6800	+3.05	+3.18	+3.31
7000	+3.21	+3.36	+3.50

wave length. For this procedure it is found advantageous to regard temperature as the fictitious argument, and table 4 gives the particular corrections of this type that should be algebraically applied to actual air temperatures before entering table 2 for wave lengths other than 4861 A.

III. REQUISITE CONSTANCY AND UNIFORMITY FOR REFRACTIVE MEDIA

Uniformity of reference conditions is a necessary, but by no means a sufficient condition for high precision in refractometry. Even the absolute indices of refraction of optical media are ratios which may, for divers reasons, vary from time to time or with respect to coordinates within the media. Temperature and pressure are obviously important factors and, in the case of optical glass, heat treatment may introduce not only a variation of index with direction of travel of the light, but also a marked physico-chemical change [24] in index that often varies from point to point in the glass although it is not a function of orientation. Moreover, the writer has found that striae in glass, when they exist in well-stratified layers, can affect measurements of index and of dispersion in a systematic manner that may prove misleading in some investigations. Therefore, for consistent results in index determinations, it is necessary carefully to consider the conditions which cause real or apparent changes in the optical density of media that are subject to measurement.

[24] A. Q. Tool and C. G. Eichlin, J.Opt.Soc.Am. 4, 359 (1920); BS J.Research 6, 525 (1931). A. Q. Tool, L. W. Tilton, and E. E. Hill, J.Opt.Soc.Am. and Rev.Sci.Inst. 12, 490-491 (1926). L. W. Tilton, J. Wash. Acad.Sci. 20, 12-13 (1930).

1. EFFECTS OF TEMPERATURE ON REFRACTIVE MEDIA

Comparatively little recent progress has been made in the theory of temperature effects on absolute index. It is considered, however, that this effect consists of at least two well-defined components which may, and often do, have opposite signs. The more frequently mentioned component is the direct result of thermally produced changes in density and its sign is always as indicated by the equation

$$\frac{\partial \bar{n}}{\partial t}=f(\bar{n})\frac{dd}{dt} \tag{24}$$

where $f(\bar{n})$ is a positive function, which may be variously approximated (for example as $\bar{n}-1$, $\bar{n}^2-1$, $\frac{\bar{n}^2-1}{\bar{n}^2+2}$, etc.). Consequently, for media which expand when heated, this change in index is negative as temperature increases, and this effect of temperature is not a function of wave length but relates primarily to conditions for infinite wave length. As mentioned for air in part 1 of section II, the second component is ascribed to changes in the resonance frequencies of the media. The increase in volume which usually takes place with rising temperature is often accompanied by a capacity for absorbing radiation of longer wave length. Thus, under such circumstances, ultraviolet absorption bands are "shifted" toward the longer wave lengths of the visible region while the long wave lengths of infrared bands become still longer. Obviously, then, these absorption components of temperature coefficients of index are functions of wave length.

The velocity of propagation of radiation of a wave length longer than that corresponding to the effective absorbing frequency is decreased, while that of shorter wave length is increased. Consequently, for media whose principal effective absorption bands lie in the ultraviolet, this total absorption-band component of refractive index usually increases as temperature increases, while the reverse is true for media with predominant infrared absorption. In most optical glass the index change (in and near the visible spectral region) that is ascribable to the ultraviolet absorption bands predominates over the combined effects of the directly produced density change and the change caused by shifting of the infrared bands. For some crown glasses, however, the opposing tendencies are about equal in absolute value, and the index sensitivity to temperature is quite negligible. For a few crowns of very low index and somewhat unusual chemical composition the density and infrared effects predominate and negative temperature coefficients of absolute index are found, especially for the longer wave lengths of the visible spectrum.

(a) THERMAL COEFFICIENTS OF ABSOLUTE REFRACTIVE INDEX

For a few optical glasses temperature coefficients of absolute index are readily available through the work of C. Pulfrich [25] and of J. O. Reed.[26] Also, such coefficients can be readily computed from coeffi-

[25] Ann.Physik 45, 609-665 (1892).
[26] Ann.Physik 65, 707-744 (1898).

cients of relative index that have been determined by C. S. Hastings,[27] G. Müller,[28] H. Fizeau,[29] and J. W. Gifford;[30] or they may be estimated from the graphs published by C. G. Peters.[31] From a critical examination of these data, and also from unpublished results obtained in this laboratory, it is inferred that chemical composition is a predominantly important factor in determining this thermo-optical relationship for glass. Accordingly, it is difficult to make concise general statements, with a useful degree of precision, about the quantitative effects of temperature variations on the refractivity of glasses.

At usual room temperatures the values of $10^6 \times \Delta\bar{n}/\Delta t$, for the sodium-lines indices of flint glasses vary from approximately 0 for some light flints containing barium to +14 for those with highest lead content. For low-index glasses it is often considered that the temperature coefficients of absolute index are small and that they lie within a narrow range as compared with flint glasses. Available data confirm this view for the optical crowns in frequent use, the range in their coefficients being −1 to +3. Measurements in this laboratory show, however, that certain fluor crowns have values of −3 while for Pyrex and fused-quartz glasses of similar refractivity the coefficients are +5 and +9, respectively.

For crystals the range in refractive-index sensitivity to temperature is greater than for glasses. It is sufficient to mention sylvite, which, at room temperature, has a coefficient of approximately -34×10^{-6}, and calcite with its value of $+11 \times 10^{-6}$ for the extraordinary ray.

Numerous published data concerning the temperature coefficients of refractive index of liquids show that such values are all negative (direct density effect predominating over that of the absorption shift in the ultraviolet) and very much larger in numerical value than those for most solids.[32] It is useful to remember, however, that the percentage variations among such temperature coefficients for different liquids are, in general, relatively small compared with similar variations for solids. For a large number of transparent and semitransparent liquids (including numerous oils) the value of $\Delta\bar{n}/\Delta t$ is approximately expressed as -4×10^{-4}, although a number of organic liquids have values ranging from −4 to -6×10^{-4}. Water with a value of -1×10^{-4} is a most important exception and, on the other hand, there are some optically dense liquids which have unusually high (negative) temperature coefficients of index, such, for example, as −7 and -8×10^{-4} for methylene iodide and carbon disulphide, respectively.

(b) APPROXIMATE RANGE IN TEMPERATURE TOLERANCES FOR SOLIDS AND LIQUIDS

When considering temperature tolerances for general refractometry it must be remembered that one may have no prior knowledge of the media and their temperature coefficients of index. From the values given above it is evident that even for glasses known to be the usual crowns and light flints, the temperature of the sample during refrac-

[27] Am.J.Sci. [3] 15, 269-275 (1878).
[28] Publicationen Astrophysikalischen Observatoriums Potsdam 4, 140-216 (1885).
[29] Ann.chim.phys. [3] 66, 429-482 (1862).
[30] Proc.Roy.Soc. (London) A91, 319-321 (1915); A100, 621-626 (1921-22).
[31] BS Sci.Pap. 20, 635 (1926) S521.
[32] The absolute value of the ratio between temperature coefficients of liquids on the one hand and of glasses and common optical crystals on the other may be very roughly expressed as 100. Consequently, for liquids, coefficients of relative and of absolute refractive index are practically identical, whereas for glasses and crystals they frequently are quite different and may be of opposite sign.

tive-index determinations must be known within the limits of approximately ± 0.3° C to correspond to a limiting error of ± 1 × 10^{-6} in index. For Pyrex glasses and for fused quartz the corresponding temperature tolerances are ± 0.2 and ± 0.1° C, respectively. Possibly, precision and accuracy in temperature measurement within approximately these limits can with care be obtained when a glass prism is mounted on a spectrometer table and exposed to the air of a room in which the temperature is constant or changes but slowly. When, however, the denser flint glasses are considered, it is found that the tolerances are reduced to ± 0.07° C (and even to ± 0.03° C for determinations with the shorter wave lengths of the visible spectrum).

Consequently, without considerable prior knowledge concerning a particular transparent solid, it is not advisable to attempt sixth-decimal-place refractometry unless temperature is controlled and measured to within ± 0.1° C or better. It is, therefore, highly desirable to provide a constant-temperature prism housing, even for work in a constant-temperature room, and to determine all indices with respect to a stirred air bath surrounding the prism.

Tolerances in temperature control and measurement during refractive-index determinations on liquids to within one unit of the sixth decimal place are so strict that there seems to be no published account of a serious attempt to reach such precision. The requirements are ± 0.01° C for water at room temperature, and approximately ± 0.002° C for liquids of average thermo-optical sensitivity.

2. EFFECTS OF PRESSURE ON REFRACTIVE MEDIA

L. Zehnder [33] found that water under pressure increases in absolute index by 15×10^{-6} per atmosphere at 20° C, and he favored the Gladstone and Dale relation, the constancy of $(n-1)/d$, when the density of liquids is varied by pressure. W. C. Röntgen and L. Zehnder [34] give data on similar increases in index for several other liquids. Although they do not choose between the Gladstone and Dale relation and that of Lorenz-Lorentz, their results are in general somewhat more favorable to the former.

F. Pockels [35] investigated the effects of unidirectional pressure on seven kinds of glass, using Neumann's equations for expressing the influence of elastic deformation on the propagation of light, and he found the Newton relation, $(n^2-1)/d$, somewhat more nearly constant than that of Gladstone and Dale or of Lorenz-Lorentz. L. H. Adams and E. D. Williamson [36] also have published results on the relation between birefringence and stress in various types of glass and have confirmed Pockels' results where comparisons can be made.

There is, however, a general lack of direct experimental evidence relating to the effect of atmospheric-pressure variations on the absolute refractivity of transparent media, and particularly so for optical glasses. Both Hovestadt [37] and Adams and Williamson have mentioned the computation of such effects from Pockels' work and the equations of Neumann but, unfortunately, the necessary constants

[33] Ann. Physik [3] 34, 91-121 (1888).
[34] Ann. Physik [3] 44, 49-50 (1891).
[35] Ann. Physik [4] 7, 745-771 (1902); also, for important corrections, see 9, 221 (1902,) and 11, 652 (1903).
[36] J. Wash. Acad. Sci. 9, 609-623 (1919).
[37] H. Hovestadt, Jena Glass, appendix H, page 409 (Macmillan and Co., London, 1902).

518-743 O - 73 - 3

characteristic of different media are known only for the seven glasses used by Pockels, and two [38] of these are unusual alumino-borates.

(a) VARIATIONS IN ATMOSPHERIC PRESSURE

Some indications concerning the effects of atmospheric-pressure changes on liquid and solid refractive media may, however, be obtained from a consideration of the index-density relationships together with the published data on β, the compressibility at constant temperature, and on its reciprocal, K, the "bulk modulus" or modulus of volume elasticity. If, in accord with Pockels' results for glass, the Newton form $(n^2-1)=C'd$ is written to express the index-density relation for a medium having a modulus $K=-V\,dP/dV$, and a density $d=m/V$, then

$$d\bar{n}=\frac{(n^2-1)}{2nK}dP \tag{25}$$

is obtained as the differential relation between refractive index and pressure. Or, if the index-density relation is $(n-1)=C''d$, one finds according to this Gladstone and Dale premise

$$d\bar{n}=\frac{(n-1)}{K}dP \tag{26}$$

as an alternative index-pressure relationship. Similarly, from the Lorenz-Lorentz expression

$$d\bar{n}=\frac{(n^2-1)(n^2+2)}{6nK}dP \tag{27}$$

would be written.

In view of the analysis which Pockels made from his experiments on glass under pressure, equation 25, rather than 26 or 27, is preferable for solids, while from the work of Zehnder and of Röntgen and Zehnder with pressure applied to liquids it may be inferred that equation 26 is more suitable for liquid media. The writer has found, however, that Pockels' data agree with an index-density relation

$$\left(\frac{n^2-1}{n^2+2}\right)^6=C'''d^3 \tag{28}$$

much more closely than with the Newton relation, and consequently the differential equation

$$d\bar{n}=\frac{(n^2-1)(n^2+2)}{10nK}dP \tag{29}$$

is preferable to equation 25. Since $K=E/[3(1-2\sigma)]$, where E is Young's modulus of elasticity and σ is Poisson's ratio, equation 29 may be written

$$d\bar{n}=\frac{3(1-2\sigma)(n^2-1)(n^2+2)}{10nE}dP \tag{30}$$

[38] The Pockels glass listed as 0.428, $n=1.5123$, was not, as originally supposed, an alumino-borate like his two other crown glasses but, according to subsequently published corrections (see footnote 35, p. 409), was a borosilicate glass. The revised data are free from peculiarities which marked the original exhibit. Nevertheless the errors continue to be propagated in the literature or, as in the International Critical Tables, 2, page 106, this important glass is omitted from consideration.

or, the change in index may be computed approximately from

$$\Delta\bar{n}=\frac{0.15(n^2-1)(n^2+2)}{nE}\Delta P \tag{31}$$

if the constant value $\sigma=0.25$ is used instead of experimentally determined values [39] which with few exceptions lie within the range from 0.21 to 0.27.

Equation 26 has been used in computing $\Delta\bar{n}$ for water, ethyl ether, ethyl alcohol, benzene, and carbon disulphide, using for $1/K$ averages of several published values for the compressibility of these liquids. The results agree within the limits of ± 10 percent with published experimental values. Equation 29, which is based on the new index-density relation, has been used in computing $\Delta\bar{n}$ for Pockels' glasses when subjected to changes in hydrostatic pressure, and the results lie within approximately ± 5 percent of those computed from the actual experimental constants in the manner suggested by Pockels. These 5 percent differences compare with 15 and 20 percent differences that are obtained when a similar computation is made by equation 25, which is based on the Newton index-density relation as favored by Pockels. If, however, the approximate form (equation 31) is used instead of equation 29 the differences are as large as 15 percent.

From equation 26 the tolerance in atmospheric- or other hydrostatic-pressure changes for work on liquid media is

$$T_{\Delta P}=\pm\frac{0.000001}{(n-1)\beta} \tag{32}$$

for a limiting error of $\pm 1\times10^{-6}$ in refractive index. For ethyl ether or carbon disulphide the limits ± 12 mm are found, and for ethyl alcohol or benzene they are ± 17 mm. Consequently, pressure corrections of this nature should not be entirely neglected. The less compressible liquids water and glycerin have the correspondingly more liberal tolerances of ± 47 and ± 72 mm, respectively, and they require corrections only for extremes of pressure variations or for reductions to standard conditions after measurements are made where the prevailing barometer level is appreciable different from 760 mm.

From equation 31 one may write

$$T_{\Delta P}=\pm\frac{0.000007nE}{(n^2-1)(n^2+2)} \tag{33}$$

as the tolerance in ΔP for sixth-decimal-place refractometry of glass. The values of E seem to range within ± 10 percent of the average 0.72×10^6 atmospheres for borosilicate glasses not containing lead oxide or phosphoric acid,[40] and except for a few extremes all other glasses seem to have values of E which lie within ± 10 percent of 0.58×10^6 atmospheres. Hence, considering only the absolute index of glass, it is evident that the tolerance in measurement or control of air pressures is seldom if ever [41] less than about 1/2 atmosphere, and therefore of no moment.

[39] R. Straubel, Ann. Physik. [3] 68, 409 (1899); or see page 193 of reference cited in footnote 37, page 409.
[40] See page 159 of reference cited in footnote 37, page 409.
[41] Glasses containing approximately 74 percent of PbO may perhaps be exceptions. See part b of this section.

In connection with the subject of pressure effects on absolute index it should be remembered, however, that corrections of this nature should be considered if one attempts a precise comparison between relative indices and those observed under actual vacuum conditions. In the case of liquids, it becomes imperative, even for accuracy in the fifth decimal place, to distinguish between values actually measured in a near vacuum (under their own vapor pressure) and those which are reduced from relative measurements by considering only the effect of pressure on the ambient air.

(b) STRESS AND BIREFRINGENCE

In addition to hydrostatic-pressure effects there should also be considered those effects which are due to unequally distributed stresses within media. Such stress distribution is caused, for example, by improper annealing of glass and by the improper mounting or tight clamping of optical components, especially under varying temperature conditions. The birefringence that is produced in this manner is commonly used as a measure of the internal stress and it seems desirable, if possible, to use birefringence as a measure of that change in refractive index which accompanies the stresses.[42]

From Pockels' [43] work it is possible to express the ratio between the change in index, Δn_y or Δn_z, and the absolute birefringence, $\Delta n_y - \Delta n_z$. For the ray which vibrates parallel to the axis of stress

$$\frac{\Delta n_y}{\Delta n_y - \Delta n_z} = \frac{2\sigma}{1+\sigma} - \frac{(1-2\sigma)\frac{q'}{v'}}{(1+\sigma)\left(\frac{p'}{v'} - \frac{q'}{v'}\right)} \tag{34}$$

and for the transverse ray

$$\frac{\Delta n_z}{\Delta n_y - \Delta n_z} = \frac{2\sigma}{1+\sigma} - 1 - \frac{(1-2\sigma)\frac{q'}{v'}}{(1+\sigma)\left(\frac{p'}{v'} - \frac{q'}{v'}\right)} \tag{35}$$

where p'/v' and q'/v' are experimentally determined constants for a given medium and σ is Poisson's ratio. For glass in the index range 1.5 to 1.75 the Pockels' data lead to higher absolute values for the transverse ray ratio, while for his glass of $n=1.96$, equation 34 gives the larger ratio. Consequently, it may be said that the tolerance in absolute stress-birefringence $\Delta n_b = (\Delta n_y - \Delta n_z)$, in millimicrons per centimeter, for sixth-decimal-place measurements of refractive index is at least as large as

$$T_{\Delta n_b} = \pm \frac{10}{f\left(\sigma, \frac{p'}{v'}, \frac{q'}{v'}\right)} \text{ millimicrons per cm} \tag{36}$$

[42] For clearness it may be necessary to mention that any piece of unannealed glass in which such stresses exist is a balanced system involving both tensions and compressions so that the index changes which are produced are of opposite sign and offset each other. (See L. H. Adams, J. Franklin Inst. 216, 39-71 (1933).) Nevertheless, portions of the glass may show an average birefringence of a given sign and therefore experience a correspondingly changed index. Such limited portions of the medium should be individually considered because such a procedure corresponds closely to conditions under which optical-glass components are often measured and used, namely, at partial rather than at full apertures.

[43] Ann. Physik [4] 7, 745-771 (1902); also for important corrections, see 9, 221 (1902), and 11, 652 (1903).

where the denominator is computed according to equation 35 for glasses of index $<1.8+$ and by equation 34 for glasses of index >1.9, approximately.

Equation 36 with each of these denominators written in turn, has been used in computing the corresponding values of this tolerance for the seven Pockels glasses, and figure 1 shows the chief results. It does not seem clear where the curves of this figure should cross [44] and accordingly they are not drawn in full. In any event, it is evident that stress-birefringence differs greatly in the degree of sensitivity with which it can serve to measure the variations in index which may

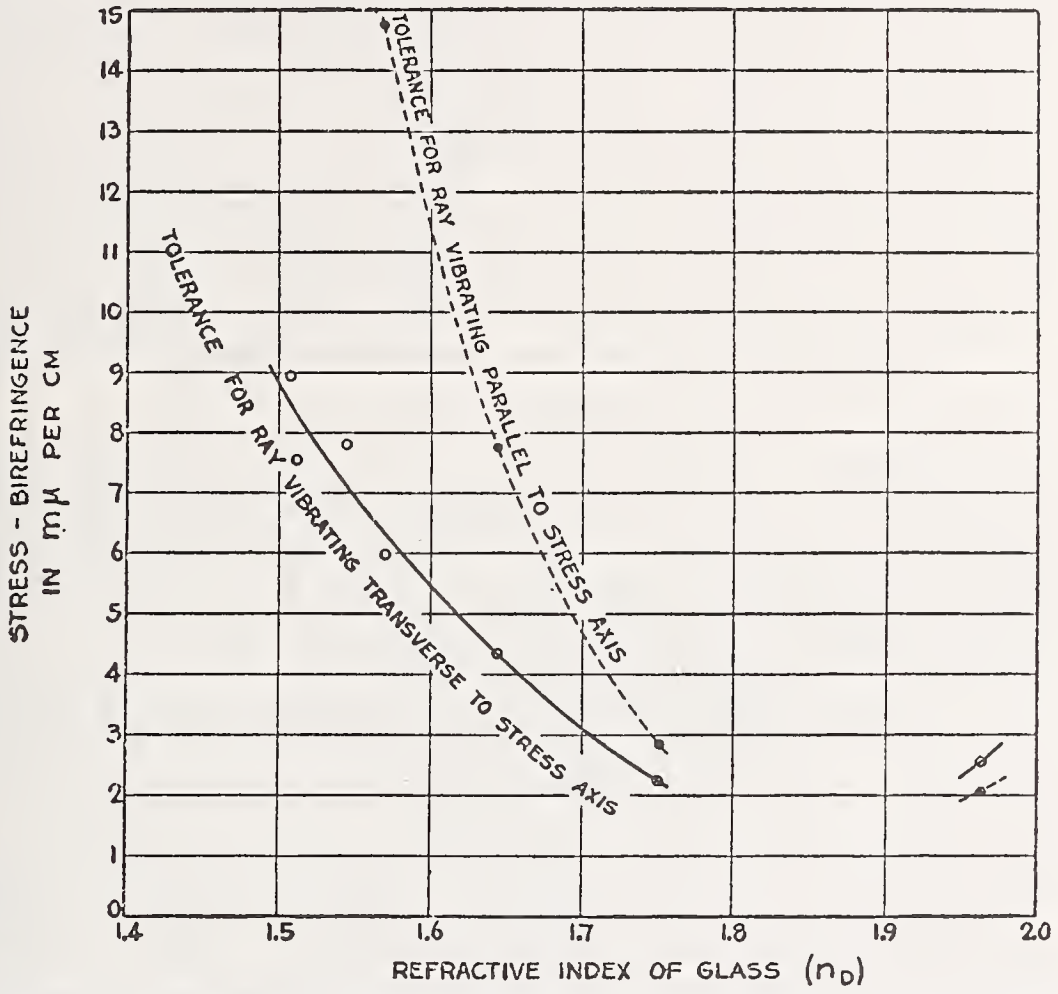


FIGURE 1.—*Tolerances in stress-birefringence for optical glass on which refractive index is measured to* $\pm 1\times 10^{-6}$.

Where the orientation is indefinite, the safer tolerance is obviously that indicated by the lower curve.

exist because of the stresses in various kinds of glass. For flints it is obvious that the birefringence present should be much lower than for crowns, and consequently the annealing requirements and schedules for flints should not be too readily neglected because of the comparative ease with which their stress-birefringence seems to be reduced. Nevertheless, remembering the Rayleigh limit of one-quarter wave

[44] Pockels concludes that p' must equal q' for some glass containing about 74 percent of PbO (index 1.8+) and consequently from equations 34, 35, and 36 it seems that for p' and q' not zero the crossing of the curves of figure 1 should occur at zero tolerance (at least for all values of σ other than 0.5, which is the limiting value for the case of an incompressible medium). On the other hand, for $\sigma=0.5$ when $p'=q'$, or whenever $p'=q'=0$, the ordinate of the point of crossing appears indeterminate as judged from these equations.

length of permissible phase difference and the resulting requirement [45] that optical glass be uniform in index within ±0.000007 per cm, it seems that tolerance limits of from 5 to 10 mμ birefringence per cm are of a proper order of magnitude for most of the requirements of optical glasses.

3. HEAT TREATMENT, CHEMICAL HETEROGENEITY, AND STRIATION OF GLASS

When a piece of glass has been annealed in such manner that the stress-birefringence is within the tolerances shown in figure 1 it is very likely that it is also homogeneous in physico-chemical sense. The work of Tool [46] and others has shown, however, that there is a possibility of residual inhomogeneity that can be caused by furnace-temperature gradients during the annealing and which may not be accompanied by appreciable birefringence in the finished product. The sensitivity of optical glass to variations in annealing temperature is such that decreases in refractive index of from 2 to 6×10^{-5} are found in various glasses at ordinary temperatures corresponding to increases of 1° C in the annealing temperature.[47] These annealing-temperature coefficients of index of glass are in general approximately ten times as large as the ordinary temperature coefficients of refractive index, which were considered in part 1 of this section.

In measuring refractivities to five or six decimals, and when making intercomparisons of such results, it is necessary, therefore, before drawing inferences and conclusions, to consider the possibility of small temperature gradients in the annealing furnaces and also to remember the great difficulties of securing sufficiently exact duplications of annealing temperatures and cooling rates during runs in different furnaces or even in successive runs of a given furnace.

Before the optical effects of nonuniform heat treatment were known, it was considered difficult or perhaps impossible to secure chemical homogeneity to the extent necessary for refractive-index uniformity to the sixth or even to the fifth decimal place, even when concerned with distances of only a few centimeters in good optical glass. A careful study and analysis of Fritz Eckert's [48] investigation shows, however, that after "fine annealings" the maximum spread chargeable to chemical heterogeneity within each of three melts of a barium flint glass was $\pm1\times10^{-5}$ or less in refractive index. Also, a report on careful annealings of six lens blanks at this Bureau [49] shows that the chemical heterogeneity, if any, was confined entirely to a few units of the sixth decimal place of refractive index.

Such gross and intensely localized heterogeneities as striae do not occur to any large extent in the best optical glass, but it is sometimes desirable to extend minimum-deviation measurements to less perfect glasses. Consequently, it would be of interest to know the extent to which striation can be present without seriously affecting index measurements. Unfortunately little definite information is available.

Striae commonly exist in two well-known forms—as "cords" or "veins" and as sheet striae or "ream." Their optical properties are,

[45] See page 720 of BS Sci.Pap. 22, (1928) S572.
[46] A. Q. Tool and C. G. Eichlin. J.Opt.Soc.Am. 4, 359 (1920); BS J.Research 6, 525 (1931) RP292.
[47] A. Q. Tool, L. W. Tilton, and E. E. Hill, J.Optical Soc.Am. and Rev.Sci.Inst. 12, 490-491 (1926). L. W. Tilton, J.Wash.Acad.Sci. 20, 12-13 (1930). See also Annual Report of the Director of the Bureau of Standards, U. S. Department of Commerce, pages 11, 23, 21, and 25, of years 1925, 1927, 1928, and 1929, respectively.
[48] Z.tech.Physik 7, 282-287 (1926).
[49] L. W. Tilton, A. N. Finn, and A. Q. Tool, BS Sci.Pap. 22, 719 (1928) S572.

of course, different from those of the adjacent media, and apparently there is a rather abrupt transition [50] at the interfaces. According to prevailing ideas [51] it may perhaps be assumed that in general the principal and only important effect is a scattering of light, which results in loss of contrast and consequent poor definition of the image of the collimator slit. This is probably true if the striae are present mainly as cords and if they do have abrupt boundaries.

On the other hand, when the striae consist of well-stratified layers or extensive sheets of "ream" the conditions are quite different. Unless the planes of stratification are approximately parallel to the light paths, all of the rays traverse striae and suffer a consequent acceleration or retardation, but it is usually found that an image of the slit is fairly well outlined even when one uses prisms in which such striae are very numerous. If the striae are so thin, or the difference in refractive index so small that beams emerging from adjacent kinds of glass differ in phase by less than one-quarter wave length, then the measured deviation produced by such a composite prism corresponds to that for a similar prism having a fictitious index, n_3, intermediate in value between n_1 and n_2, indices of the "normal" glass and the striae, respectively. In fact, if s is the proportion by volume of the striae present, one may write

$$n_3 = n_1 + s(n_2 - n_1) \tag{37}$$

as an estimated condition which seems probable for the case of a quasi regular composite prism of this type.

Since striae may differ in index from the normal glass by 3×10^{-4} (an average of some values which were observed by Smith, Bennett, and Merritt) it seems that the presence of certain types of striae in amounts exceeding 0.3 percent of the whole volume of glass should be considered a potential source of errors that may appreciably affect the sixth decimal place of refractive-index determinations.

IV. REQUISITE PRECISION IN SOURCE OF RADIATION

It is desirable to know to what extent doublets and complex lines may be used in sixth-decimal-place refractometry, and to decide whether or not it is necessary to specify the conditions under which a source of radiation is operated. For such purposes it is sufficient quantitatively to consider the dispersion in relative index for a few highly dispersive substances.

If the simple Cauchy dispersion formula

$$n - 1 = a + \frac{b}{\lambda^2} \tag{38}$$

is written to approximate (with two constants) the result of experiments over a limited spectral interval it is evident that

$$\frac{dn}{d\lambda} = -\frac{2b}{\lambda^3} \tag{39}$$

[50] A. A. Michelson, BS Sci.Pap. 15, 41 (1919) S333.
[51] T. T. Smith, A. H. Bennett, and G. E. Merritt, BS Sci.Pap. 16, 75 (1920) S373. A. Arnulf, Rev. optique 6, 6-20 (1927).

can be written to show how refractive index varies with wave length within the specified interval. Consequently, one may consider

$$T_{\Delta\lambda}=\pm\frac{5\lambda^3}{b\times10^7} \tag{40}$$

as a suitable tolerance in wave length of source for a limiting error of $\pm1\times10^{-6}$ in index.

For use with wave lengths measured in angstroms the constant b is seldom as large as $+25$ or $+30\times10^5$ (values for carbon disulphide and densest flint glass of index $n_D=1.92$, respectively, computed for the shortest wave-length region of the visible spectrum) and thus any uncertainty or variation in the source of less than ±0.01 angstroms is not of consequence for wave lengths longer than approximately 4000 A. For the great majority of substances on which refractive indices are precisely determined, the tolerance is appreciably more liberal. Water and fluorite, for example, have the corresponding minimum or safe tolerances of ±0.10 and ±0.13 A, respectively.

V. PRECISION AND ACCURACY ATTAINABLE IN MINIMUM-DEVIATION REFRACTOMETRY

In this and in previous papers [52] many of the difficulties of precise prism refractometry, particularly by the minimum-deviation method, have been discussed. The number, N, of sources of error is fairly large. At least nineteen of them have been quantitatively discussed in these investigations and the discussions are briefly summarized in table 5. Several of the errors, say $(N-N_1)$ of them, are easily held within limits which are small compared with $\pm1\times10^{-6}$ in index. In no respect is it found impossible, and in few if any instances does it seem unreasonably difficult to control or correct within that tolerance limit. With the exception of error arising from incorrect prism orientation, practically all contributions to the final error in index are equally likely to have positive or negative signs in any extensive and well-arranged observational program. Consequently the algebraic sum is never large.

If, as an approximation to the worst case, N_1 of the contributed errors are large compared with the rest and are all about equal in absolute value, say 1×10^{-6}, then the probable error in the final index is $\sqrt{N_1}\times10^{-6}$. For several years the individual tolerances corresponding to $\pm1\times10^{-6}$ have been used in this refractometric laboratory and it seems that probable errors in refractive indices precisely determined by the minimum-deviation method seldom exceed ±2 or 3×10^{-6}. Hence the existence of some five or ten fairly important sources of accidental error is inferred. Whenever less precision is obtained after the expenditures of reasonable effort in the use of fairly good equipment, the difficulty is probably to be sought not in any one or two particularly important sources of error but in a combination of several matters which have been overlooked or deliberately neglected as inconsequential.

[52] See references listed in table 5.

TABLE 5.—*Summary of sources of refractive-index error that have been quantitatively discussed*

References to BS J. Research			Quantity evaluated or controlled			±Tolerances for various refractivities of sample ($\Delta n = \pm 1 \times 10^{-6}$, $A = 60°$)				
Volume	Page	Formula	Name	Symbol	Unit	$n = 1.3$	$n = 1.5$	$n = 1.7$	$n = 1.9$	Provisory remarks [1]
1929										
2	921	8	Prism angle	ΔA	Second of arc	0.57	0.33	0.22	0.16	
2	921	9	Double deviation	$\Delta 2D$	do	0.56	0.62	0.80	1.5	
1931										
6	72	14	Prism orientation	$P.E.\ i$	Minute of arc	3.9	3.1	2.7	2.4	(Index error always +).
1933										
11	41	23	Prism surface sagitta	s'	Wave length	0.27	0.14	0.08	0.04	1 cm diam. of surface area.
11	42	24	Prism translation	$P.E.\ \epsilon_0$	Millimeter	1.3	0.7	0.4	0.2	$s' = 0.02\lambda$.
11	47	33	Eccentricity of prism-table axis.	ϵ	do	0.3	(Very large)	0.2	0.1	s' = tabulated values.
11	52	45	Collimator focusing	ΔF_C	do	2.9	2.0	1.8	2.7	$f_e = 400$ mm, $\epsilon = 0.2$ mm.
1934										
13	117	------	Air composition	------	Proportion by volume.	(Relatively large tolerances—see reference cited)				CO_2-free dry air.
1935										
14	400	10	Air temperature	Δt_a	Degree C	0.77°	0.67°	0.59°	0.53°	$t = 0°$ C.
14	400	12	Air pressure	Δp_a	Millimeter of Hg.	2.2	1.9	1.6	1.5	
14	401	16	Air humidity	Δv	do	15.	13.	12.	11.	
14	403	21	Abnormal CO_2 in air	Δc	Proportion by volume.	0.005	0.004	0.004	0.003	
14	408	------	Temperature of certain glasses	------	Degree C	----------	Soda lime 0.3° Fused quartz 0.1°	Dense flint 0.2°	Densest flint 0.07°	$\lambda = 5893$ A.
14	408	------	Temperature of certain liquids.	------	do	Water 0.01°	Many oils 0.002°	CS_2 0.001°	----------	$\lambda = 5893$ A.
14	411	32	Hydrostatic pressure on glass.	ΔP	Atmosphere	----------	0.5	0.5	0.5	
14	411	33	Hydrostatic pressure on certain liquids.	ΔP	Millimeter of Hg.	Water 47. Ether 12.	Glycerin 72. Benzene 17.	CS_2 12.	----------	
14	412	36	Stress-birefringence in glass	Δn_b	$m\mu$/cm	----------	9.	3.	2.	(? at $n = 1.8+$).
14	415	37	Striae in glass	s	Proportion by volume.	0.003	0.003	0.003	0.003	$n_2 - n_1 = 0.0003$.
14	416	40	Wave length of light source	$\Delta\lambda$	Angstrom	Water 0.10	Fluorite 0.13	CS_2 0.014	Densest flint 0.010	$\lambda = 4000$ A.

[1] For explanations and for other pertinent limitations and conditions, see discussions given in references cited in columns 1 to 3

It is a much more difficult matter to state precisely what accuracy is attainable. Tests of the accuracy of angle measurement by the use of a prism polished on all three faces are of fundamental importance. Torsion and mechanical strains in the goniometer constitute a particularly dangerous source of systematic error which is seldom entirely constant and cannot always be effectively eliminated. Fortunately, those errors which would be systematic if neglected are automatically replaced by residual accidental errors when proper corrections can be and are applied. In this laboratory no reason has been found for suspecting that residual systematic errors in precise index measurement with a spectrometer are materially in excess of the precision which is attained.

WASHINGTON, January 26, 1935.

U. S. Department of Commerce National Bureau of Standards

RESEARCH PAPER RP1575

Part of Journal of Research of the National Bureau of Standards, Volume 32, February 1944

A PRECISION APPARATUS FOR THE RAPID DETERMINATION OF INDICES OF REFRACTION AND DISPERSION BY IMMERSION

By Conrad A. Faick and Bernard Fonoroff

ABSTRACT

A new immersion method for determining indices of refraction and ν values, employing the double-diaphragm method for securing oblique illumination, is described. The average error in the determination of indices of refraction based upon 144 measurements is 2×10^{-5}; the maximum error is 5×10^{-5}. Complete measurements of the indices of refraction for the sodium D line and the hydrogen F and C lines may be made in approximately 1½ hours, and from these measurements the ν values may be calculated with an average error of 0.1 and a maximum error of 0.8.

CONTENTS

I. INTRODUCTION

In the production of optical glass, careful control over the optical properties, especially the index of refraction (n_D) and the Abbe value (ν), of the product is essential. Most methods for measuring indices of refraction are sufficiently accurate for the control of the n_D value. The ν value, on the other hand, presents a special problem in index measurement. It is defined as $(n_D-1)/(n_F-n_C)$, and in order to determine this value within certain specified limits, the index of refraction must be measured very accurately. Some typical tolerances which have to be met and the errors in the index measurements corresponding to these tolerances are shown in table 1.

TABLE 1.—*Typical tolerances for n_D and ν values*

Glass	Typical values and tolerances		Error corresponding to tolerance in ν
	n_D	ν	
Borosilicate crown	1.517 ±0.001	64.5 ±0.3	$\pm2\times10^{-5}$
Light barium crown	1.572 ±0.001	57.4 ±0.3	$\pm3\times10^{-5}$
Dense flint	1.649 ±0.0015	33.8 ±0.3	$\pm9\times10^{-5}$

It is obvious from table 1 that the required accuracy of index measurements is determined mainly by the ν value tolerances. If the errors in index measurements are as great as those listed in column three, the ν value may be in error by the amount of the permissible tolerances. In order to be certain that the ν values are well within the required limits, the best spectrometer measurements [1] are necessary for final control of crown glasses, and are desirable for flint glasses. However, there are many useful routine checks and measurements that can be made on refractometers.

In normal times at the optical glass plant of the National Bureau of Standards, the problem of control was solved by making two independent sets of measurements. The immersion method of Faick and Finn [2] was used for immediate production control work. The final and more accurate values of optical constants were later measured on a spectrometer and a precision refractometer.

During wartime, manufacturers are called upon to make new glasses and a greater variety of them. This requires frequent changes in batch compositions and makes necessary more frequent and careful determination of index of refraction and dispersion. Under these conditions the above-mentioned immersion control method was deficient both as to speed and accuracy. The more accurate index measurements as they are generally made, on spectrometers or on carefully calibrated Pulfrich refractometers, require ground and pitch-polished samples. The time-consuming operations of grinding and polishing cause delay in the processing of the glass or in the revision of batch compositions, especially in glass plants that are not equipped with grinding and polishing facilities for the preparation of samples. Most immersion methods permit the use of rough, unprepared samples, but either do not yield sufficiently accurate results or lack the simplicity and speed desired.[3] It was, therefore, felt desirable to develop an immersion method for rough samples which would have the required accuracy and permit speed and ease of manipulation.

II. METHOD OF MATCHING

One of the chief requirements of immersion refractometry is a sensitive method for ascertaining the match between the indices of the sample and the immersion liquid. Of the various methods that could be used for this, one embodying the principles of oblique illumination, well known to petrographers, seemed the most promising both for sensitivity and simplicity of operation. A system of oblique illumination, using double diaphragms, has been described by Saylor [4] and shown to be superior to usual methods, at least when used on a microscope. Another double-diaphragm method, employing a telescope and a long focus lens as parts of the optical system, has been used by Cheshire [5] for measuring the index of refraction of glass by immersion. This system is too cumbersome for rapid control work.

[1] H. L. Gurewitz and L. W. Tilton, J. Research NBS 32, 39 (1944) RP1572.

[2] C. A. Faick and A. N. Finn, BS J. Research 6, 993-1002 (1931) RP320.
This method consists essentially in matching the index of refraction of the glass with that of a suitable liquid in which the sample is immersed and then measuring the index of the liquid. The apparatus consists of a telescope and a grid, illuminated by a sodium light, which are placed on opposite sides of a water-cooled tank containing the immersion liquid and sample. The match between the glass and the liquid is obtained by altering the index of the liquid until no distortion of the grid pattern is observed through the telescope. Accuracies within $\pm 3\times10^{-4}$ are obtainable with this method.

[3] For a discussion of immersion refractometry, see Glazebrook's Dictionary of Applied Physics 4, 130 (The Macmillan Co., New York, N. Y., 1923).

[4] C. P. Saylor, J. Research NBS 15, 277-294 (1935) RP829.

[5] R. W. Cheshire, Phil. Mag. [6] 32, 409-420 (1916).

Basically, the method finally chosen consists in inserting at the focus of a condensing lens (fig. 1, A) a diaphragm which shades out almost one-half of the beam of light. The emerging beam is then passed through the immersion liquid and again brought to a focus by means of a second converging lens. At this focus (fig. 1, B) a second

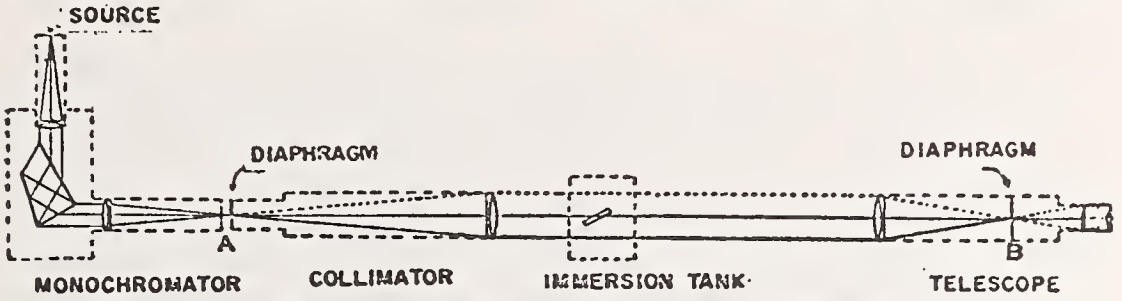


FIGURE 1.—*Schematic drawing of the optical system for matching the indices of the glass and immersion liquid.*

diaphragm is placed so that most of the remaining bright portion of the beam is shaded out. If a telescope is then inserted in the system, the field, as seen through the eyepiece, will appear dimly but uniformly lighted. Now, when a specimen is immersed in a liquid of the same index of refraction as that of the specimen, the light rays passing through the specimen will not be deviated and, since the second

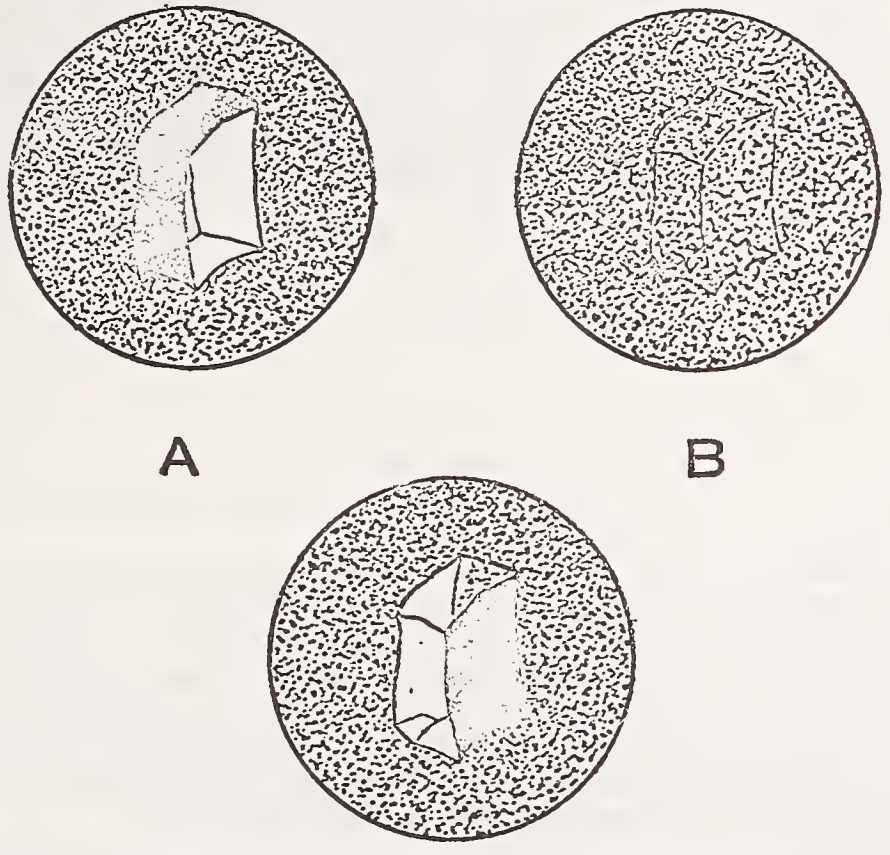


FIGURE 2.—*Field through the eyepiece of the telescope—diaphragm on right side of the instrument.*

A, When the index of the liquid is greater than the index of the sample; B, When the liquid and sample have the same index; C, When the index of the liquid is lower than that of the sample.

diaphragm takes out the bright portion of the beam, the sample will be practically invisible against the dimly lighted field seen through the eyepiece (fig. 2, B). However, if the match between the sample and liquid is altered, the path of light passing through the sample will be deviated so that the sample no longer blends with the field, but becomes light on one side and dark on the other. When the index of the immersion liquid is greater than that of the glass, the side of the sample (as seen through the telescope) that is on the same side as the diaphragms will be light (fig. 2, A); when the index of the liquid is lower, the opposite side of the sample will be light (fig. 2, C). By altering the index of refraction of the immersion liquid, which is a mixture of two liquids, one of a low and the other of a high refractive index, the match between the specimen and immersion medium can be easily established. After a match has been obtained, the index of refraction of the liquid is measured on a precision refractometer by comparison with a standard of approximately the same index.

III. APPARATUS

The apparatus (figs. 1 and 3) consists essentially of light sources, a monochromator with the exit slit removed, a collimator, an immersion tank, a telescope, a tested precision refractometer, and accessory equipment. Referring to figure 1, the light from the source is first passed through a monochromator,[6] where a particular spectrum line is isolated and is brought to a focus at the end of the collimator where the first diaphragm is placed. This diaphragm is an adjustable knife-edge that covers part of a circular opening, replacing the usual collimator slit. The light then passes through the collimator, from which it emerges as a parallel beam (or a very slightly converging one) that passes through the tank and then enters the telescope. Near the focal plane, the second diaphragm, which is also adjustable, is located. The objective is selected so that in combination with the eyepiece it forms a telescope that can be focused on the sample in the immersion tank at a distance which is convenient for the operator. As an added convenience, the telescope is fitted with a right-angle prism and mounted in a vertical position.

The light sources used are a sodium-vapor lamp and a hydrogen-discharge tube. They are mounted on a sliding support to facilitate changing from one to the other (fig. 3).

The immersion tank is equipped with plate-glass windows and a water jacket for temperature control. The water jackets of this tank and of the refractometer are connected in parallel, and water from the same source is circulated through both. The accessory equipment includes a water reservoir and circulating pump, a motorized stirrer for mixing the immersion liquids, thermometers[7] in the immersion tank and refractometer and a vacuum pump and a hydrogen generator for frequent refilling of the discharge tube.

[6] By use of a monochromator, it was possible to avoid difficulties arising from the other lines in the sodium source, and from the continuous spectrum in the hydrogen tube. Filters like those used on the refractometer were found unsatisfactory on the matching instrument.

[7] Mercury-in-glass type graduated to a tenth of a degree, permitting the estimation of hundredths of a degree. The thermometers should be carefully checked against each other, since the difference in temperature is the important factor.

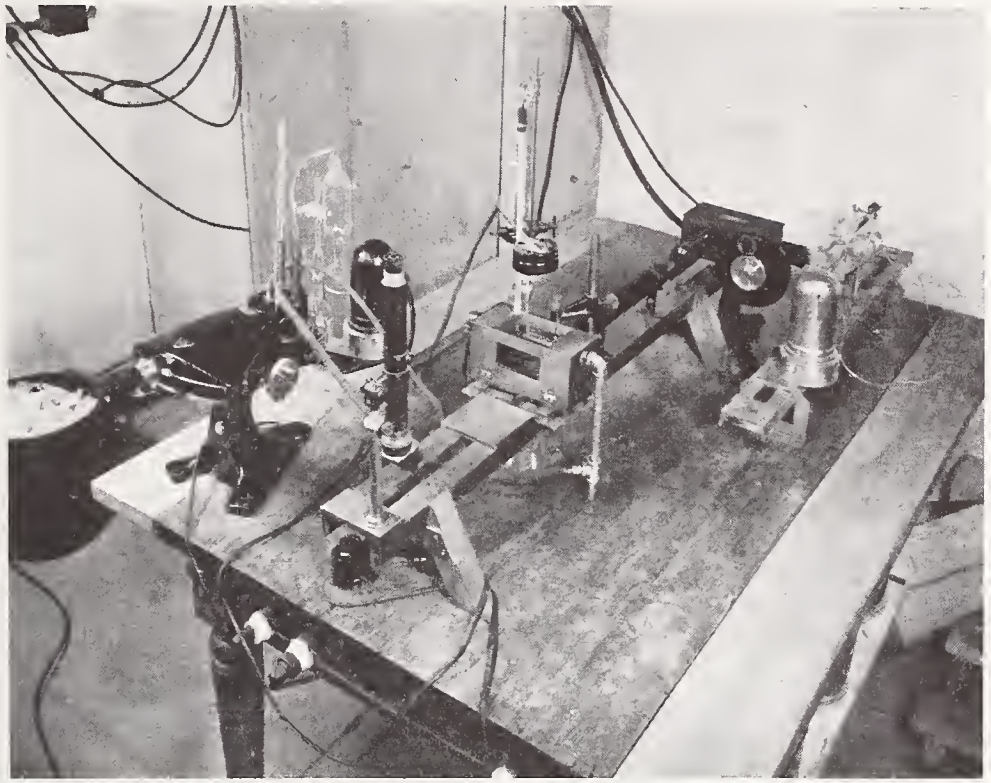


FIGURE 3.—*Complete immersion apparatus.*

IV. PROCEDURE

Six immersion liquids were prepared. The indices of refraction of these liquids were approximately 1.517, 1.572, 1.604, 1.617, 1.620, and 1.649, respectively. The first five liquids were mixtures of liquid petrolatum (n_D=1.456) and α-chloronaphthalene (n_D=1.632). The sixth liquid was a mixture of α-chloronaphthalene and α-bromonaphthalene (n_D=1.656). After isolating a particular spectrum line by means of the monochromator, the sample was immersed in the appropriate liquid, and the two were matched for index by adding small quantities of one of the pure liquids. When a good match had been obtained, the index of the immersion liquid was measured on the refractometer. This process was repeated four times for each sample. In each series of measurements, care was taken to approach the correct value from both a higher index immersion liquid and a lower one. Measurements of the temperature of the immersion liquid and of the refractometer prism block were made with a precision of $\pm 0.02°$ C. Corrections for very small temperature differences were applied by using an average change of -0.00035 in index per degree rise in temperature.

In order to obtain the highest possible accuracy in index measurements, a comparison method recommended by Tilton[8], was used. By this method, instrumental errors, such as a shift in the zero setting, can be readily compensated. Essentially this method consists in checking the refractometer, before and after each series of immersion measurements, with a standard of known index approximately the same as the rough sample being measured and making the necessary corrections. This check was made for each type of glass and for each wavelength used, by measuring the standard as a Pulfrich slab on the refractometer. In several cases brought to the attention of the authors, considerable difference was found in the readings of refractometers when measuring solid and liquid samples of the same index. Should this be the case, an alternative procedure should be used in checking the refractometer. In the alternative procedure a liquid which in each case is matched with a known sample in index is used as the "standard" instead of the solid sample.

A series of 12 glass samples was tested to obtain the accuracy of the instrument. These samples were cut from prisms that had been previously measured on a spectrometer. The indices of refraction of these test pieces were known to the fifth decimal place. Since these samples were in the shape of rectangular prisms approximately 2 by 2 by 0.2 cm in size and had polished faces, and since the instrument was designed for measuring rough, irregularly shaped specimens, it was necessary to compare the accuracies of measurement obtainable with rough pieces and finished samples. This was done by measuring the indices of refraction of four rough pieces of glass and then grinding these pieces to the same shape as the standards and then remeasuring the indices. The results of this test show that any difference that does exist is within the error of measurement.

[8] L. W. Tilton, J. Opt. Soc. Am. 32, 371-381 (1942).

TABLE 2.—*Comparison of measurements made on rough, irregularly shaped samples and on fine-ground, rectangular specimens*

Index of refraction (n_D)			Index of refraction (n_D)		
	Rough sample	Ground sample		Rough sample	Ground sample
	SAMPLE I			SAMPLE III	
1	1.51580	1.51578	1	1.60179	1.60178
2	1.51582	1.51581	2	1.60182	1.60180
3	1.51578	1.51583	3	1.60180	1.60180
4	1.51585	1.51580	4	1.60184	1.60181
Average	1.51581	1.51581	Average	1.60181	1.60180
	SAMPLE II			SAMPLE IV	
1	1.57187	1.57188	1	1.61960	1.61960
2	1.57189	1.57187	2	1.61961	1.61963
3	1.57190	1.57186	3	1.61963	1.61961
4	1.57188	1.57190	4	1.61962	1.61962
Average	1.57189	1.57188	Average	1.61962	1.61962

V. RESULTS

The values obtained in the test of the rough blanks are given in table 2. A total of 48 measurements were made on the 12 regular samples for each of the wavelengths, corresponding to the sodium *D* line, and hydrogen *F* and *C* lines. The results of these measurements are tabulated in table 3. Table 3 also shows the errors in the measured

TABLE 3.—*Comparison of immersion determinations of index of refraction and dispersion with spectrometer values*

Method	n_D	Error $\times 10^5$	n_F	Error $\times 10^5$	n_C	Error $\times 10^5$	μ	Error
			SAMPLE I					
Spectrometer	1.51651	-------	1.52211	-------	1.51411	-------	64.6	-------
Immersion 1	1.51653	+2	1.52213	+2	1.51412	+1	64.6	0
Immersion 2	1.51653	+2	1.52209	−2	1.51413	+2	64.9	+0.3
Immersion 3	1.51650	−1	1.52210	−1	1.51410	−1	64.6	0
Immersion 4	1.51651	0	1.52208	−3	1.51412	+1	64.9	+.3
Average	1.51652	1	1.52210	2	1.51412	1	64.8	0.2
			SAMPLE II					
Spectrometer	1.51748	-------	1.52316	-------	1.51507	-------	64.0	-------
Immersion 1	1.51746	−2	1.52318	+2	1.51509	+2	64.0	0
Immersion 2	1.51747	−1	1.52316	0	1.51510	+3	64.2	+0.2
Immersion 3	1.51750	+2	1.52317	+1	1.51508	+1	64.0	0
Immersion 4	1.51746	−2	1.52313	−3	1.51508	+1	64.3	+.3
Average	1.51747	2	1.52316	2	1.51509	2	64.1	0.1
			SAMPLE III					
Spectrometer	1.57205	-------	1.57906	-------	1.56913	-------	57.6	-------
Immersion 1	1.57202	−3	1.57909	+3	1.56914	+1	57.5	−0.1
Immersion 2	1.57202	−3	1.57904	−2	1.56909	−4	57.5	−.1
Immersion 3	1.57206	+1	1.57905	−1	1.56911	−2	57.6	0
Immersion 4	1.57206	+1	1.57906	0	1.56911	−2	57.5	−.1
Average	1.57204	2	1.57906	2	1.56911	2	57.5	0.1

TABLE 3.—*Comparison of immersion determinations of index of refraction and dispersion with spectrometer values*—Continued

Determination	n_D	Error $\times 10^5$	n_F	Error $\times 10^5$	n_C	Error $\times 10^5$	ν	Error
SAMPLE IV								
Spectrometer	1.57151		1.57850		1.56859		57.7	
Immersion 1	1.57150	−1	1.57849	−1	1.56858	−1	57.7	0
Immersion 2	1.57151	0	1.57853	+3	1.56860	+1	57.6	−0.1
Immersion 3	1.57154	+3	1.57847	−3	1.56861	+2	58.0	+.3
Immersion 4	1.57151	0	1.57846	−4	1.56858	−1	57.8	+.1
Average	1.57152	1	1.57849	3	1.56859	1	57.8	0.1
SAMPLE V								
Spectrometer	1.60423		1.61419		1.60027		43.4	
Immersion 1	1.60431	+3	1.61416	−3	1.60025	−2	43.4	0
Immersion 2	1.60430	+2	1.61416	−3	1.60024	−3	43.4	0
Immersion 3	1.60431	+3	1.61420	+1	1.60023	−4	43.3	−0.1
Immersion 4	1.60428	0	1.61419	0	1.60028	+1	43.4	0
Average	1.60430	2	1.61418	2	1.60025	3	43.4	0
SAMPLE VI								
Spectrometer	1.60454		1.61446		1.60053		43.4	
Immersion 1	1.60454	0	1.61450	+4	1.60056	+3	43.4	0
Immersion 2	1.60457	+3	1.61449	+3	1.60054	+1	43.3	−0.1
Immersion 3	1.60456	+2	1.61450	+4	1.60053	0	43.3	−.1
Immersion 4	1.60454	0	1.61447	+1	1.60057	+4	43.5	+.1
Average	1.60455	1	1.61449	3	1.60055	2	43.4	0.1
SAMPLE VII								
Spectrometer	1.61804		1.63018		1.61320		36.4	
Immersion 1	1.61804	0	1.63020	+2	1.61322	+2	36.4	0
Immersion 2	1.61803	−1	1.63019	+1	1.61323	+3	36.4	0
Immersion 3	1.61801	−3	1.63019	+1	1.61324	+4	36.5	+0.1
Immersion 4	1.61806	+2	1.63017	−1	1.61325	+5	36.5	+.1
Average	1.61804	2	1.63019	1	1.61324	3	36.5	0.1
SAMPLE VIII								
Spectrometer	1.61731		1.62942		1.61249		36.5	
Immersion 1	1.61734	+3	1.62941	−1	1.61250	+1	36.5	0
Immersion 2	1.61729	−2	1.62943	+1	1.61252	+3	36.5	0
Immersion 3	1.61728	−3	1.62937	−5	1.61252	+3	36.6	+.1
Immersion 4	1.61729	−2	1.62940	−2	1.61252	+3	36.6	+.1
Average	1.61730	3	1.62940	2	1.61252	3	36.6	.1
SAMPLE IX								
Spectrometer	1.62024		1.63248		1.61536		36.2	
Immersion 1	1.62023	−1	1.63249	+1	1.61536	0	36.2	0
Immersion 2	1.62024	0	1.63251	+3	1.61536	0	36.2	0
Immersion 3	1.62027	+3	1.63252	+4	1.61540	+4	36.2	0
Immersion 4	1.62027	+3	1.63246	−2	1.61539	+3	36.3	+0.1
Average	1.62025	2	1.63250	3	1.61538	2	36.2	0
SAMPLE X								
Spectrometer	1.62023		1.63247		1.61536		36.2	
Immersion 1	1.62022	−1	1.63246	−1	1.61539	+3	36.3	+0.1
Immersion 2	1.62024	+1	1.63247	0	1.61535	−1	36.2	0
Immersion 3	1.62025	+2	1.63242	−5	1.61537	+1	36.4	+.2
Immersion 4	1.62025	+2	1.63248	+1	1.61536	0	36.2	0
Average	1.62024	2	1.63246	2	1.61537	2	36.3	+0.1

TABLE 3.—*Comparison of immersion determinations of index of refraction and dispersion with spectrometer values*—Continued

Method	n_D	Error $\times 10^5$	n_F	Error $\times 10^5$	n_C	Error $\times 10^5$	μ	Error
			SAMPLE XI					
Spectrometer	1.64918	.....	1.66296	.....	1.64373	.....	33.8	.....
Immersion 1	1.64918	0	1.66300	+4	1.64375	+2	33.7	−0.1
Immersion 2	1.64917	−1	1.66296	0	1.64370	−3	33.7	−.1
Immersion 3	1.64915	−3	1.66296	0	1.64371	−2	33.7	−.1
Immersion 4	1.64917	−1	1.66293	−3	1.64373	0	33.8	0
Average	1.64917	1	1.66296	2	1.64372	2	33.7	0.1
			SAMPLE XII					
Spectrometer	1.64845	.....	1.66220	.....	1.64301	.....	33.8	.....
Immersion 1	1.64845	0	1.66215	−5	1.64301	0	33.9	+0.1
Immersion 2	1.64847	+2	1.66217	−3	1.64299	−2	33.8	0
Immersion 3	1.64842	−3	1.66218	−2	1.64296	−5	33.7	−.1
Immersion 4	1.64845	0	1.66218	−2	1.64306	+5	33.9	+.1
Average	1.64845	1	1.66217	3	1.64300	3	33.8	0.1

values as compared with the spectrometer values. An examination of these errors shows that the maximum error, or degree of accuracy, for any single measurement is $\pm 5 \times 10^{-5}$, but the average error for the 144 measurements is only $\pm 1.9 \times 10^{-5}$. The reproducibility, or precision, of these measurements, expressed as the average of the deviations from the mean value of each set of four immersion measurements, is $\pm 1.5 \times 10^{-5}$.

A summary of the average accuracy and precision of all the determinations is given in table 4.

TABLE 4.—*Summary of accuracy and precision of determinations listed in table 3*

	Average error	Average deviation from mean values
n_D	1.6×10^{-5}	1.4×10^{-5}
n_F	2.1×10^{-5}	1.6×10^{-5}
n_C	2.1×10^{-5}	1.6×10^{-5}
All values	1.9×10^{-5}	1.5×10^{-5}

The accuracy attained in these measurements is almost as high as the precision attainable on a refractometer. This indicates that the errors introduced in determining the match between the indices of the glass and liquid are for the most part negligible. According to Tilton,[9] it would be very difficult to obtain results of this accuracy without the strict use of a comparison or substitution method. The success attained may be largely attributed to the fact that in measuring a known standard, on the refractometer both before and after each group of four immersion measurements, a comparison method was used that approximates a substitution method. Other factors which should be mentioned are (1) temperature differences between the refractometer and immersion tank were rarely greater than 0.1°, and these temperature effects were considered in obtaining the final

[9] L. W. Tilton, J. Research NBS 30, 311-328 (1943) RP1535.

values, and (2) the recommended tests to eliminate shielding of rays, setting on false edges, etc. (see footnote 9) were frequently made.

Although the errors introduced in matching are, in general, negligible, the experience of the operator plays a part in these determinations. This is due to surface effects that interfere with the determination of a match. The ideal condition is one which produces a uniformly illuminated field against which the sample is invisible. This exists very rarely in actual practice. Usually, under conditions of match, the main body of the sample will blend with the field, but the edges of the sample will be outlined by very thin brilliant lines. In these cases, a match is determined by rotating the sample through 180°. There should be no change in the thickness or brilliance of these lines when this is done. Thus any confusion due to surface effects may be avoided. The experience of the operator will also determine the speed with which measurements are made. An experienced person can measure about six samples in an 8-hour day. Thus it is possible to secure complete information on the index of refraction and ν value for a particular melt within 1½ hours after a pot of glass is opened.

VI. SUMMARY

An improved apparatus for determining indices of refraction and dispersion of glass by immersion is described. The results of 144 measurements indicate that the index of refraction may be determined with an average error of $\pm 1.9\times10^{-5}$ and a maximum error of $\pm 5\times10^{-5}$ Dispersion as represented by ν values may be calculated from these measurements with an average error of 0.1 and a maximum error of 0.8. This apparatus may, therefore, be used both for routine production control work and for the determination of the optical constants of finished optical components, especially of flint glasses. It may also be applied profitably to research in the determination of the effect of composition on the optical properties of glass, especially where the glasses studied may be unstable. The full possibilities of the matching instrument have not been determined. More accurate refractometers and temperature controls are necessary for this purpose.

WASHINGTON, October 1, 1943.

Refractive Index of Potassium Bromide for Infrared Radiant Energy*†

ROBERT E. STEPHENS, EARLE K. PLYLER, WILLIAM S. RODNEY, AND ROBERT J. SPINDLER
National Bureau of Standards, Washington, D. C.

(Received October 23, 1952)

The index of refraction of potassium bromide at 22°C has been measured for the infrared region from 1 micron to 25 microns by means of two methods on two different instruments. The two resulting sets of indexes agree to within the probable error of either one except at the two longest wavelengths. A five constant expression has been found that represents the index as a function of wavelength over the region from 0.4047 microns to 25 microns. Indexes were measured at several of the wavelengths at temperatures higher and lower than 22°C and thermal coefficients of index were computed.

ALTHOUGH potassium bromide has long been one of the important materials for prisms for use in infrared spectroscopy, particularly in the range of wavelengths from 15 to 25 microns, the discrepancies among published values of the index of refraction have prevented its service as a standard material for precise use in calibrations. The inconsistency may perhaps be a consequence of different methods being used in different laboratories and the fact that no one investigator covered the total range of wavelengths. In order to provide such consistent data, a large prism of this material was made available to the authors by the Perkin-Elmer Corporation.

The index of refraction and the thermal coefficients of index of this prism, for a number of wavelengths in the visible region of the spectrum have been reported by Rodney and Spindler.[1] This paper reports indexes of refraction principally for the infrared region from 1 micron to 25 microns.

At the time the index determination became necessary the only instrument available that was considered sufficiently accurate was a Perkin-Elmer infrared spectrometer. This instrument is of the Littrow type in which the prism remains fixed and the Littrow mirror is rotated by means of a screw. The use of this instrument for measuring index has been described in connection with measurements of index of thallium bromide-iodide[2] and silver chloride.[3] Data were obtained with this spectrometer at only one temperature, 22°C. The resulting indexes are shown in Table I, column five, designated "Known incidence."

Meanwhile, a new Gaertner precision spectrometer was acquired by the National Bureau of Standards. With this spectrometer were a pair of concave mirrors and associated parts with which the collimator and telescope could be replaced, making the spectrometer usable in the infrared. Mountings were constructed so that either of two physical detectors could be attached to the movable arm and receive the flux from the exit slit. A photograph of this instrument and associated apparatus is shown in Fig. 1.

The physical detectors used were a lead sulfide photoconducting cell for the range from the visible region to a wavelength of approximately 2.4 microns (it is also sensitive in the visible region and the near ultraviolet) and a Golay pneumatic detector, which is usable to a wavelength of about 40 microns.

The scale for measuring angles on this Gaertner spectrometer is a highly precise circle on a disk of stainless steel, which was ground and polished optically flat and ruled at the National Bureau of Standards. It is believed that this circle, when used with opposite reading microscopes can be used to measure angles with errors no larger than 1 second.

This spectrometer is equipped with a gear system, which may be engaged at will, that rotates the table through half the angle through which the collector arm is turned. This makes it possible to turn the spectrometer to any spectrum line while maintaining the prism in the position of minimum deviation once this condition has been established for any one wavelength. A visible spectrum line is used in establishing this condition.

The ability of this spectrometer to measure the angle of minimum deviation for an invisible spectrum line is not quite as good as it is for visible lines where the two positions of minimum deviation can be realized and the angle between them (twice the deviation) measured; also the precision with a physical detector is not equal to that with the eye. With the mirrors for collimating and collecting, the collecting arm can be used only on one side of zero deviation and even there the position of zero deviation cannot be reached. We can measure only the difference in the angles of minimum deviation for two spectrum lines, then if the angle for one of them is known that for the other may be obtained by addition.

In this investigation the angles of minimum deviation for several lines in the visible spectrum were measured on a Watts spectrometer. These lines were then observed on the Gaertner spectrometer and the scale reading corresponding to zero deviation was computed for all

* The work described was carried out in part under the sponsorship of the U. S. Air Force and U. S. Army Signal Corps.

† The results of this work were presented by Robert E. Stephens at the New York meeting of the Optical Society of America, March 21, 1952.

[1] W. S. Rodney, Masters dissertation, Catholic University of America, Washington, D. C. (1951). Also W. S. Rodney and R. J. Spindler, J. Research Natl. Bur. Standards 49, 253 (1952).

[2] Tilton, Plyler, and Stephens, J. Research Natl. Bur. Standards, R. P. 2008, 43, 81 (1949).

[3] Tilton, Plyler, and Stephens, J. Opt. Soc. Am. 40, 540 (1950).

TABLE I. Index of refraction of potassium bromide.

Wavelength (microns)	Source	Index computed by Eq. (1)	Observed index: Minimum deviation	Observed index: Known incidence	Observed minus computed: Minimum deviation	Observed minus computed: Known incidence
0.404656	Hg	1.589752	1.589753		-0.1×10^{-5}	
0.435835	Hg	1.581479	1.581475		−0.4	
0.486133	H(F)	1.571791	1.571789		−0.3	
0.508582	Cd	1.568475	1.568479		−0.4	
0.546074	Hg	1.563928	1.563930		+0.2	
0.587562	He	1.559965	1.559966		+0.1	
0.643847	Cd	1.555858	1.555853		−0.5	
0.706520	He	1.552447	1.552447		0.0	
1.01398	Hg	1.54408	1.54410	1.54408	+2	0×10^{-5}
1.12866	Hg	1.54258	1.54265	1.54264	+7	+6
1.36728	Hg	1.54061	1.54063	1.54060	+2	−1
1.7012	Hg	1.53901	1.53904	1.53905	+3	+4
2.44	1,2,4-trichlorobenzene	1.53733	···	1.53738	···	+5
2.73	methanol	1.53693	···	1.53695	···	+2
3.419	polystyrene	1.53612	1.53614	1.53616	+2	+4
4.258	CO_2	1.53523	···	1.53523	···	0
6.238	polystyrene	1.53288	1.53286	1.53284	−2	−4
6.692	polystyrene	1.53225	1.53220	1.53219	−5	−6
8.662	polystyrene	1.52903	1.52903	1.52901	0	−2
9.724	polystyrene	1.52695	1.52689	1.52689	−6	−6
11.035	polystyrene	1.52404	1.52401	1.52403	−3	−1
11.862	methyl-cyclohexane	1.52200	···	1.52199	···	−1
14.29	polystyrene	1.51505	···	1.51495	···	−10
14.98	CO_2	1.51280	1.51282	1.51286	+2	+6
17.40	1,2,4-trichlorobenzene	1.50390	1.50378	1.50400	−12	+10
18.16	1,2,4-trichlorobenzene	1.50076	1.50066	1.50070	−10	−6
19.01	water vapor	1.49705	1.49701	1.49704	−4	−1
19.91	water vapor	1.49288	1.49280	1.49293	−8	+5
21.18	water vapor	1.48655	···	1.48664	···	+9
21.83	water vapor	1.48311	1.58305	1.48307	−6	−4
23.86	water vapor	1.47140	1.47123	1.47138	−17	−2
25.14	water vapor	1.46324	1.46303	1.46322	−21	−2

of these lines. These values were averaged for greater accuracy and the deviations for the invisible lines were computed from this reference. It is estimated that the total error in the measured value of the angle of minimum deviation does not exceed two seconds. The corresponding error in index would be less than 7×10^{-6}. The prism angle was also measured visually on the Watts spectrometer to within ± one-half second. The total resulting error in index should be less than 1×10^{-5}.

For those indexes corresponding to the sharp emission lines of the mercury spectrum whose wavelengths are accurately known, the corresponding indexes should be accurate to $\pm1\times10^{-5}$ as indicated above. However, because of the broadness of some of the absorption bands used to identify the longer wavelengths, there is considerable difficulty in locating the peaks; consequently, the wavelength to which the measured index corresponds may be in error to some extent. The infrared wavelengths tabulated are mostly from the list published by Plyler and Peters.[4] A few additional values are included and several have been adjusted because the continuous background varied with wavelength. The wavelength 2.44 microns is an estimate of the center of an absorption band that is unresolved into its components because of the low dispersion of potassium bromide in this neighborhood. The same is true of the wavelength 2.73 microns which is produced by methanol dissolved in carbon tetrachloride.

The movable arm of the spectrometer could be moved at approximately constant angular speed by a synchronous motor. The response of the detector could be recorded by a Speedomax recorder in which the record paper was also propelled by a synchronous motor. Thus, a chart showing detector response as a function of spectrometer position could be made. A second pen on the recorder, actuated by a relay, permits marks to be made on the chart at points corresponding to known positions of the spectrometer. By the use of these charts,

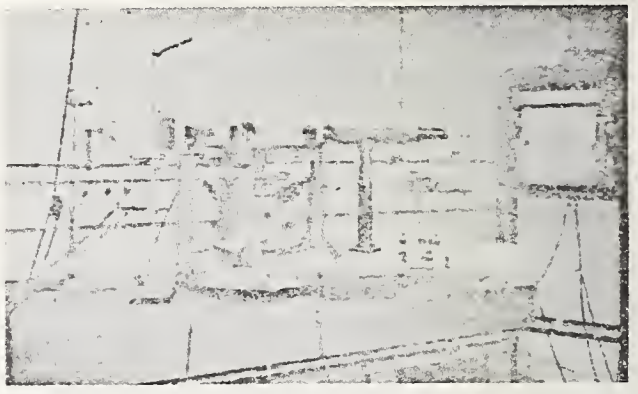


FIG. 1. Gaertner precision spectrometer and associated equipment for infrared measurements.

[4] Earle K. Plyler and C. Wilbur Peters, J. Research Natl. Bur. Standards 45, 462 (1950).

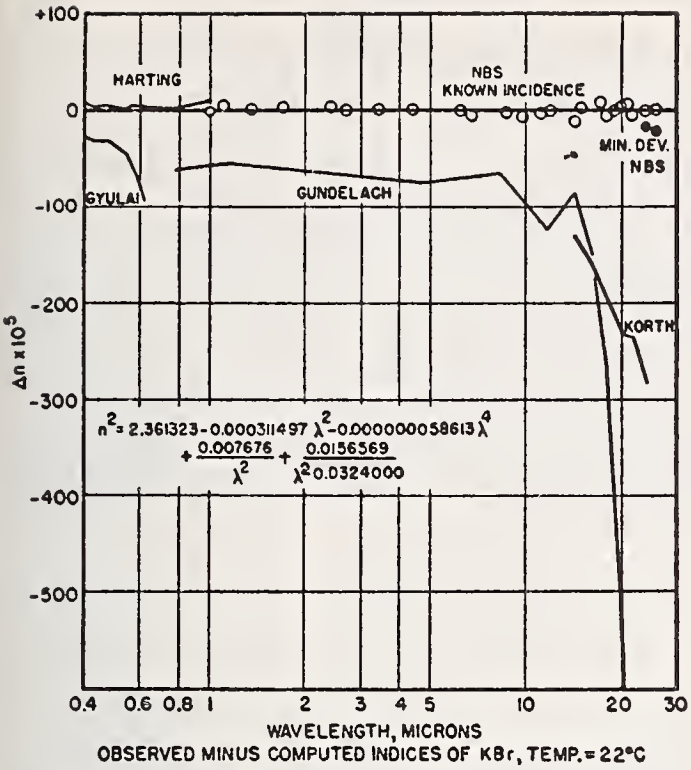


FIG. 2. Comparison of National Bureau of Standards values of index of potassium bromide with those of other observers.

absorption bands and emission lines may be identified and the approximate corresponding positions of the spectrometer located.

The scale-reading corresponding to an emission line or an absorption band was determined accurately by making spectrometer settings by hand for maximum or minimum response. This was done repeatedly and the results averaged for greater accuracy.

Dispersion formulas of several forms were fitted to the indexes obtained with the Perkin-Elmer spectrometer in the infrared region together with indexes at eight wavelengths in the visible region from the work of Rodney and Spindler. The following equation, whose constants were evaluated by least squares, was found to fit the data satisfactorily:

$$n^2 = 2.361323 - 0.00311497\lambda^2 - 0.000{,}000{,}058613\lambda^4 + \frac{0.007676}{\lambda^2} + \frac{0.0156569}{\lambda^2 - 0.0324}. \quad \text{(I)}$$

After the infrared measurements with the Gaertner spectrometer were completed, it was found that the formula also fitted those indexes almost as well except at the two longest wavelengths.

Some measurements were also made at temperatures of 13.5°C and 31.0°C and temperature coefficients of index were computed. These data are not sufficiently precise to determine dependence on wavelength but averaged about 4.0×10^{-5} per centigrade degree. This compares with an average of 3.9×10^{-5} per degree as found by one of the authors[2] for the visible region. These values are lower than the value, 4.4×10^{-5} reported by Forrest[5] and higher than the value of 3.6×10^{-5} found by Korth,[6] both for 5461 angstroms. The temperature coefficients given by Harting[7] range both higher and lower than any of the values here mentioned.

For comparison the indexes reported by Korth,[6] Harting,[7] Gundelach,[8] and Gyulai[9] have been adjusted to 22°C and the differences from the values of Eq. (I) are shown graphically in Fig. 2, together with the observed minus computed values of the present work.

---

[5] J. W. Forrest, J. Opt. Soc. Am. 32, 382 (1942).
[6] K. Korth, Z. Physik 84, 677 (1933).
[7] H. Harting, Z. Instrumentenk. 63, 125 (1943).
[8] E. Gundelach, Z. Physik 66, 775 (1930).
[9] Z. Gyulai, Z. Physik 46, 80 (1927).

Journal of Research of the National Bureau of Standards Vol. 59, No. 2, August 1957 Research Paper 2779

Relative Stress-Optical Coefficients of Some National Bureau of Standards Optical Glasses[1]

Roy M. Waxler and Albert Napolitano

The relative stress-optical coefficients of 27 optical glasses made at the National Bureau of Standards have been determined. The glasses were loaded in compression, and the amount of birefringence was measured with a polarimeter, using a quarter-wave plate compensator. The stress-optical coefficients determined for the crown glasses ranged between +2.95 and +1.82 brewsters. For the flint glasses, a regular variation of the stress-optical coefficient with weight percent of lead oxide was found with values ranging from +3.18 to −1.16 brewsters.

1. Introduction

The double refraction produced in isotropic solids by the application of stress was discovered by Brewster [1][2] who, incidentally, used glass for the greater part of his experiments. From Brewster's early work, the theory of photoelasticity has been developed [2] and is the basis of a recognized method of stress analysis for isotropic transparent solids.

For a given material, the relationship between the optical-path difference and the applied stress is

$$r = Ctd. \tag{1}$$

C is called the relative stress-optical coefficient and is characteristic of the type of glass under investigation. The above equation holds true for both compression and tension [3], and it is presumed to be valid up to the breaking stress of glass [4]. When r, the retardation, is expressed in angstroms; t, the stress, in bars; and d, the thickness, in millimeters, the resulting coefficient, C, is measured in brewsters.

The effect of the chemical composition of glass on the photoelastic effect was first investigated by Pockels [5], who came to the conclusion that an increase of lead oxide, or decrease of boric oxide, always lowers the relative stress-optical coefficient. In compression most glasses act as uniaxial negative crystals, but Pockels succeeded in observing a reversal of this effect for very heavy lead silicate glasses. The effect of the composition was more accurately studied by Filon [6] and Adams and Williamson [7], who reached the same general conclusions.

In more recent work, Vedam [8] measured the elastic and photoelastic properties of 18 optical glasses, and Schaefer and Nassenstein [9] made a very extensive study of these properties for 154 optical glasses produced by Schott & Co. Unfortunately, the compositions of the glasses measured by Vedam are given only to a rough approximation, and these data are not mentioned in the paper of Schaefer and Nassenstein.

As the relative stress-optical coefficient relates the stress exerted upon a glass to the optical-path difference produced, a knowledge of this coefficient permits one to carry out a stress analysis of a glass specimen on the basis of the optical evidence of the existing strain. Furthermore, Lillie and Ritland [10] have shown that a knowledge of this coefficient for optical glasses is needed in order to develop fine annealing schedules that will keep the amount of birefringence within prescribed limits. In the course of developing annealing schedules for large optical elements, it was felt desirable to evaluate these coefficients for several glasses made at the Bureau. Because these are similar to commercial types of optical glasses, the results are offered for their potential interest to glass technologists.

2. Optical System

In most investigations of the relative stress-optical coefficient, the stressed specimen is placed in a 45-deg position between crossed nicols. A Babinet compensator, also placed between the crossed nicols and at the same angle as the specimen, is used to measure the optical-path difference. However, the use of a quarter-wave plate as a compensator affords greater accuracy, and was used in this investigation [11]. With this arrangement, the polarizer is set at a 45-deg position with respect to the principal directions of the stressed specimen, and the elliptically polarized light emerging from the glass is restored to a linear polarization by a properly oriented quarter-wave plate. The quarter-wave plate and stressed glass in combination then have the same effect as an optically active material that rotates the plane of polarization of the incident light. The angular difference in the extinction positions of the analyzer may be noted if the analyzing nicol is mounted in a graduated circular scale. The amount of rotation is equal to one-half the phase angle, and may be used to measure the optical-path difference in the birefringent glass.

The optical arrangement and loading device used in this investigation are shown in figure 1. The light source consisted of a laboratory sodium arc lamp. The light was collimated and filtered through a glass cell 15 mm in length that contained a 6-percent solu-

[1] This work was sponsored by the Air Research and Development Command, U. S. Air Force under Contract No. AF 18 (600)-139.

[2] Figures in brackets indicate the literature references at the end of this paper.

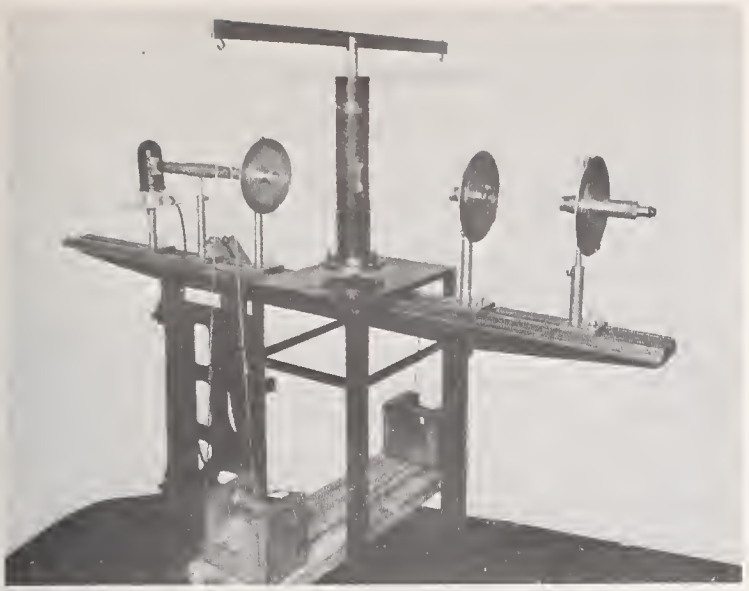


FIGURE 1. *Optical system and loading device for determining the stress-optical coefficient of glasses.*

tion of potassium dichromate. The wavelength was then taken as the mean of the two sodium D lines, or 5893 A.

The polarizer, quarter-wave plate, and analyzer were each mounted in a brass tube 3¼ in. long and 1¼ in. in diameter. Nicol prisms were used for polarizer and analyzer, and the quarter-wave plate was a Babinet compensator adjusted to give the proper retardation.

The circles for measuring the amount of rotation were 8½ in. in diameter and graduated to 0.5 deg, with a vernier permitting readings to be made to 0.1 deg. A small telescope was mounted in the tube that held the analyzing nicol and was focused upon a ⅛-in.-diameter hole, which formed a stop for the light beam as it emerged from the polarizer.

3. Preparation of Samples

The optical glasses tested are shown in table 1, together with their chemical compositions as computed from the raw materials in the batch. Column 4 of table 1 identifies the glass according to optical type, refractive index, and Abbe valve [12].

These glasses were made in clay pots holding from 1,000 to 2,000 lb of glass, depending on the type of glass. A 5-lb piece of high quality glass,[3] which was representative of the type of glass under investigation, was selected from each melt. A test specimen was cut from each piece of glass and annealed. During heat treatment, each specimen was enclosed in a metal box in order to eliminate temperature gradients. The specimen was supported on small bits of refractory material so that it was not in contact with any part of the metal. The glasses were heated to a temperature about 20 deg C below the deformation point [13], held at this temperature for several hours, and then cooled at a rate of 1 deg C/hr through the transformation region. Once through the transformation region, the glass was rapidly cooled to room temperature.

Each specimen of glass was annealed and then carefully ground to the shape of a rectangular block, the approximate dimensions being 3 cm by 3 cm by 10 cm. Two opposite faces 3 cm by 10 cm were

[3] High-quality glass—seed free, striae free, having a nominal index of refraction, and a nominal dispersion.

TABLE 1. *Relative stress-optical coefficient and compositions (computed from batch) of NBS optical glasses*

Specimen	C, brewsters, $\frac{10^{-13}\ cm^2}{dyne}$	Standard error of C	Glass type	Melt	Component oxides (weight %)									Other oxides	Specimen
					SiO_2	PbO	BaO	B_2O_3	Na_2O	K_2O	ZnO	As_2O_3	Sb_2O_3		
1	2.93	0.027	CF 529/516	8217	65.4	10.0	0.2	----	13.2	5.6	3.6	0.2	1.8		1
2	2.97	.029	F 5725/425	7841	55.1	31.7	1.0	----	5.0	6.9	----	.3			2
3	3.18	.029	F 5795/410	7073	53.1	35.5	0.6	----	0.4	9.6	----	.3	0.5		3
4	2.98	.016	F 605/380		47.6	40.9	----	----	2.2	8.8	----	.5	----		4
5	2.77	.014	F 617/366	8853	45.6	43.2	----	----	4.6	6.1	----	.5	----		5
6	2.89	.031	F 620/362	8622	45.6	45.5	----	----	2.7	5.7	----	.5	----		6
7	2.65	.015	F 649/338	8865	41.2	51.1	----	----	0.7	6.5	----	.5	----		7
8	2.45	.026	F 666/324	7686	39.3	54.4	----	----	----	6.0	----	.3	----		8
9	2.18	.020	F 689/309	8652	37.0	58.1	----	----	----	4.6	----	.3	----		9
10	1.77	.010	F 720/293	8968	34.1	62.4	----	----	----	3.2	----	.3	----		10
11	1.33	.013	F 754/277	8001	31.2	66.2	----	----	----	2.3	----	.3	----		11
12	−1.16	.039	F 920/211	A2059	20.0	79.6	----	----	----	0.4	----	----	----		12
13	2.73	.026	BaC 541/599	8700	58.8	----	19.9	3.8	2.8	10.3	4.1	.3	----		13
14	2.56	.025	BaC 5725/574	8702	45.2	----	28.7	5.8	1.8	7.0	7.5	.3	.7	Al_2O_3 3.0	14
15	2.65	.019	BaC 574/577	8573	47.4	----	30.0	4.9	0.9	7.0	7.5	.3	.5	Al_2O_3 1.5	15
16	1.85	.012	BaC 6109/572	8462	38.2	0.2	42.9	6.7	----	----	4.2	.4	.2	Al_2O_3 4.9, CaO 2.3	16
17	1.83	.013	BaC 611/588	8621	38.3	.2	42.6	10.7	----	----	----	.4	.2	CaO 4.5, Al_2O_3 3.1	17
18	2.03	.013	BaC 617/550	8890	37.3	1.4	44.8	4.4	.2	0.4	5.6	.3	.7	Al_2O_3 4.9	18
19	1.82	.018	BaC 620/600	7903	37.8	----	44.2	11.3	----	----	----	.3	.2	CaO 3.6, BeO 2.3, ZrO_2 0.3.	19
20	2.95	.022	BaF 584/460	7631	49.8	18.8	13.4	----	1.5	8.2	7.8	.5	----		20
21	2.35	.026	BaF 588/534		45.8	5.6	28.0	4.4	2.3	6.0	7.5	.4	----		21
22	2.84	.037	BaF 605/436	8610	45.7	23.3	14.3	----	----	8.2	8.1	.4	----		22
23	2.78	.016	LC 512/605	8691	71.9	----	----	5.0	14.7	5.0	----	----	1.2	CaO 2.2	23
24	2.76	.016	LC 523/586	8695	70.2	----	----	1.5	14.0	2.5	----	----	1.2	CaO 9.4, Cl 0.7, SO_3 0.5.	24
25	2.66	.016	LC 528/580	7617	68.3	----	----	1.5	14.0	2.5	----	----	1.2	CaO 11.3, Cl 0.7, SO_3 0.5.	25
26	2.75	.021	BSC 511/635	8713	68.8	----	----	8.5	7.8	14.7	----	.2	----		26
27	2.81	.019	BSC 517/645	8843	66.3	0.5	----	12.5	7.5	12.8	----	.4	----		27

polished, in order to transmit the light beam. Attention is called to the fact that the length of the specimen was made three times its width to avoid end effects [14].

After annealing, grinding, and polishing, the residual strain was measured. An optical path difference of no more than 2 or 3 mμ/cm was observed. This nominal residual strain was presumed to remain constant during the tests. A zero correction was unnecessary as the relative stress-optical coefficient is determined from differences in applied load.

4. Method of Loading

The load was applied by means of a vertical shaft which supported a horizontal yoke, and suspended loading platform, as shown in figure 2. The shaft and the two vertical guides were machine polished to insure a smooth and close fit, thus allowing only vertical motion to the shaft. A ⅜-in.-diameter steel ball was seated in the center of a metal disk at both ends of the glass specimen. Seats were also made on the end of the shaft and the bottom steel support to accommodate the ball. When this assembly was inserted under the shaft, the axis of the shaft and glass coincided. By taping the glass near both ends, the specimen was centrally located within each disk. A light film of oil was placed at the glass and metal interface. This method has been described in a previous paper [15], and it has proved quite satisfactory in producing a uniform stress.[4]

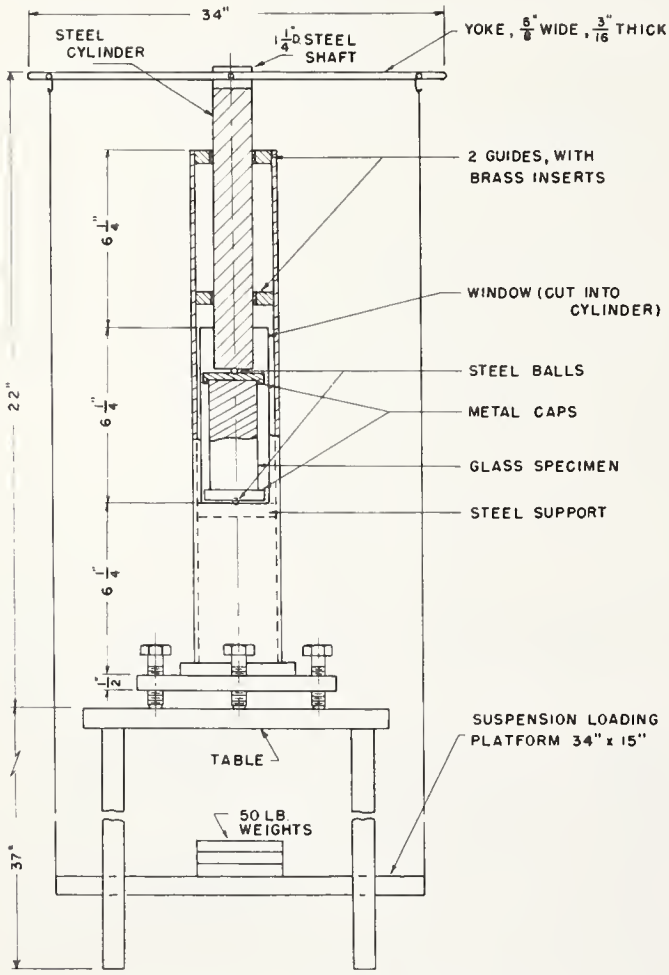


FIGURE 2. *Loading apparatus.*

By means of leveling screws and a bubble level, the steel cylinder was adjusted into a vertical position, making it possible to rest the glass on a horizontal surface. The glass, being then in vertical position, could be placed in alinement with the direction of force created by the load. This line of force established the directions of the planes of polarization of light transmitted by the glass.

To apply the stress, 50-lb weights were supplied by the Mass and Scale Section of the Bureau. Upon calibrating these weights, it was found that they were accurate to within 0.007 lb, which was felt to be more than satisfactory because it was desired to determine the values of the stress-optical coefficients only to three significant figures.

5. Procedure

Each part of the optical train was adjusted so as to be perpendicular to the light path. This was accomplished by noting the reflection of light from each surface and making certain that the reflected beam of light coincided with the incident beam.

It was necessary also to have some base point on the graduated circle of the analyzer in order to orient the polarizer and quarter-wave plate in 45-deg positions relative to the optic axes of the stressed glass. For this reason each specimen was loaded to 300 lb, and one of the optic axes was found by alternately rotating the polarizer and analyzer until a position of minimum intensity was reached. Extinction positions noted on the analyzer and polarizer could be reproduced to within ±0.1 deg of arc. Both axes were found in this manner, and a point midway between the two, as noted on the analyzer, was used as a base point for subsequent settings of the polarizer and quarter-wave plate. These elements then could be set to within ±0.1 deg. This method was felt to be satisfactory, considering the magnitude of errors that might possibly be introduced by missettings [11].

The birefringence produced in stressed glass may be used to ascertain the uniformity of stress distribution, as well as its magnitude. Before making any measurements, several of the specimens were loaded and examined in a broad beam of white light between sheets of Polaroid in a crossed position. With the aid of a sensitive red-tint plate it could be seen that the whole central portion of the specimen had a uniform color.

In making measurements, the glass was weighted with a tare consisting of the steel shaft, yoke, and loading platform, and the extinction position of the analyzer was noted. Then a 50-lb weight was added to the loading platform, and another extinction position noted. This process was repeated with 50-lb weight increments until the maximum load was 300 lb. Waiting about 5 min before readings,

[4] With this method of loading, the vertical thrust passes exactly centrally through the block, and it is a well-known principle in mechanics that at a sufficient distance from the point of application of a load, the stress system depends only on the statical resultant of the load.

the measurements were made in a darkened room so that the eye might be more sensitive to the intensity of the light beam viewed through the analyzer. The room was kept at a temperature of approximately 20° C.

It was noted that a specimen would exhibit the same extinction position before and after the application of the 300-lb load. This was checked with several glasses, and no difference could be detected outside the error of the instrument.

In investigating a particular specimen, 10 measurements were made for each extinction position, making a total of 70 measurements.

6. Results and Discussion

The results of this investigation are given in table 1, which lists the values of the relative stress-optical coefficient and their standard errors. The values were determined from the measurements in the following manner.

If Δw is the load increment in pounds, d the thickness of the specimen in inches, and b the width of the specimen in inches, we obtain

$$r=\frac{1.752C(\Delta w)d}{bd}, \tag{2}$$

where r is the retardation in angstroms, and C is the stress-optical coefficient in brewsters. The number, 1.752, is a conversion factor for the units employed.

As discussed in section 2, the retardation can also be expressed in terms of the angle through which the plane of polarization is rotated. The equation reads

$$r=\frac{2(\Delta y)\lambda}{360.0}, \tag{3}$$

where $2\Delta y$ is twice the rotation of the analyzer in degrees of arc and λ is the wavelength of the light in angstroms.

The angular rotation of the plane-polarized light is a linear function of the applied load. The increment in angular rotation can be represented by the relation

$$\Delta y=B(\Delta w) \tag{4}$$

from eq (2) and (3), where B is the slope of the straight line. It is appropriate to consider increments rather than accumulated total rotations, because, for each glass, the measurements corresponding to the various loads were all made on the same specimen, and consequently a cumulative effect may be expected in the errors of the accumulated rotation values. Errors in the optical measurements, on the other hand, are effectively reduced through the compensating effect of averaging 10 duplicate measurements at each load. A detailed statistical analysis shows that this reduced error in the optical measurement is negligible in comparison with the experimental errors due to the method of applying the load. Under these conditions, the correct least-squares estimate of slope B is obtained by averaging the five values of Δy, corresponding to the five consecutive 50-lb increments. From these five values it is also possible to calculate the precision of the evaluated slope. Each slope estimate was converted into a C value, expressed in brewsters, and the standard error of C was derived from that of the corresponding slope. Table 1 lists the C values and their standard errors. A measure of uncertainty (95% confidence interval) can be obtained for these values by considering the interval $C \pm$(2.8 times standard error).

As shown in table 1 most of the glasses studied in this investigation have values of stress-optical coefficient that are fairly close together, ranging from 2 to 3 brewsters. Two glasses in table 1, Nos. 16 and 17, vary only slightly in composition, and may be seen to have almost the same value of stress-optical coefficient, 1.85 and 1.82 brewsters, respectively.

Because optical glasses are in general made up of so many different chemical components, it is difficult to correlate changes in physical properties with a particular constituent. As the series of flint glasses consists principally of silica and lead oxide, it is interesting to note that, when the values of the stress-optical coefficient are plotted against percentage of lead oxide, a regular variation is indicated. This confirms the observations of other investigators, notably Pockels[5] [5] and Filon [6], whose results are shown in figure 3, together with the values obtained in this investigation. Furthermore, flint glass F920/211 is seen to depart from the general rule that glass in compression acts as a uniaxial negative crystal.

[5] Pockels determined the absolute changes in the refractive index of a stressed glass for light polarized, parallel, and perpendicular to the direction of force, and expressed his results by the coefficients p/r and q/r. With knowledge of these values, the relative stress-optical coefficient, C, may be calculated for purposes of comparison. This was done by Filon, and his results are shown in figure 3. Adams and Williamson also calculated C for Pockels' glasses and found that the results showed good agreement with their own determinations.

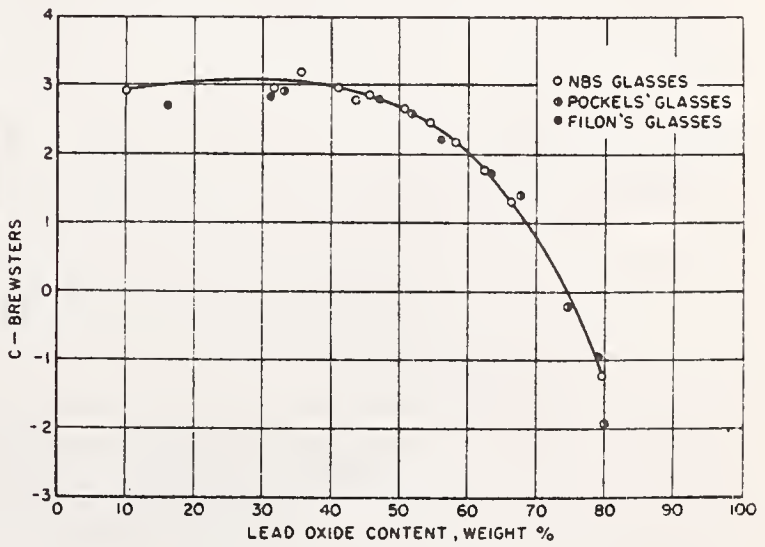


FIGURE 3. *Variation of the stress-optical coefficient with lead oxide content in flint glasses.*

Pockels [5] and Filon [6] also concluded that lowering the amount of B_2O_3 in a series of glasses decreased the relative stress-optical coefficient. It was impossible to verify this from the results of the present investigation because the glasses studied here were multicomponent and contained comparatively little boric oxide, the amount being, at most, only 12.5 percent.

The authors are deeply appreciative of the statistical analysis of the data by John Mandel and Mary N. Steel. They are also grateful to Joseph M. Nivert, Jr., for preparation of the glass specimens for this work.

7. References

[1] Sir David Brewster, Phil. Trans. (105), 60 (1815); Phil. Trans. (106), 156 (1816); Trans. Roy. Soc. Edinburgh **8,** 369 (1818).
[2] E. G. Coker and L. N. G. Filon, Treatise on photo-elasticity (Cambridge University Press, London, 1931).
[3] S. R. Savur, Phil. Mag. **50,** 453 (1925).
[4] L. N. G. Filon, A manual of photo-elasticity for engineers, p. 33–36. (Cambridge University Press, London, 1936).
[5] F. Pockels, physik Z. **2,** 693 (1901); Ann. Physik **7, 745** (1902); **9,** 220 (1902); **11,** 651 (1903).
[6] L. N. G. Filon, Proc. Roy. Soc. **79,** 440 (1907); **83,** 572 (1910).
[7] L. H. Adams and E. D. Williamson, J. Wash. Acad. **9,** 609 (1919).
[8] K. Vedam, Proc. Indian Acad. Sci. [A] **31,** 450 (1950).
[9] Clemens Schaefer und Heinrich Nassenstein, Z., Naturforsch. [A] **8,** 90 (1953).
[10] H. R. Lillie and H. N. Ritland, J. Am. Cer. Soc. **37,** 466 (1954).
[11] R. W. Goranson and L. H. Adams, J. Franklin Inst. **216,** 475 (1933).
[12] ASTM standards on glass and glass products ASTM designation: C 162–52, page 68 (April 1955).
[13] F. W. Glaze, Optical glass at the National Bureau of Standards, NBS Circ. 469.
[14] M. M. Frocht, Photo-elasticity, vol. II, p. 30 (John Wiley & Sons, Inc., New York, N. Y., 1948).
[15] R. M. Waxler, Glass Ind. **34,** 258 (1953).

Washington, January 31, 1957.

Reprinted from JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, Vol. 51, No. 3, 368–370, March, 1961
Printed in U. S. A.

Rapid Method for Interpolating Refractive Index Measurements

ORESTES N. STAVROUDIS AND LOYD E. SUTTON
National Bureau of Standards, Washington, D. C.
(Received September 1, 1960)

A DESIGNER of optical systems must know the refractive index of the materials he may be required to use. The range of sensitivity of modern photographic emulsions and other image-recording devices as well as the requirements of modern technology demands that this knowledge extend over a wide wavelength range, often far beyond the visible. On the other hand, the index of refraction is usually measured only at discrete points for which strong calibrated line sources either in absorption or emission are available. To bridge the gap between what is available and what is required is clearly a problem of interpolation.

It was decided to apply a two-termed Sellmeier formula[1] to the fitting of experimental data:

$$N=\frac{A_1\lambda^2}{\lambda^2-\lambda_1^2}+\frac{A_2\lambda^2}{\lambda^2-\lambda_2^2}, \tag{1}$$

where $N=n^2-1$. Clearing this expression of fractions one obtains

$$N\lambda^4-\alpha N\lambda^2+\beta N-\gamma\lambda^4+\delta\lambda^2=0, \tag{2}$$

where

$$\begin{aligned}\alpha&=\lambda_1^2+\lambda_2^2\\ \beta&=\lambda_1^2\lambda_2^2\\ \gamma&=A_1+A_2\\ \delta&=A_1\lambda_2^2+A_2\lambda_1^2.\end{aligned} \tag{3}$$

This is now a linear expression in α, β, γ, and δ. Using least-squares techniques, these four quantities can be determined and, from the auxiliary equations (3), λ_1, λ_2, A_1, and A_2 can be found.

The uniqueness of the determination of λ_1, λ_2, A_1, and A_2 from Eq. (3) is easy to verify. From the first two of these, λ_1^2 and λ_2^2 are the two solutions of the quadratic

$$u^2-\alpha u+\beta=0.$$

The remaining two equations of (3) are a simultaneous pair in A_1 and A_2 which in general possess a unique solution.

The occurrence of imaginary values of λ_1 and λ_2 is an embarrassing possibility. However, it is sufficient to say that in the dozen or so applications of this program, imaginaries appeared only when the data contained gross errors of a clerical nature or when too few data were used. Needless to say, the difficulties disappeared when the data were corrected.

A least-squares solution to Eq. (2) is not necessarily a best fit to the Sellmeier equation (1) in the sense of least squares. This may be seen from the fact that (2) is obtained by multiplying Eq. (1) by $(\lambda^2-\lambda_1^2)(\lambda^2-\lambda_2^2)$, which is not a constant. Thus, when Eq. (2) is used, measurements at various values of λ are in effect weighted unequally by this nonconstant multiplier. However, dividing Eq. (2) by λ^2 would change the multiplier to $[1-(\lambda_1^2/\lambda^2)](\lambda^2-\lambda_2^2)$. This latter multiplier was found in practice to be more nearly a constant function of λ than the former. Thus, a better approximate solution to Eq. (1) is obtained by fitting not Eq. (2) but

TABLE I. Parameters after successive iterations.

	Run 1		Run 2		
	Initial solution	Iteration 1	Iteration 2	Iteration 3	Iteration 4
A_1	1.149463	1.149675	1.149650	1.149675	1.149675
A_2	2.392486	3.030655	3.031170	3.238736	3.253283
λ_1	0.0862619 μ	0.0861695 μ	0.0861695 μ	0.0861692 μ	0.0861693 μ
λ_2	37.1848 μ	41.5683 μ	41.5683 μ	42.9010 μ	42.9929 μ
ΣR^2	0.110468 $\times10^{-5}$	0.071797 $\times10^{-5}$	0.071330 $\times10^{-5}$	0.069328 $\times10^{-5}$	0.069315 $\times10^{-5}$

$$N\lambda^2-\alpha N+\beta N\lambda^{-2}-\gamma\lambda^2+\delta=0. \tag{4}$$

Use of the above equation has been found to be a practical expedient toward fitting Eq. (1) for a wide variety of materials. The solutions obtained in each case have been sufficiently accurate to provide an adequate starting point for the linear improvement process described below.

To obtain a more exact fit of the Sellmeier equation, the values of the parameters of the least-squares solution to (4) were used as a set of initial values in a least-squares linear improvement routine. Let $\bar{A}_1$, $\bar{A}_2$, $a_1=\bar{\lambda}_1^2$, $a_2=\bar{\lambda}_2^2$ represent these initial values, and let

$$N_0=(n_0^2-1)=\frac{\bar{A}_1\lambda^2}{\lambda^2-a_1}+\frac{\bar{A}_2\lambda^2}{\lambda^2-a_2}. \tag{5}$$

Let $N_m=(n_m^2-1)$ denote the measured values which are to be fitted in the sense of least squares by proper choice of the four parameters on the right side of Eq. (5). That is, values of ΔA_1, ΔA_2, Δa_1, and Δa_2 are sought so that $(N_0+\Delta N_0)$ will give a better fit to N_M. At this point, we must assume that the initial solution (5) is sufficiently close to the final solution to allow ΔN_0 to be replaced by a linear function of ΔA_1, ΔA_2, Δa_1, and Δa_2 for the purposes of an iterative improvement process. Setting

$$\Delta N_0=\left(\frac{\partial N}{\partial A_1}\right)_0\Delta A_1+\left(\frac{\partial N}{\partial A_2}\right)_0\Delta A_2+\left(\frac{\partial N}{\partial a_1}\right)_0\Delta a_1+\left(\frac{\partial N}{\partial a_2}\right)_0\Delta a_2 \tag{6}$$

in the desired equality $(N_0+\Delta N_0)=N_m$, and algebraically deriving the partial derivatives of N_0 from (1) leads to

$$\left(\frac{\lambda^2}{\lambda^2-a_1}\right)\Delta A_1+\left(\frac{\lambda^2}{\lambda^2-a_2}\right)\Delta A_2+\left(\frac{\bar{A}_1\lambda^2}{(\lambda^2-a_1)^2}\right)\Delta a_1+\left(\frac{\bar{A}_2\lambda^2}{(\lambda^2-a_2)^2}\right)\Delta a_2=N_m-N_0, \tag{7}$$

from which values of the four unknowns ΔA_1, ΔA_2, Δa_1, Δa_2 may be found by least squares. These increments are then applied to the values of the parameters found from (4) and the process may then be repeated.

Measurements of the refractive index of a substance at 46 wavelengths (Table II) were fitted to a two-term Sellmeier equation using this process. Because these measurements are only preliminary results, the substance will not be identified. More importantly, these data are not intended to provide information about a specified substance but are used solely to furnish a numerical example.

Two runs were made using an IBM 704 computer. In the first, solutions to (4) were found to which the improvement routine was applied once. The results of this run verified that the quality of fit had been improved by this first iteration. The machine program had been coded to allow the option of starting the computation with initial values of λ_1^2 and λ_2^2 specified in the input data. Since further iterations appeared necessary, another run was made, which used this option. For this second run, the final values of λ_1^2 and λ_2^2 from the first run were read in, along with a number specifying two iterations of the improvement routine. Thus, for the second run the machine first computed values of the parameters A_1 and A_2 directly by least squares, and the resulting solution was then adjusted by two applications of the improvement routine to all four parameters. Figure 1 gives the residuals (calculated minus measured values) at each of the 46 wavelengths for the initial solution and the results of the final iteration. Table II gives the values of the parameters and the sum of the squares of the residuals obtained at each iteration. Computing time, including readin and the reading out of each iteration, was 3 min for run 1 and 4 min for run 2.

The process described herein provides a rapid and economical means of obtaining an interpolation formula for refractive index versus wavelength of moderate accuracy. Although it falls short of what is desirable in such a formula by as much as an order of magnitude, it does provide a starting point for obtaining an interpolation formula of greater complexity, such as has been already developed.[2]

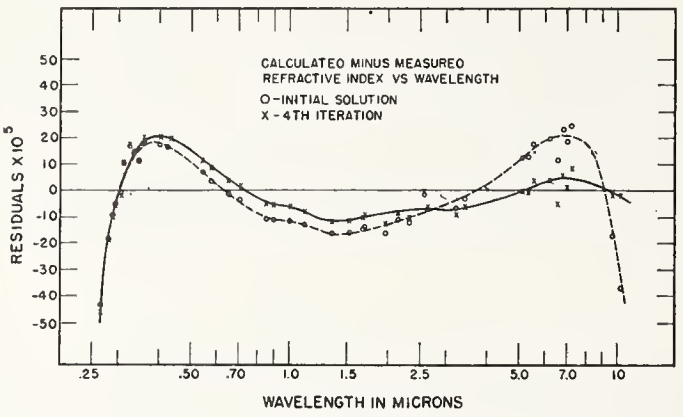


FIG. 1. Residuals of fit after initial solution and fourth iteration.

TABLE II. Final results—iteration 4.

Wavelength μ	Refractive index Measured	Calculated	Wavelength μ	Refractive index Measured	Calculated
0.26520	1.51217	1.51171	1.12866	1.46779	1.46771
0.28035	1.50668	1.50648	1.36728	1.46673	1.46662
0.28936	1.50390	1.50380	1.52952	1.46613	1.46602
0.29673	1.50186	1.50180	1.681	1.46561	1.46551
0.30215	1.50044	1.50043	1.701	1.46554	1.46545
0.3130	1.49782	1.49793	1.9701	1.46472	1.46459
0.32546	1.49520	1.49538	2.1526	1.46410	1.46402
0.33415	1.49363	1.49378	2.32542	1.46356	1.46346
0.34036	1.49257	1.49272	2.5766	1.46262	1.46261
0.34662	1.49158	1.49171	2.6738	1.46234	1.46227
0.361051	1.48939	1.48959	3.2434	1.46018	1.46009
0.3663	1.48869	1.48889	3.422	1.45940	1.45934
0.404656	1.48438	1.48459	5.138	1.45012	1.45012
0.435834	1.48173	1.48193	5.3034	1.44904	1.44904
0.54607	1.47586	1.47598	5.343	1.44878	1.44877
0.57801	1.47478	1.47486	5.549	1.44732	1.44736
0.589262	1.47443	1.47451	6.238	1.44216	1.44220
0.64385	1.47302	1.47306	6.6331	1.43899	1.43894
0.65628	1.47274	1.47278	6.8559	1.43694	1.43700
0.706519	1.47177	1.47179	7.0442	1.43529	1.43531
0.85212	1.46984	1.46979	7.268	1.43314	1.43322
0.89440	1.46942	1.46937	9.724	1.40514	1.40512
1.01398	1.46847	1.46841	10.346	1.39636	1.39634

There is an obvious generalization of Eq. (2) to include Sellmeier equations containing more than two terms. Attempts have been made to obtain fits using three and four terms, but these have failed. Invariably the program located values of λ_1 and λ_2 dispersed among the values of wavelength used as input data. Since the Sellmeier equation has a pole at each such wavelength, the resulting formula can not be used for interpolation. No satisfactory explanation of this phenomenon has been found.

The authors wish to thank Mr. Irving Malitson of this laboratory for his invaluable assistance in providing data for our numerous trials and for releasing for publication the tables presented herein.

[1] R. W. Ditchburn, Light 1953, 456.
[2] I. Malitson and L. E. Sutton (to be published).

Reprinted from JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, Vol. 51, No. 8, 901–905, August, 1961
Printed in U. S. A.

Fitting Refractive Index Data by Least Squares

LOYD E. SUTTON AND ORESTES N. STAVROUDIS
National Bureau of Standards, Washington, D. C.

(Received March 13, 1961)

The accessibility of high-speed computing machinery makes practicable the use of a routine for the least-squares fitting of a three-term Sellmeier equation to a set of experimentally determined values of index of refraction. The constants of a two-term Sellmeier equation are evaluated by a method described previously [O. N. Stavroudis and L. E. Sutton, J. Opt. Soc. Am. **51**, 368 (1961)]. These are then used in a preliminary fitting of another term. The rough fit is then improved by an iterative process which includes an acceleration technique to speed convergence to the final result. In a typical example the average residual of index is only about 2×10^{-5} for 46 wavelengths from 0.2652μ to 10.346μ.

INTRODUCTION

THE index of refraction of optical materials is usually measured only at discrete wavelength values for which strong calibrated line sources either in absorption or emission are available. The problem of interpolating for intermediate wavelengths would be solved ideally by obtaining a formula exhibiting *completely* the physical relationship of refractive index with wavelength. On the other hand, this laudable goal, because of the inherent complexity of this relationship, is at present unattained, at least for routine application.

In this paper, the Sellmeier formula [Eq. (2)] is used as a compromise model for the ideal equation. This formula is based on theoretical considerations which facilitate a more compact expression of the fitting equation than would generally be expected from a completely arbitrary formulation. Also, a purely mathematical fitting of the Sellmeier formula often provides at least a rough guide to the values of certain other physical parameters of the optical material, in addition to the measured refractive indices. The degree of accuracy possible in such conclusions involving further physical interpretation of the results is not presently known in a general sense, however. Primary emphasis is given here to obtaining a mathematical fit of measured data useful for interpolation.

A detailed background for the theoretical considerations leading to the Sellmeier formula may be found in such standard texts as Ditchburn's.[1] Following his terminology, one obtains for the real part of the classical Helmholtz formula for the complex index of refraction

$$n^2(1-K^2)=1+\sum_i \frac{A_j\lambda^2}{\lambda^2-\lambda_j^2+g_j\lambda^2/(\lambda^2-\lambda_j^2)}, \qquad (1)$$

where n is the real part of the complex refractive index and K is the extinction coefficient. Wavelengths of maximum absorption are denoted by λ_j. The g_j are related to the damping coefficients of the dipole oscillators comprising the medium, and the A_j are constants related to the strengths of the absorption bands. Over a range in which λ differs considerably from all of the λ_j, one may, by assuming K and the g_j to be negligibly small within that range, pass to the Sellmeier equation

$$n^2=1+\sum_j A_j\lambda^2/(\lambda^2-\lambda_j^2). \qquad (2)$$

METHOD

In our previous paper,[2] a method was described for fitting a two-term Sellmeier formula to data consisting of values of refractive index measured at given wavelengths. The fitting equation was of the form

$$n^2-1=\frac{A_1\lambda^2}{\lambda^2-\lambda_1^2}+\frac{A_2\lambda^2}{\lambda^2-\lambda_2^2}, \qquad (3)$$

where n=refractive index, λ=wavelength (in microns), and A_1, A_2, λ_1^2, and λ_2^2 were unknown constants to be determined mathematically so as to minimize the sum of squared residuals. A process was given for obtaining sufficiently accurate initial values of the four constants so that a few iterations of a linear improvement routine would converge to a set of values of the constants with the desired minimal property. The entire calculation was programmed for a high-speed digital computer.

If the wavelength range of the measurements is over only the visible region or perhaps somewhat wider, it has been established, in the years since Cauchy's original investigations in 1836, that the index curve for many materials can be represented by Eq. (3) or by its cruder approximations which use fewer than four parameters, with a fit adequate to the accuracy of the observations. On the other hand, there are many cases today in which a much broader range of wavelengths is covered and four parameters are found to be too few to provide a reasonable fit. This paper is a continuation of our previous paper[2] and deals with this latter situation. A method is given for using the solution to Eq. (3) as a means for fitting to measured indices the three-term Sellmeier equation

$$n^2-1=\frac{B_1\lambda^2}{\lambda^2-l_1^2}+\frac{B_2\lambda^2}{\lambda^2-l_2^2}+\frac{B_3\lambda^2}{\lambda^2-l_3^2}, \qquad (4)$$

[1] R. W. Ditchburn, Light **1953**, 456.

[2] O. N. Stavroudis and L. E. Sutton, J. Opt. Soc. Am. **51**, 368 (1961).

where the B's and l^2's represent a set of six constants to be determined by least squares.

In a solution of Eq. (3), the values obtained for λ_1 and λ_2 are found, in ordinary practice, to lie on opposite sides of the range of measured wavelengths.

By convention, λ_1 is used to denote the singularity of Eq. (3) on the low wavelength side of the measured range (usually called the "ultraviolet"), while λ_2 denotes the singularity on the high side (the "infrared"). However, for Eq. (4), two configurations might normally arise:

Case A (l_1 and l_2—in the "ultraviolet")
(l_3 —in the "infrared")

Case B (l_1 —in the "ultraviolet")
(l_2 and l_3—in the "infrared").

The following method applies to case A and, with a minor modification, to case B also. Our machine program was coded to be useable for either case. However, the method will be described in more detail for case A since this configuration has been indicated for all nine optical materials for which we have so far used the code.

The process starts from a knowledge of the values of the four constants in Eq. (3). We first rewrite Eq. (3) rather arbitrarily to obtain an equivalent case A equation of three terms, as follows

$$n^2-1=\frac{(\frac{1}{2}A_1)\lambda^2}{\lambda^2-\lambda_1^2}+\frac{(\frac{1}{2}A_1)\lambda^2}{\lambda^2-\lambda_1^2}+\frac{A_2\lambda^2}{\lambda^2-\lambda_2^2}. \tag{5}$$

Comparison of the constants of Eq. (5) with the corresponding constants of Eq. (4) suggests a way of obtaining an initial fit for Eq. (4) which must be at least as accurate as Eq. (5). We can assign numerical values to four of the constants in Eq. (4) as follows:

$$\begin{aligned} &B_{1(0)}=\tfrac{1}{2}A_1 \\ &B_{2(0)}=\tfrac{1}{2}A_1 \\ &B_{3(0)}=A_2 \quad l_{3(0)}^2=\lambda_2^2, \end{aligned} \tag{6}$$

where the second subscripts denote initial values for the corresponding constants. We can then solve for initial values of the remaining constants of Eq. (4), denoted by $l_{1(0)}^2$ and $l_{2(0)}^2$.

To carry out the solution outlined above, we write Eq. (4) as

$$N=n^2-1=h\left[\frac{\lambda^2}{\lambda^2-l_{1(0)}^2}+\frac{\lambda^2}{\lambda^2-l_{2(0)}^2}\right]+T_2, \tag{7}$$

where $h=\frac{1}{2}A_1$ and $T_2=A_2\lambda^2/(\lambda^2-\lambda_2^2)$. This is equivalent to

$$M=\frac{\lambda^2}{\lambda^2-l_{1(0)}^2}+\frac{\lambda^2}{\lambda^2-l_{2(0)}^2}, \tag{8}$$

where $M=(N-T_2)/h$.

For case B, our machine program was so coded as to apply formally the same method of case A to the equation obtained by writing the two terms on the right side of Eq. (3) in reverse order. The two cases are clearly symmetrical, provided allowance is made for one exception to the symmetry. The multiplier used in proceeding from Eq. (8) to Eq. (9) should here be considered roughly a constant, because (for case B) the constants now represented by $l_{1(0)}^2$ and $l_{2(0)}^2$ are assumed to be large relative to the values of λ^2. For this case, we thus omit the division by λ^4 and use Eq. (9), rather than Eq. (11), to solve for values of S and P.

It may be noted at this point that, for either of the two cases, the solution of Eq. (10) involves solving a quadratic equation, which theoretically might produce an imaginary value for $l_{1(0)}$ or $l_{2(0)}$. Although such a result has not yet arisen in practice, a special exit to signal such an occurrence was included in the code. Solutions for $l_{1(0)}^2$ and $l_{2(0)}^2$ have always been somewhat lower and higher, respectively, than the λ_1^2 of Eq. (3), and the resulting quality of the initial fit of Eq. (4) has always been decidedly better than that of Eq. (3), as one would, of course, expect from the use of more parameters.

Values of M are easily computed for each measured wavelength. Thus, the fitting of Eq. (8) reduces to a special case of the two-term method described in our previous paper.[2] Multiplying Eq. (8) algebraically by $[\lambda^2-l_{1(0)}^2][\lambda^2-l_{2(0)}^2]$ yields

$$[(M-1)\lambda^2]S+[-M]P=[M-2]\lambda^4, \tag{9}$$

where

$$\begin{aligned} S&=l_{1(0)}^2+l_{2(0)}^2 \\ P&=l_{1(0)}^2\cdot l_{2(0)}^2. \end{aligned} \tag{10}$$

We now need to compensate for the weighting effect caused by the nonconstant multiplier introduced in the step above. Since, for case A, this multiplier is roughly equal to λ^4, we will divide Eq. (9) by λ^4 before the next step. Experience has shown that the accuracy maintained through this step is fully adequate for obtaining an initial six-parameter fit. A solution for S and P is thus obtained through a linear least-squares fit using the following equation:

$$\left[\frac{M-1}{\lambda^2}\right]S+\left[\frac{-M}{\lambda^4}\right]P=[M-2]. \tag{11}$$

The resulting values of S and P are substituted into the auxiliary Eqs. (10) to obtain $l_{1(0)}^2$ and $l_{2(0)}^2$. These two values, together with the four given by Eq. (6), provide our initial solution for the constants of Eq. (4), under the assumption of a case A configuration.

Improvement of Initial Solution

Having obtained initial values for the six constants appearing in Eq. (4), we may now use the values of the $l_{(0)}^2$'s to obtain improved values for the $B_{(0)}$'s by the usual least-squares fitting. This step increases the accuracy of our starting point, thus speeding up the rate

of convergence for the remaining calculations. The machine program also allows one the option of inserting specific estimates of the $l_{(0)}{}^2$'s and starting with this last step, if desired.

Now the code begins applying a linear improvement process to all six constants of the starting-point solution as obtained above. This iterative process is basically the same as that given in detail for four parameters in our previous paper.[2] We may condense the equations here somewhat by letting p_i ($i=1, 2, \cdots 6$) represent one of the six initial parameters, and by letting (Δp_i) represent the linear estimate of its adjustment, for which we must solve. If there are k measurements, we will have k linear equations in the form

$$n_m-(n_{c0})=\sum_{i=1}^{i=6}\left(\frac{\partial n_c}{\partial p_i}\right)_0(\Delta p_i), \tag{12}$$

where n_m=measured index, n_c=calculated index, and the zero subscripts indicate quantities to be evaluated from the starting-point solution. Solving these k Eqs. (12) by least squares yields six values of (Δp_i). The first adjusted solution $(p_i)_1$ is thus obtained from

$$(p_i)_1=(p_i)+(\Delta p_i), \tag{13}$$

where $i=1, 2, \cdots 6$. This adjusted solution is taken as a new starting point and the improvement process may be iterated a specified number of times.

This linear improvement process produced rapid convergence to an optimum fit during numerous applications of the four-parameter code described in our previous paper.[2] Usually, about 4 or 5 iterations were sufficient. In this paper, we are seeking a considerably closer fit by use of two additional parameters, and it was found that Eq. (13) tended to "overshoot," usually during the first few iterations. That is, the magnitude of the adjustment step (Δp_i) was found to be too large because of the nonlinearity of n_c as a function of the parameters.

This difficulty was originally overcome by using the equation

$$(p_i)_1=(p_i)+c\cdot(\Delta p_i), \tag{14}$$

where $i=1, 2, \cdots 6$ and $0<c<1.0$ during the first few cycles of iterations. Thus, for example, one could specify 4 iterations of Eq. (14) with $c=0.3$, to be followed by 5 iterations of Eq. (13). The code includes provisions for automatic termination of a run if a widely divergent iteration occurs, and also for starting the iterations of a new machine run from the best result of a previous run, when further improvement is indicated. The sum of squared residuals and current values of the six parameters are given in a short printout after each iteration. This code, as so far described, proved adequate for obtaining optimum fits fairly easily in most cases. In some instances, however, several machine runs were necessary, and it appeared desirable to seek to increase the accuracy of the method of iteration, while still solving only linear equations.

An improvement was thus made in the accuracy of a single iteration step by use of a scheme which may conveniently be called the "midpoint method." The method involves a modified use of Eq. (13) or Eq. (14), the choice remaining fixed for this description. Just before use of Eq. (13) or Eq. (14), the machine treats the solution to Eq. (12), (Δp_i), as tentative values to be replaced by better values, $(\Delta' p_i)$. To obtain these $(\Delta' p_i)$, a midpoint is first computed, which is defined as

$$(p_i)_{\text{mid}}=(p_i)+\frac{\alpha}{2}(\Delta p_i), \tag{15}$$

where $\alpha=1.0$ for Eq. (13), or $\alpha=c$ for Eq. (14) and $i=1, 2, \cdots 6$.

The machine next obtains new partial derivatives, $(\partial n_c/\partial p_i)_{\text{mid}}$, where the subscript indicates that these partials are to be evaluated at the midpoint just defined. In Eq. (12), the quantities $(\partial n_c/\partial p_i)_0$ are at this stage replaced by $(\partial n_c/\partial p_i)_{\text{mid}}$, while the left side remains unchanged. The values obtained by solving this modified form of Eq. (12) constitute our better set of values, $(\Delta' p_i)$, which are then used in Eqs. (13) or (14) instead of the tentative values, (Δp_i). This completes a single step of the improved iteration process.

This "midpoint method" was found in practice to give a definite acceleration in convergence of the iterations, particularly during the last few iterations, near the final solution. To see why this method was used, we should first realize that the ordinary use of Eqs. (12) and (13) is a generalization of Newton's method for solving the equation, $f(x)=0$, given an approximate solution, x_0. For that equation, Newton's method consists of iterations of the formula

$$x_1=x_0-f(x_0)/f'(x_0), \tag{16}$$

where x_1 is the next approximation.

This may be written as

$$-f(x_0)=f'(x_0)(\Delta x_0), \tag{17}$$

where $(\Delta x_0)=(x_1-x_0)$ will be solved for, and

$$x_1=x_0+(\Delta x_0). \tag{18}$$

In using Eq. (12), we are seeking a (least squares) solution for the function $f(p_i)\equiv(n_c-n_m)=0$. Clearly, Eq. (12) is a generalization of Eq. (17), and Eq. (13) is the corresponding generalization of Eq. (18). The application of least squares in this manner is originally due to Gauss.

We have found that the accuracy of Eq. (16) may be increased through the following considerations, which apply when $f(x)$ is a quadratic function, ax^2+bx+c. The following identity then holds for two arbitrary values of x, denoted by x_0 and x_1

$$\frac{f(x_1)-f(x_0)}{x_1-x_0}\equiv f'\left(\frac{x_0+x_1}{2}\right). \tag{19}$$

518-743 O - 73 - 5

It would, of course, be difficult to compute directly the left side of Eq. (19) for small values of (x_1-x_0), i.e., (Δx_0). Taking x_1 to be determined from x_0 by using Eq. (16), we see from Eq. (19) that the slope at the midpoint of our x interval is the same as the slope of the chord joining the two ordinates, $f(x_0)$ and $f(x_1)$. If we now reapply Eq. (16), this time using $f'[(x_0+x_1)/2]$ in the place of $f'(x_0)$, we will be performing a linear extrapolation along that chord to locate its intersection with the x axis. This second application of Eq. (16) may be written as

$$-f(x_0)=f'[(x_0+x_1)/2](\Delta' x_0), \tag{20}$$

where $(\Delta' x_0)=(x_1'-x_0)$ will be solved for

$$x_1'=x_0+(\Delta' x_0), \tag{21}$$

and x_1' represents a better solution than the x_1 of Eq. (18).

The preceding geometric interpretation of the "midpoint method" is exact only when $f(x)$ is a quadratic function, as stated. For more general functions, we may examine a Taylor's series expansion with the midpoint as origin

$$f(P\pm h)=f(P)\pm hf'(P)+(1/2)h^2f''(P)\pm\cdots, \tag{22}$$

where $P=(x_0+x_1)/2$ is the midpoint and

$$h=(1/2)(x_1-x_0).$$

Application of Eq. (22) to the left side of Eq. (19) shows that the slope at the midpoint still equals the actual chord slope, if we consider only the first three terms in the above expansion of the function. Thus, in the region near the final solution, where the value of h is quite small, the improvement of the "midpoint method" over Newton's method is mathematically well defined and quantitatively important. Elsewhere, the improvement is, in general, only moderate, but still helpful.

The "midpoint method," as applied toward the solution of $f(x)=0$, may be generalized for application to equations with several independent variables.[3] For our problem, this generalization takes the form of Eq. (15), with $\alpha=1$, and its use in conjunction with Eqs. (12) and (13), as already described. Here, Eqs. (20) and (21) provide the model for the generalization, and a full step of correction is performed. When a shorter (partial) step of correction is specified, the code in that case will obtain the midpoint by setting $\alpha=c$ in Eq. (15) and use the resulting secondary solution of Eq. (12) in Eq. (14) rather than Eq. (13). The value of c and the number of iteration steps of both types are specified by the user of the code as part of the input data.

EXAMPLE

In Table I, we give a summary of results obtained from application of the three-term code to the same substance used as an example in our previous paper.[2] (The substance will remain unidentified until separate publication of the results in detail.) To obtain the three-term fit, the values of A_1, A_2, and λ_2^2 from Table I and a number specifying a case A fit were included in the input data. Five short steps ($c=0.16$), followed by four full steps of correction, were also specified. For the initial three-term fit the value of Q (see Table I) was 15.214×10^{-8}. After the eighth iteration step, this value was down to 3.466×10^{-8}, and after the ninth and final step reached 3.463×10^{-8}. This decrease was monotonic except for the sixth step, where a minor oscillation occurred. The "midpoint method," introduced into the code later, tends to remove such effects. In any case, the very small adjustments in the parameters, caused by the final step, showed that a minimum Q had been reached, to the accuracy of the measured data.

TABLE I. Parameters of equations; fitted to 46 measurements.

Best 2-term fit by Eq. (3)	Best 3-term fit by Eq. (4)
	$B_1=$ 0.643356
$A_1=$ 1.149675	
	$B_2=$ 0.506762
$A_2=$ 3.253283	$B_3=$ 3.826131
	$l_1=$ 0.0577888μ
$\lambda_1=$ 0.0861693μ	
	$l_2=$ 0.1096757μ
$\lambda_2=$42.9929μ	$l_3=$46.3864μ
$Q=69.315\times10^{-8}$	$Q=$ 3.463$\times10^{-8}$

[a] The quantity "Q" is the sum of squares of the residuals.

The solid line in Fig. 1 represents the general magnitude of the residuals of the two-term equation. The forty-six residuals of the three-term equation are shown as individual points, a few of these overlapping one another. The average of absolute values of the residuals, for the two-term fit is 9.3×10^{-5}, and, for the three-term fit, is 1.9×10^{-5}. Improvement at small wavelengths is clearly greater than would be reflected by such averages.

Machine time, for such a three-term calculation as the above, averages 2 or 3 minutes, or about twice that

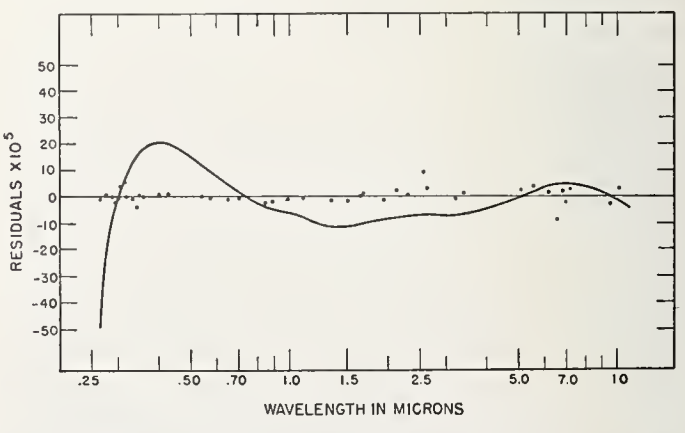


FIG. 1. Residuals (calculated minus measured refractive indexes). —Two-term equation. ·Three-term equation.

[3] M. B. Wilk, Ann. Math. Stat. 29, 618A (1958).

if printing of the results is included. The printing may be done "off-line," if desired.

Purely as a test problem, the present three-term code was applied to data for another substance, which had long before been fitted accurately by a five-term Sellmeier formula. In the five-term fit, the parameters λ_1 through λ_4 had been given fixed values, equally spaced across a band of short-wavelength absorption located by physical measurements. Only the remaining constants, including a value for λ_5 in the infrared, had been determined by least-squares adjustment. The three-term formula, as obtained by the present code, was found to fit the data to the same accuracy as the five-term formula. It was interesting to note that, in this three-term fit, l_1 turned out to be just 4% above the average of λ_1 and λ_2, l_2 was 1% below the average of λ_3 and λ_4, and l_3 was 1% above the value of λ_5. Such agreement is encouraging, and speaks well for both formulas. The economy of the three-term fit as an interpolation formula was surprisingly good. However, we certainly cannot conclude, from any single example of this type, that the results of the present code will necessarily allow physical interpretations with a specific degree of accuracy.

CONCLUSION

Our experience with this code indicates that it will be a labor-saving and powerful tool. The interpolation formulas obtained have consistently fitted the data to an accuracy adequate to that of the measurements, at least over the quite wide wavelength ranges of the nine materials so far considered. No strict physical interpretation of the parameters obtained can presently be claimed, in a general sense. If reliable, fixed values of the three denominator parameters were known from physical measurements, the code could be used in a single step to compute corresponding numerator constants by linear least squares. Such an ideal situation has not arisen in practice, however, and we have thus determined all six parameters mathematically in each case so far.

ACKNOWLEDGMENT

The authors wish to thank Irving Malitson of this laboratory for providing the refractive index measurements used in our numerical examples.

JOURNAL OF RESEARCH of the National Bureau of Standards—A. Physics and Chemistry
Vol. 67A, No. 2, March–April 1963

Effect of Pressure and Temperature on the Refractive Indices of Benzene, Carbon Tetrachloride, and Water

R. M. Waxler and C. E. Weir

(October 31, 1962)

An interferometer for measuring change in index of refraction with pressure is described. Absolute indices of refraction are reported to five decimals for benzene, carbon tetrachloride, and water at pressures as high as 1100 bars over a small temperature range. The results of replicate measurements agree to within ±0.0001. Various equations relating index and specific volume show systematic deviations in all cases. At constant specific volume, the index of carbon tetrachloride increases with increasing temperature, while the index of water decreases with increasing temperature. The refractive index of benzene shows no effect due solely to temperature within the experimental error. Possible explanations for this behavior are discussed.

1. Introduction

Recently there has been great interest in studying optical properties of materials at high pressure as a method of examining short range molecular interactions, the present emphasis being confined largely to spectroscopic measurements. To a large extent index of refraction measurements have been neglected in these studies, although it would appear that index studies with their inherent precision might be of assistance in interpreting spectroscopic data. Index-density studies on condensed phases have been made previously, but the objectives of the previous experiments were usually to determine the applicability of index-density relationships or the usefulness of index as a measure of volume change. Most of the data published have been given to about the fourth decimal, and although such precision is excellent for many measurements, it is not particularly good for index of refraction. The present work was undertaken in an attempt to increase the precision of index measurements at elevated pressures. For these studies an interferometer for use at high pressures has been developed, and this report contains the results of measurements on three liquids at pressures up to about 1000 bars.

Poulter, Ritchey, and Benz [1] [1] used the minimum deviation method for studying the refractive indices of liquids to 13,600 atm. The windows [2] of their pressure vessel were mounted at an angle of 30° to each other and with the enclosed liquid constituted a 30° prism. Poulter and Benz [3] discussed the errors arising from distortion of the windows. Rosen [4] used a similar method for studying water to 1800 bars. This equipment as described by Lyons and Poindexter [5] utilized windows forming a prism with a 51° angle. These workers estimated the errors arising from distortion to be of the order of 0.0003 in index but concluded that this error was less than other experimental errors. Gibson and Kincaid [6] used the immersion method for studying the index of benzene. In these studies glass particles were immersed in benzene and pressure was used to match the index of the benzene to that of the glass. A sensitivity of 0.00004 is reported for index but the data are reported to the fourth decimal in index. In these studies the index of the glass was known at 1 bar but its value at elevated pressure was calculated.

As early as 1857, Jamin [7] described the use of an interferometer for measuring change in refractive index of fluids under pressure. Most subsequent work has been done on gases, the most notable being that of Michels and Hamers [8]. The windows limit the pressure range of interference methods. The present report describes an interferometer which is completely enclosed in the pressure vessel so that distortion of the windows at elevated pressure has a negligible effect on the interference pattern. In the present experiments, change in refractive index was measured with a sensitivity of 0.00001 and estimated limits of reproducibility, expressed as twice the standard deviations about the mean, of the order of ±0.0001. With modifications it is believed that both the precision and the pressure range can be extended by at least one order of magnitude. It is also apparent that the principle can be used to obtain precise data on PVT behavior of solids.

2. Apparatus and Experimental Method

2.1. Interferometer Construction

The interferometer, figure 1, was constructed using a stainless steel bar, A, which was about 130 mm long and 13 mm square. The optical portion consisted of two optical flats separated by a steel spacer.

[1] Figures in brackets indicate the literature references at the end of this paper.

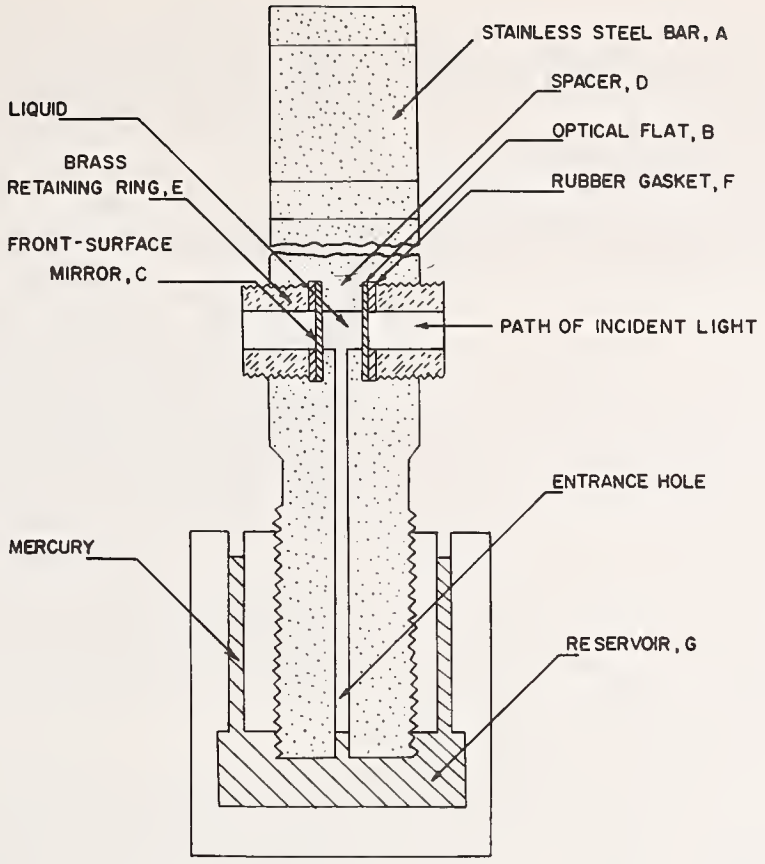


FIGURE 1. *Cross section of the interferometer.*

Recesses 11 mm in diameter to accommodate the flats were made in opposite sides near one end of the bar. The recesses were of such depth as to leave a web, D, of steel about 3 mm thick in the center of the bar. The steel web served as a seat and spacer for the optical flats and was machined so that its surfaces were parallel. The flats were held firmly against the web by threaded brass rings, E, screwed into the upper threaded portion of the recesses. A small neoprene rubber gasket, F, under each ring was used to seal the cell against leakage and to prevent cracking of the flat. It was necessary to silver both flats to produce bright fringes because they were in contact with liquids having indices of refraction not appreciably different from that of the glass. Thus the front flat, B, was half-silvered on its rear surface and the back plate, C, was a front surface mirror, the light being incident from the right as shown in figure 1. The small glass flats 11 mm in diameter were cut from larger optical flats obtained commercially.

A small hole 3 mm in diameter drilled through the web served as the interferometer chamber. To permit introduction of liquid into the chamber, a small axial hole about 1.5 mm in diameter connected the chamber with the lower end of the bar. This end of the bar was threaded and screwed into a stainless steel reservoir, G, containing mercury to separate the liquid under study from the oil used as a pressure-transmitting medium. A small hole was drilled into the reservoir to permit transmission of pressure. The upper end of the bar was slotted to facilitate alinement of the interferometer inside the pressure vessel before applying pressure.

2.2. Pressure Vessel

The pressure vessel was mounted on a large optical bench with the associated optical equipment. This vessel was the one used by Gibson [9] in his classic experiments on liquids. The heavy-walled steel container was fitted with two heavy plate glass windows. The windows were supported by hardened optically flat steel plugs of a design proposed by Poulter [2]. The windows provided an aperture of 8 mm which was adequate for viewing the fringes in the 3 mm interferometer chamber placed opposite the windows inside the vessel. The pressure vessel was thermostatted by means of a temperature-controlled oil bath capable of control to within ± 0.02 °C. To obtain temperatures below room temperature, the bath was cooled with the immersion coil of a portable refrigerator.

In studying each liquid, the oil bath was brought to constant temperatures of 25, 35, and 55 °C. In the case of water, additional experiments were conducted where the bath was brought to temperatures of 2 and 8 °C. By probing with a calibrated thermocouple it was discovered that temperature gradients existed within the pressure vessel itself. A cali-

brated thermocouple was fastened to the interferometer at a position very close to the liquid under study to determine the actual temperatures at which the experiments were conducted.

Pressures were measured to ± 0.3 bar with a thermostatted manganin pressure gage mounted in the oil line external to the pressure vessel. The pressure gage was calibrated at the freezing point of benzene as reported by Gibson [10].

2.3. Optics

The optical system was similar to the one described by Saunders [11] and the measurements were made photographically. The light source was a helium discharge tube used in conjunction with a constant deviation prism to permit isolation of individual helium lines for illumination. In these studies the 5875.62 A line was used because it was found to produce fringes best suited for photographing. The optical design permitted visual observation of the interference fringes while photographing, and incorporated a right-angle prism so that the observer was out of direct line of the windows in the pressure vessel. This right-angle prism was turned 90° from its normal position [11] so that all the elements of the optical system were in a horizontal plane. In these experiments one window of the pressure vessel and the oil used to transmit pressure were in the path. They were not expected to affect the fringe pattern even when the window became distorted at elevated pressures. To check this prediction, a thin microscope cover glass (0.015 mm thick) was placed in the vessel. The surfaces of the cover glass were sufficiently flat and parallel to produce internal interference, while the glass was so thin that a negligible change in thickness would be produced by a pressure change of 1000 bars. No change in the fringe system was detected at pressures between 1 bar and 1000 bars and it was concluded that distortion of the windows would have no appreciable effect on the measurements.

3. Experimental Technique

A typical experiment was performed as follows: The interferometer was assembled and adjusted to produce a fringe system. The instrument was then filled with the liquid under study. In order to remove air and to ensure that air bubbles were not formed at elevated temperatures it was found necessary to fill the interferometer under vacuum with freshly boiled liquid. The fringe system was rechecked and the interferometer placed inside the pressure vessel and alined. After waiting for about 20 min for attainment of thermal equilibrium, the pressure was raised to some value between 1 bar and 1000 bars and an interval of about 20 min permitted for dissipation of the heat generated by compression. The pressure was then decreased slowly to 1 bar and the fringes photographed as they passed a reference mark on one of the flats. Another 20 min interval was allowed for restoration of thermal equilibrium at 1 bar. Each measurement, therefore, represented an independent experiment between an elevated pressure and 1 bar. Pressure intervals were chosen to give essentially four measurements between 1 bar and 1000 bars.

4. Materials

Measurements were made on benzene, carbon tetrachloride, and water at approximately 25, 35, and 55 °C. Additional measurements were made on water at approximately 2 and 8 °C. The water was freshly distilled while the organic liquids were commercially available materials of high purity used with no further purification. In all cases the purity was corroborated by determination of the index of refraction as described later.

5. Index Calculation

The change in index under pressure was calculated from the equation

$$\Delta\mu = \Delta n \lambda / 2t \sin\theta, \tag{1}$$

where $\Delta\mu$ is the change in refractive index, Δn—the number of fringes passing the reference mark, t—the spacing between the optical flats and θ—the angle of incidence of the light. In these experiments $\theta = 90°$, $\lambda = 5875.62$A, and $t = 0.2664$ cm; so that

$$\Delta\mu = 1.1028 \times 10^{-4} \Delta n. \tag{2}$$

Therefore, a shift of one interference fringe represents a change of approximately 1 in the fourth decimal place of refractive index. Estimates of fractional fringes were made to one-tenth of a fringe so that the measurements were made to approximately 1×10^{-5} in the index. With this sensitivity it was necessary to correct for the compressibility of the steel spacer. The volume compressibility of the steel was taken to be 0.59×10^{-6}/bar [12] with the linear compressibility being one-third of this value. At 1000 bars the correction for the compression of the steel was approximately 3 fringes.

It was feared that unequal differential compressions might cause tilting of the interferometer plates and introduce an error in the fringe count. The error from this source was largely eliminated by counting the number of fringes in the field and noting the fringe orientation before and after each run. In some experiments changes were noted in number or orientation of fringes and these data were discarded. Only those experiments showing no changes in the fringe system were considered acceptable.

6. Initial Index Measurement

In order to ascertain the purity of the materials, the indices were first measured to the fourth decimal place with an Abbé refractometer using sodium light. Refractive indices for the 5875.62 A He line were calculated from these results and published dispersion data on the liquids [13, 14, 15].

To use the data of the present investigation for change in index it was necessary to know the initial indices to five decimals. For benzene and water, refractive indices to the fifth decimal were taken from published data on dispersion over a range of temperatures [13, 14]. The initial refractive index of carbon tetrachloride was determined by the method of minimum deviation, the liquid being enclosed in a hollow glass prism [14].

For benzene and carbon tetrachloride the initial refractive indices were only known at temperatures very close to 25 °C. The refractive index at other temperatures of interest was determined in a manner completely analogous to the pressure measurements. In these experiments it was necessary to correct for the expansivity of the steel spacer. This expansivity was determined by experiment using air in the interferometer to be 14.9×10^{-6}/°C.

It may be noted that all changes in refractive index as determined with the interferometer were absolute values. The initial measurements of refractive index referred to air but were converted to absolute values from a knowledge of the refractive index of air at various temperatures [16].

7. Results and Discussion

The experimental results are given in tables 1 to 4. Tables 1 to 3 include specific volumes in addition to the indices of refraction, while table 4 shows only the index data. Specific volumes given in columns 3 of tables 1 to 3 were computed from values of the constants of the Tait equation given by Gibson and his coworkers [6, 17] who studied these liquids over the same pressure and temperature range. Data of equal precision do not appear to be available for water at 1.56 and 7.64 °C. The data of several workers [18] who studied water at the lower temperatures were compared with the results of Gibson at higher temperatures where there was overlapping of experimental conditions. Although most of the data agreed reasonably well it was considered that the agreement was not sufficiently close to permit calculations involving specific volumes from different sources to be compared. Consequently only index data are reported for water at 1.56 and 7.64 °C and no further discussion of these data are contemplated at this time.

TABLE 1. *Effect of temperature and pressure upon the refractive index and specific volume of benzene*

Pressure	Absolute refractive index	Specific volume
24.80 °C		
Bars		*ml/g*
1.0	1.49859	1.14433
250.1	1.51031	1.11973
489.0	1.51982	1.10057
666.1	1.52593	1.08828
34.50 °C		
1.0	1.49221	1.15805
246.2	1.50438	1.13187
484.8	1.51445	1.11138
757.2	1.52418	1.09192
1107.7	1.53489	1.07112
54.34 °C		
1.0	1.47910	1.18757
274.5	1.49422	1.15350
484.4	1.50382	1.13305
770.2	1.51478	1.11014
1124.5	1.52633	1.08705

TABLE 2. *Effect of temperature and pressure upon the refractive index and specific volume of carbon tetrachloride*

Pressure	Absolute refractive index	Specific volume
24.80 °C		
Bars		*ml/g*
1.0	1.45791	0.63093
276.9	1.47089	.61486
523.6	1.48037	.60347
797.4	1.48934	.59299
1116.7	1.49847	.58276
34.50 °C		
1.0	1.45197	0.63855
255.4	1.46481	.62234
480.9	1.47426	.61094
756.1	1.48381	.59948
1119.9	1.49441	.58710
54.34 °C		
1.0	1.44067	0.65496
248.7	1.45493	.63623
452.7	1.46428	.62423
751.9	1.47616	.61006
1059.2	1.48617	.59830

TABLE 3. *Effect of temperature and pressure upon the refractive index and specific volume of water*

Pressure	Absolute refractive index	Specific volume
24.80 °C		
Bars		*ml/g*
1.0	1.33293	1.00288
259.6	1.33652	.99152
463.6	1.33942	.98318
762.8	1.34322	.97180
1108.6	1.34747	.95973
34.50 °C		
1.0	1.33177	1.00580
259.3	1.33531	.99463
480.9	1.33832	.98572
799.2	1.34228	.97385
1110.0	1.34588	.96318
54.34 °C		
1.0	1.32866	1.01415
241.6	1.33192	1.00372
489.7	1.33517	.99371
785.1	1.33883	.98267
1127.7	1.34274	.97089

TABLE 4. *Effect of temperature and pressure upon the refractive index of water at two lower temperatures*

Pressure	Absolute refractive index
1.56 °C	
Bars	
1.0	1.33439
269.5	1.33859
507.4	1.34213
768.5	1.34581
1049.7	1.34948
7.64 °C	
1.0	1.33423
256.1	1.33805
497.8	1.34161
730.7	1.34495
1088.9	1.34929

7.1. Internal Consistency of the Data

Several replicate measurements were made where the interferometer was completely disassembled and refilled with liquid before each measurement. In each experiment the system was brought to the same pressure as nearly as possible, and then small interpolations were made in order to compare the data directly. On the basis of two experiments of five and four determinations each, the standard deviation was computed to be 0.00005 in refractive index, corresponding to limits of reproducibility of measurements of ± 0.0001, or twice the standard deviation about the mean. Inspection of large scale graphs of index versus pressure, figures 2 and 3, shows that the points lie on a smooth curve with random deviations in the fifth decimal. As a matter of interest, it may be noted that the pressure measurement although assumed exact in this discussion may be subject to a greater relative error than the measurement of index. Of the present data those for water and benzene appear to be slightly more consistent than those for carbon tetrachloride.

7.2. Comparison With Other Data

For purposes of comparing the present data there appear to be only two previously published sets of data, those of Gibson and Kincaid on benzene [6] and those of Rosen on water [4]. Gibson and Kincaid covered essentially the same pressure and temperature range as the present experiments, but Rosen restricted his measurements to 25 °C. Gibson and Kincaid used the immersion method of determining index while Rosen employed the minimum deviation method. Both sets of data are reported to four decimals. Both authors made measurements for more than one wavelength, all differing from the wavelength used here. These data may best be compared with the data of the present investigation by plotting change in refractive index versus pressure. If allowance is made for dispersion, temperature, and vacuum corrections, it is found that the present data agree with the data of Gibson and Kincaid [6] and

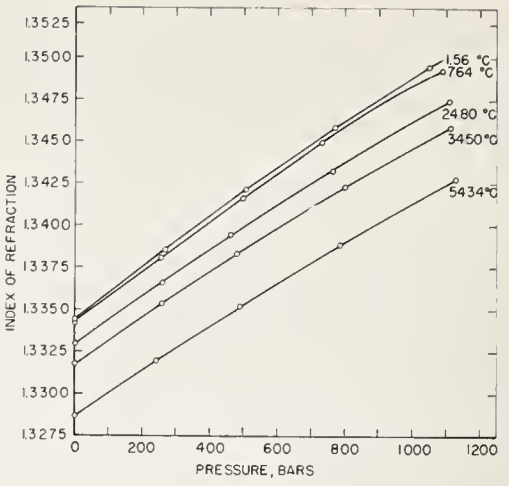


FIGURE 2. *Pressure dependence of index of refraction of water.*

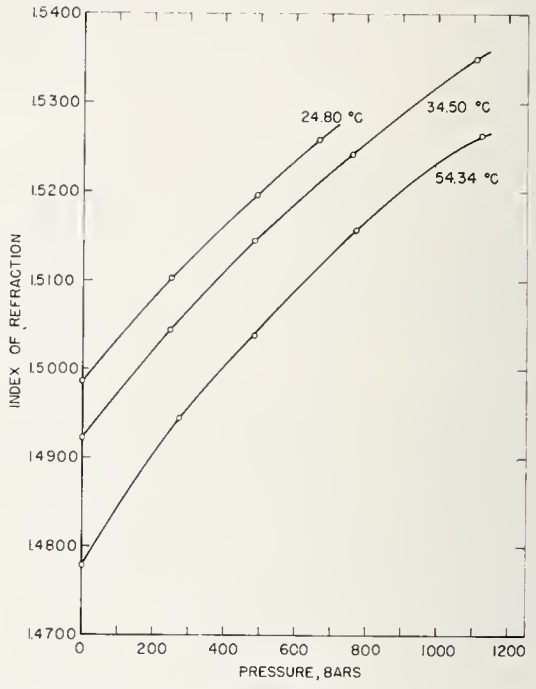


FIGURE 3. *Pressure dependence of index of refraction of benzene.*

Rosen [4] exactly to the third decimal with maximum deviations of two or three units in the fourth decimal or last significant figure reported by these authors.

A similar approach was followed in comparing the data of the present investigation for change of refractive index with temperature at 1 bar with that of Pesce for benzene [19]. Pesce found the change in absolute refractive index in going from 24.80 to 54.34 °C to be −0.01934. The measurement from the present study over the same temperature range was −0.01949.

7.3. Variation of Index at Constant Volume

In the case of the present data, a graph of index versus specific volume is of considerable interest.

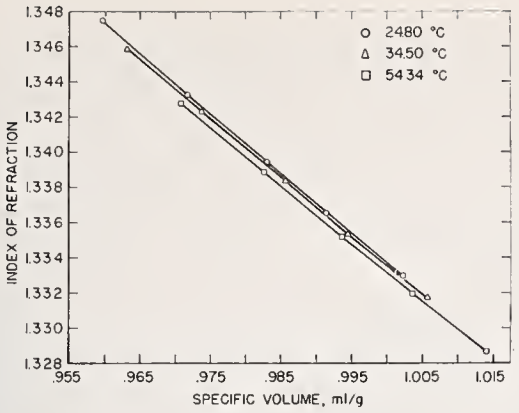


FIGURE 4. *Index of refraction-specific volume relationship in water.*

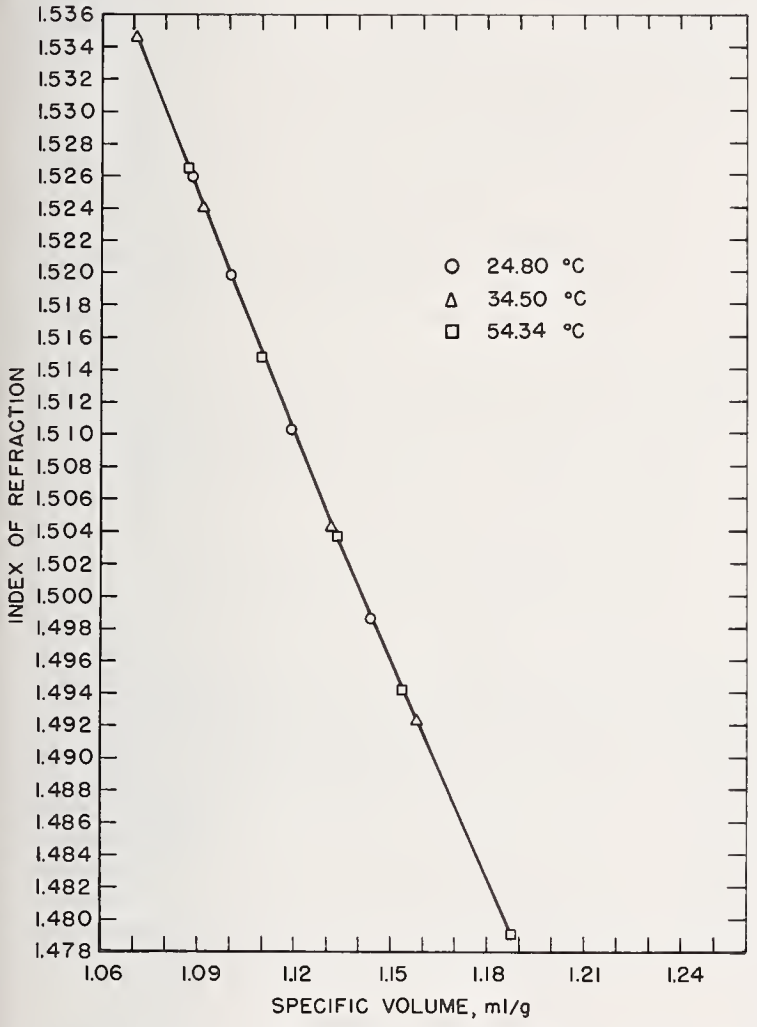


FIGURE 5. *Index of refraction-specific volume relationship in benzene.*

These graphs are given in figures 4, 5, and 6 for water, benzene, and carbon tetrachloride. The relationship is not linear but rather that of a curve convex to the specific volume axis with a very large radius of curvature. In the case of water, the refractive index decreases with increasing temperature at constant specific volume, while in carbon

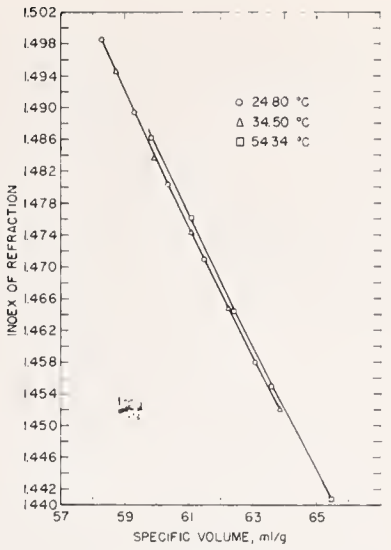


FIGURE 6. *Index of refraction-specific volume relationship in carbon tetrachloride.*

tetrachloride the refractive index increases with increasing temperature, the changes in index in both cases being small—of the order of a few parts in the fourth decimal for a temperature change of 10 °C. It appears that, in the case of benzene, there is no significant dependence of refractive index upon temperature at constant volume, although a larger scale graph of the data shows a separation into three curves of constant temperature with the refractive index decreasing with increase of temperature at the same specific volume. For each liquid there appears to be a tendency for the refractive index to vary less with temperature at higher specific volumes.

These results show that in a liquid where specific volume is held constant, it does not necessarily follow that the refractive index is constant. Despite the fact that this appears to be the first report of this effect, it is not completely unexpected. Gibson and Loeffler [20] observed a shift in the absorption band of solutions of chromophoric materials with temperature at constant volume. Oksengorn [21, 22] has demonstrated a shift of the ultraviolet absorption band of C_6H_6 to longer wavelengths with increasing temperature at constant volume. Robertson [23] has given analogous data for solutions of organic materials in an inert solvent. Gibson and Loeffler [20] attributed the shift to the fact that at higher temperatures the molecules spend a greater fraction of time at close distances to their neighbors. The increased perturbation arising in the close approaches would produce a shift in position of the electronic levels and shift the band. Robertson [23], although attributing the shift with changing specific volume at constant temperature to perturbations arising from Van der Waal forces, believes the shift with temperature at constant volume arises from changes in band shape as a result of increased population of the higher vibrational levels at high temperatures. It is well known that band shapes vary when pressure and temperature are changed [24, 25, 26].

Inasmuch as this represents the first report of change in index at constant volume, as far as can be ascertained, it is of some importance to inquire whether the effect is real or might arise from experimental errors. In order to produce coincidence in the curves for water a maximum shift of the order of 0.0006 in index, or a shift of 0.002 in specific volume or some combination of these two is required. According to the estimates of precision in index the value 0.0006 lies far outside the experimental precision and the value of 0.002 in specific volume is not comparable with the precision of ± 0.00005 in specific volume claimed by Gibson [9]. There remains the additional possibility that the temperatures reported here might be in error sufficiently to produce the observed spread in figures 4 and 6. Temperatures were measured using both a thermocouple and platinum resistance thermometer, with a maximum difference over the temperature range of 0.11 °C. This maximum uncertainty cannot attribute appreciably to the observed separation of the data. In addition, all measurements were performed in exactly the same manner at identical temperatures. The fact that the shifts for water and C Cl_4 are in opposite directions, while C_6H_6 shows no shift within experimental error, plus the estimates of errors involved requires the conclusion that the shifts of index at constant volume are real for H_2O and C Cl_4.

From the previous discussion it is concluded that at constant volume the position and shape of the absorption bands change with temperature. For the liquids studied here data at constant volume appear to be available only for benzene which was studied by Oksengorn [21, 22]. The frequency used here is on the low frequency tail of the uv absorption band as shown by the dispersion data of Gibson and Kincaid [6]. The shift in the uv band with increasing temperature at constant volume is toward lower frequencies and, at the frequency of measurement used here, a higher index should result from this shift.

If it is assumed that the shift in absorption frequency with temperature at constant volume is equivalent to a change in frequency of measurement at constant temperature and pressure, the available data can be combined to calculate the change in index resulting from the shift in the absorption band. This assumption specifically requires that the shape of the absorption band be unchanged by the shift. At a specific volume of 1.10 cm^3/g, Oksengorn's data show a shift $\Delta\nu/\Delta T \approx -1.5$ $cm^{-1}/°C$. Gibson and Kincaid report a frequency dependence $\Delta\mu/\Delta\nu \approx 2.5 \times 10^{-6}/cm^{-1}$ at 1 atm and 25 °C. Therefore, the change in index at constant volume would be $\Delta\mu/\Delta T \approx 3.7 \times 10^{-6}/°C$ with the change in sign resulting from the fact that shift of the absorption band is of opposite sense to the change in index. The calculated change in index in increasing the temperature from 24.80 to 54.34 °C, at constant volume is 1.1×10^{-4}. Considering the several measurements required for the experimental determination of change in index at constant volume, i.e., change in index with pressure at various temperatures, change in index with temperature at 1 bar and expansivity and compressibility of the interferometer, it appears that the predicted change in index of 1.1×10^{-4} would most probably not be detected. The results of the present study do not indicate any significant change in refractive index with temperature at constant specific volume for benzene.

The shift of an absorption band in the ultraviolet region of the spectrum is not the only possible explanation for change in index with temperature at constant volume. Since the dispersion of benzene is large, the index would be very sensitive to changes in band contour. According to Robertson [23] at constant volume an enhancement of the low frequency tail of the absorption band with increasing temperature is to be expected. The results of the present study do not appear to be in accordance with this expectation. Measurements of index at different wavelengths might provide a sensitive method for studying changes in contour of the band tail.

No comparable data at constant volume appear to be available for C Cl_4 or H_2O. The magnitude of the change in index of C Cl_4 with temperature at a constant volume of 0.60 ml/g is about $4.8 \times 10^{-5}/°C$. This result might be explained by a shift of the uv band to longer wavelengths with increasing temperature at constant volume. Alternatively, changes in band contour might cause the observed changes in refractive index.

The data for H_2O show a decrease in index with increasing temperature of magnitude $2.6 \times 10^{-5}/°C$ at a constant volume of 0.97 ml/g. From dispersion data on H_2O [14] it is apparent that the frequency used here is on the tail of the uv absorption band near 1800 A [27]. Therefore, it would appear that the data on index might be explained by a shift of the uv band to shorter wavelengths with increasing temperature at constant volume. It is not clear that shifts in this direction occur since no data at constant volume are available, and data of Collins [28] on the near infrared absorption of water to 5000 atm show no appreciable shifts or intensity changes in this region. However, an appropriate change in contour without a shift of the center of gravity of the band would also produce the observed result. It is possible that the effect observed in water with temperature at constant volume is due to small changes in the hydrogen binding.

7.4. Refractive Index Formulas

Various equations have been proposed which express a relationship between refractive index and specific volume in such a way that the relationship is independent of temperature. The following equations may be mentioned: the Gladstone-Dale [29], $(\mu-1)V$=constant, the Lorentz-Lorenz [30, 31], $(\mu^2-1/\mu^2+2)V$=constant, the Newton [32], $(\mu^2-1)V$=constant and the Eykman [33], $(\mu^2-1/\mu+0.4)V$=constant, where μ represents the index of refraction and V the specific volume. Upon fitting the data of the present investigation to the various

equations it was found, in agreement with the earlier conclusion of Gibson [6], that the Eykman equation for benzene gives a satisfactory fit to four significant figures. However, none of the formulations give satisfactory agreement for each liquid over the range of temperature and pressure employed. This observation is not new [4, 6, 34] and a considerable amount of work has been done, particularly on the Clausius-Mosotti equation, the dielectric analog of the Lorentz-Lorenz equation with the dielectric behavior of both gases and liquids at elevated pressures. Danforth [35] who measured the dielectric constant of liquids up to 12,000 bars believed that his data indicated both a change of polarizability and a failure of the Clausius-Mosotti equation at elevated pressures. A comprehensive survey of much of this work is given by Ten Seldam [36]. More recently it has been realized that the intermolecular forces may be investigated directly by spectroscopic studies at elevated pressures and a survey of current work in this field is available [24].

By using a more refined approximation of the local field acting upon a molecule Böttcher [37] gives as an improved form of the Lorentz-Lorenz equation

$$\frac{\mu^2}{(\mu^2-1)(2\mu^2+1)V}=\frac{m}{12\pi N_A}\left\{\frac{1}{\alpha}-\frac{1}{a^3}\left(\frac{2\mu^2-2}{2\mu^2+1}\right)\right\} \quad (3)$$

where m is the molecular weight, N_A is Avogadro's number, α is the polarizability of the molecule and a is the effective radius of the molecule. It can be seen that by plotting

$$\frac{\mu^2}{(\mu^2-1)(2\mu^2+1)V} \text{ vs } \frac{(2\mu^2-2)}{(2\mu^2+1)}$$

a straight line results with an intercept involving α and a slope involving a. The data for water, benzene, and carbon tetrachloride were plotted according to eq (3) and the graphs are shown in figures 7, 8, and 9. Values of α and a determined by least squares are given in table 5.

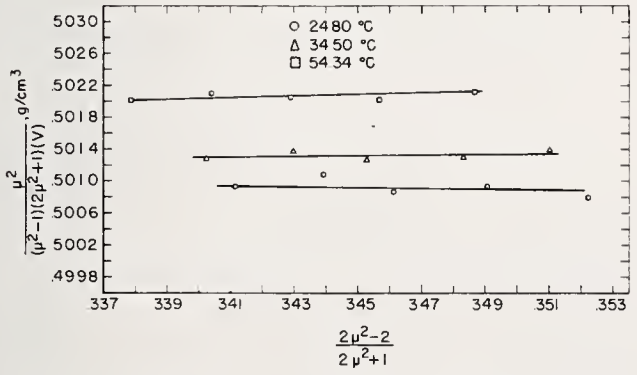


FIGURE 7. *Relationship between* $\frac{2\mu^2-2}{2\mu^2+1}$ *and* $\frac{\mu^2}{(\mu^2-1)(2\mu^2+1)V}$ $\frac{1}{}$ *for the water molecule.*

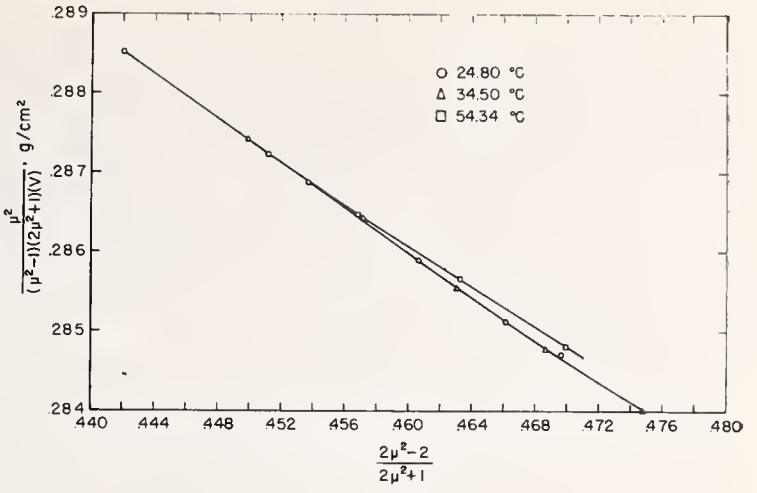


FIGURE 8. *Relationship between* $\frac{2\mu^2-2}{2\mu^2+1}$ *and* $\frac{\mu^2}{(\mu^2-1)(2\mu^2+(1)V}$ $\frac{1}{}$

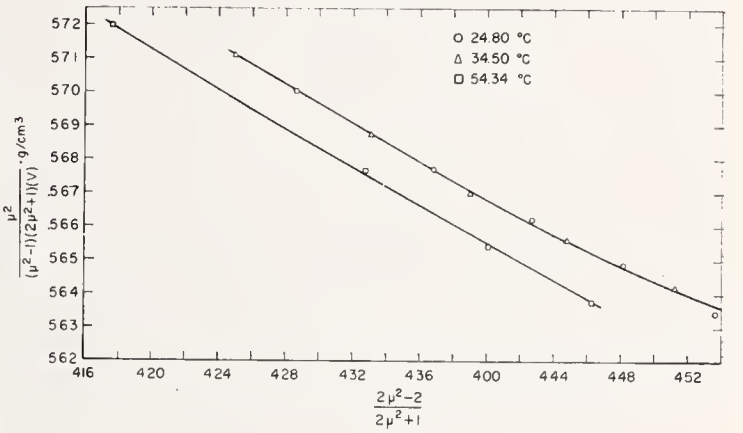


FIGURE 9. *Relationship between* $\frac{2\mu^2-2}{2\mu^2+1}$ *and* $\frac{\mu^2}{(\mu^2-1)(2\mu^2+1)V}$ $\frac{1}{}$ *for the carbon tetrachloride molecule.*

TABLE 5. *Polarizabilities and molecular radii as determined from Böttcher's equation*

Molecule	Temperature	Polarizability, α	Average polarizability, α	Molecular radius, a	Average molecular radius, a
	°C	A^3	A^3	A	A
Water	24.80	1.57	1.58	3.77	3.77
	34.50	1.59		----------	
	54.34	1.59		----------	
Benzene	24.80	9.86	9.86	2.93	2.93
	34.50	9.83		2.92	
	54.34	9.90		2.95	
Carbon tetrachloride	24.80	9.93	9.88	2.96	2.93
	34.50	9.93		2.96	
	54.34	9.79		2.87	

Böttcher used as his model a spherical molecule whose polarizability and radius were independent of temperature and pressure. It can be seen from the figures that, of the three liquids, benzene fits this model best, but even with benzene an increasing curvature is apparent at the right-hand side of the graph. This region corresponds to the physical situation of the liquid being under greater pressure. It may be inferred, therefore, that polarizability and/or radius of the benzene molecule is not completely independent of pressure.

This curvature may also be observed for carbon tetrachloride in figure 9. There is here, in addition, a lateral displacement of the isotherms. The fact that both the intercept and the slope of the lines representing the data for 24.80 and 54.34 °C are distinctly different indicates that both polarizability and radius are not independent of temperature (as well as pressure) for $C Cl_4$.

Böttcher's formula, eq (3), was developed for a nonpolar dielectric, and figure 7 shows that water fits Böttcher's model least well of the liquids studied. Despite some scattering of points there appears to be a definite separation of the isotherms. A meaningful estimate of molecular radius, a, may only be had for the data at 24.80 °C, because the slope of the curve at this temperature is negative as is necessary for the calculation. The slopes of the curves at the two higher temperatures are slightly positive.

For each liquid average values of α and a over the temperature range involved are shown in table 5. For carbon tetrachloride it was found that $\alpha=9.88$ A^3 and $a=2.93$ A. Böttcher calculated $\alpha=9.07$ A^3 and $a=2.42$ A by applying his equation to a series of mixtures of carbon disulfide and carbon tetrachloride. Average values of a for water, carbon tetrachloride, and benzene were found to be 3.77 A, 2.93 A, and 2.93 A, respectively. From molar volume data for these liquids over the same temperature range, corresponding values of a were found to be 1.94 A, 3.33 A, and 3.40 A. The large difference in the case of water is again presumed to be due to the fact that water is a polar liquid.

The authors are indebted to R. E. Gibson, Director of the Applied Physics Laboratory of the Johns Hopkins University, for the use of the pressure vessel in which the experiments were conducted. The authors also thank I. Malitson of the Refractometry Section of the National Bureau of Standards for his initial measurements of the refractive index of carbon tetrachloride.

8. References

[1] T. C. Poulter, C. Ritchey, and C. A. Benz, Phys. Rev **41,** 366 (1932).
[2] T. C. Poulter, Phys. Rev. **40,** 860 (1932).
[3] T. C. Poulter and C. A. Benz, Phys. Rev. **40,** 872 (1932).
[4] J. S. Rosen, J. Opt. Soc. Am. **37,** 932 (1947).
[5] W. J. Lyons and F. E. Poindexter, J. Opt. Soc. Am. **26,** 146 (1936).
[6] R. E. Gibson and J. F. Kincaid, J. Am. Chem. Soc. **60,** 511 (1938).
[7] J. Jamin, Compt. rend **45,** 892 (1857).
[8] A. Michels and J. Hamers, Physica **4,** 995 (1937).
[9] R. E. Gibson, J. Am. Chem. Soc. **59,** 1521 (1937).
[10] R. E. Gibson, J. Am. Chem. Soc. **57,** 284 (1935).
[11] J. B. Saunders, J. Research NBS **35,**157 (1945) RP1668.
[12] P. W. Bridgman, The Physics of High Pressure (G. Bell & Sons Ltd., London 1949).
[13] A. F. Forziati, J. Research NBS **44,** 373 (1950) RP2085.
[14] L. W. Tilton and J. K. Taylor, J. Research NBS **20,** 419 (1939) RP1085.
[15] Smithsonian Physical Tables, 9th ed. (Washington, D.C., 1954).
[16] A. Pérard, Travaux et Memoires du Bureau International des Poids et Mesures **19** (1932).
[17] R. E. Gibson and O. H. Loeffler, J. Am. Chem. Soc. **63,** 898 (1941).
[18] N. E. Dorsey, Properties of Ordinary Water Substance (Reinhold Publ. Corp., New York, N.Y., 1940).
[19] B. Pesce, Gazz. Chim. ital. **65,** 440 (1935).
[20] R. E. Gibson and O. H. Loeffler, J. Am. Chem. Soc. **62,** 1324 (1940).
[21] B. Oksengorn, Compt. rend. **242,** 2324 (1956).
[22] Ref. 24 p. 323.
[23] W. W. Robertson, J. Chem. Phys. **33,** 362 (1960).
[24] Colloques Internationaux du Centre National de la Recherche Scientifique Proprietes Optiques et Acoustiques des Fluides Comprimes et Actions Intermoleculaires (Centre National de la Recherche Scientifique, Paris, 1959).
[25] A. Kronenberger and P. Pringsheim, Z. Physik **40,** 75 (1926).
[26] E. Fishman and H. G. Drickamer, J. Chem. Phys. **24,** 548 (1956).
[27] K. Watanabe and M. Zelikoff, J. Opt. Soc. Am. **43,** 753 (1953).
[28] J. R. Collins, Phys. Rev. **36,** 305 (1930).
[29] J. H. Gladstone and T. P. Dale, Phil. Trans. **153,** 337 (1863).
[30] H. A. Lorentz, Ann. Physik **9,** 641 (1880).
[31] L. Lorenz, Ann. Physik **11,** 70 (1880).
[32] R. A. Houstoun, A Treatise on Light (Longmans, Green & Co., New York, 1938).
[33] J. F. Eykman, Rec. Trav. Chim. **14,** 177 (1895).
[34] F. E. Poindexter and L. E. James, Phys. Rev. **42,** 910 (1932).
[35] W. E. Danforth, Phys. Rev. **38,** 1224 (1931).
[36] C. A. Ten Seldam, Dissertation Univ. of Utrecht, March 1953 (Van Gorcum & Co., N. V., Assen, Netherlands, 1953).
[37] C. J. F. Böttcher, Theory of Electric Polarization (Elsevier Publishing Co., Amsterdam, 1952).

(Paper 67A2–203)

JOURNAL OF RESEARCH of the National Bureau of Standards—A. Physics and Chemistry
Vol. 69A, No. 4, July–August 1965

Effect of Hydrostatic Pressure on the Refractive Indices of Some Solids

R. M. Waxler and C. E. Weir

(March 3, 1965)

Measurements were made on refractive index changes with hydrostatic pressures between 1 bar and 1 kbar using the helium yellow line. The materials studies were: KBr NaCl, LiF, diamond, MgO, quartz, Al_2O_3, and three silicate glasses. All the materials increased in refractive index with pressure except diamond, MgO, and Al_2O_3 which decreased, and LiF which did not change. The results were compared with photoelastic measurements, and Poekel's geometric theory of photoelasticity was substantiated as well as Mueller's physical theory. The data show that the ratio of change of polarizability with density is greater for solids having stronger interatomic repulsive forces. Volume and temperature coefficients of polarizability were evaluated for the cubic crystals and glasses. The thermo-optic behavior of crystals and glasses is discussed.

1. Introduction

The relationship between index of refraction and density is of importance because of the effect of atomic interactions on the atomic polarizabilities. For transparent solids, values of $\rho d\mu/d\rho$, where μ is the index and ρ the density, have been obtained by calculation from the photoelastic constants [1].[1] For crystals as well as glasses values of $\rho d\mu/d\rho$ so calculated usually have been found to be less than the corresponding values calculated on theoretical grounds, i.e., from the Lorentz-Lorenz and the Drude relationships [1]. The most extreme example of this discrepancy is probably found in the case of diamond and MgO where $\rho d\mu/d\rho$ values calculated from photoelastic constants are negative, a result which is clearly impossible to reconcile with the assumption of constant polarizability involved in the initial derivation. The theory of photoelasticity due to Mueller [2, 3] recognizes the change in the intrinsic polarizability of an atom due to strain.

For glasses, data on $\rho d\mu/d\rho$ have been obtained by Ritland [4] (and others) in the annealing temperature range where both index and density change with time at constant temperature and on volumetric relaxation following removal of high hydrostatic pressures [5]. In the former case $\rho d\mu/d\rho$ is found to be somewhat less than theory predicts and in the latter case somewhat greater.

Direct experimental data on $\rho d\mu/d\rho$ are required to understand thermo-optic properties. Values of $d\mu/dT$, where T is the temperature, have been measured for many solids and it is known that the sign may be either positive or negative. For most crystals $d\mu/dT$ is found to be negative but in the case of diamond, MgO, and ZnS the reverse is true [6]. Recent work has shown that for Al_2O_3 $d\mu/dT$ is positive for both the ordinary and extraordinary rays [7]. Most optical glasses exhibit a minimum in the μ-T curve somewhat below room temperature so that at and above room temperature $d\mu/dT$ is positive but at some lower temperature the sign reverses [8]. Any attempt to explain such a complex thermo-optic behavior must of necessity require an understanding of the $\rho d\mu/d\rho$ effect because thermal dilatation is an important contributing factor to the value of $d\mu/dT$.

Finally, by reversal of the customary procedure, data on $\rho d\mu/d\rho$ may be used to evaluate photoelastic constants in certain instances. For glasses only two constants are required, so that knowledge of $\rho d\mu/d\rho$, plus information on the optical path difference produced by a unidirectional stress serves to determine completely the photoelastic constants.

For these several reasons it was considered of interest to obtain direct measurements of the density coefficient of refractive index, $\rho d\mu/d\rho$, in solids at constant temperature by the application of hydrostatic pressure. This is the first time that such an approach has been used. This report contains data on $\rho d\mu/d\rho$ obtained in the pressure interval between 1 bar and approximately 1000 bars for several crystals and glasses of interest. Data were obtained from specimens of the cubic ionic crystals, LiF, KBr, NaCl, and MgO; the covalent cubic crystal, diamond; and the birefracting crystals, Al_2O_3 and quartz where both the ordinary and extraordinary rays were studied. The glasses studied were fused silica, a commercial plate glass and a borosilicate crown glass, BSC 517/645. Measurements reported here were made only at 25 °C using the helium yellow line.

2. Experimental Method

In these experiments interference fringes were observed in a plate of the material having plane, polished, nearly parallel faces [9]. The fringes were viewed in reflection using collimated helium light of

[1] Figures in brackets indicate the literature references at the end of this paper.

$\lambda=5875.62$ Å at normal incidence. For a specimen of thickness, t, and index, μ, the fringe number N is given by the expression

$$N\lambda=2t\mu. \tag{1}$$

Application of hydrostatic pressure will produce a change in μ, a decrease in t, and a shift in the interference fringes. By differentiating (1) and rearranging it is found that

$$\Delta\mu=\frac{\Delta N\lambda}{2t}-\mu\frac{\Delta t}{t}. \tag{2}$$

A measurement of the number of fringes that pass a point of reference, and the thickness together with data on the initial index and the linear compression $\Delta t/t$ permits calculation of the change in index.

The experimental apparatus and method have been described earlier in detail and will only be outlined here [10]. Briefly, a plate of the test specimen is immersed in liquid in a pressure vessel equipped with glass windows. Hydrostatic pressure is generated by compressing the liquid and the number of interference fringes passing a reference line, i.e., the fringe count, was determined visually. Estimates of the number of fringes were made to the nearest tenth of a fringe.

The change in fringe number observed in approximately 1 kbar is not very large for these solids because of their rather low compressibilities compared to the liquids previously studied and because the decrease in thickness under pressure acts in such a direction as to decrease the fringe number (see eq (2)). In the most favorable case ΔN is known only to three significant figures. Extreme precision in the other quantities is not required. The most reliable values for the initial indices were used, however, and values for the index at the frequency of the He yellow line were generally obtained by interpolation of the dispersion data reported. The data on the indices of refraction of the solids may be found in references [11] through [20]. For BSC 517/645 the index was reported for the sodium D line only. However, from the low dispersion of such glasses and the small frequency separation of the sodium and helium yellow lines, the value reported was assumed to apply to the present measurements within the required precision.

Compression data for all materials except BSC 517/645 have been reported, usually in terms of volume compression ($\Delta V/V_0$). Volume compressions were converted to linear values for the isotropic materials through the relationship $\Delta t/t_0=\frac{\Delta V}{3V_0}$. For quartz and Al_2O_3, where data on linear compressions are available, $\Delta t/t_0$ was computed for the direction perpendicular to the optic axis. The data for determining the compression of the solids may be found in references [21] through [26]. (For BSC 517/645 the linear compression was calculated from elastic constant data measured at 1 bar [27].) It has been estimated that, at most, the uncertainty in values of $\Delta t/t_0$ lies in the third significant figure.

For the optically isotropic solids (cubic crystals and glasses) the measurements were straightforward but for the crystals of lower symmetry, Al_2O_3 and quartz, which are optically anisotropic the method was modified to permit measurements for both the ordinary and extraordinary ray. For these materials the specimen was oriented so that the optic axis was perpendicular to the direction of the light beam. In the measurements a large polarizer was inserted between the source and the window of the pressure vessel. On rotating the polarizer two distinct fringe systems could be seen depending on whether the electric vector of the plane polarized light transmitted by the polarizer was parallel or perpendicular to the optic axis of the specimen. (For the quartz crystal which is optically active, it should be noted that, since the fringes were obtained by reflection, the emergent polarized light was still plane polarized in the same plane at which it was incident [28]. This would not have been true if the fringes were observed in transmission [16].) On applying hydrostatic pressure the fringe shift for each separate fringe system was noted, giving data on both rays for the birefringent crystals.

3. Preparation of Specimens

Most specimens were prepared by sectioning or clearing larger samples of commercially available synthetic single crystals. However, both quartz and diamond were prepared from naturally occurring single crystals. The glasses were of commercial origin. Although the test specimen required was approximately 1 cm $\times$ 1 cm $\times$ 0.5 cm, it was found expedient to grind and polish larger specimens and then to cut the final specimen from the most perfectly ground portion of the large piece. Rough blanks of both Al_2O_3 and quartz were cut with the optic axes in approximately the correct orientation. The blanks were then oriented using the Laue back-reflection x-ray technique and ground so that the optic axis was in the surface of the specimen.

3.1. Grinding and Polishing

A brief account is given here of the preparation of the crystals because these techniques are not widely known. The specimens were first ground flat and with nearly parallel surfaces using a surface grinder for the harder materials and using hand grinding with fine abrasive for the alkali halides. The harder crystals were prepolished on a wood lap with 8μ–22μ diamond powder using olive oil lubricant and then given a final polish on cloth impregnated with 3μ diamond dust using an alcohol-water solution as a lubricant. The alkali halides were given only the latter of the two polishing steps. Al_2O_3 specimens were available only in small diameters and the grinding and polishing operation was conducted on an assemblage of five small disks, set in wax on a flat glass plate.

The diamond was a 2 carat brilliant cut gemstone and the table face was found to be satisfactory for one surface. The culet was ground down parallel to

the table face using a high speed steel lap impregnated with diamond powder to produce a surface approximately 1 mm in width. Satisfactory fringes could be observed through this small surface.

In all polishing operations it was necessary to check the fringe system at frequent intervals. If it was noted that the surfaces were deviating too much from parallelism adjustments were made in the polishing to reduce the angle between the surfaces.

3.2. Aluminizing

In order to improve the sharpness of the interference fringes the reflecting power of the surfaces of the specimens was increased by vacuum deposition of a layer of aluminum. The front surface of each specimen was covered with a partly reflecting film while the rear surface was made fully reflecting. Multiple reflection fringes were obtained in this manner but because of the large thickness to length ratio of the specimens the fringes were not extremely sharp. A series of depositions of different thicknesses were made on the front surface of plate glass specimens until the fringes appeared to reach a maximum intensity and sharpness. This film thickness was found to be satisfactory for all other specimens.

TABLE 1. *Pressure induced changes in refractive index as measured with helium yellow radiation of 5875.62 Å*

Material	Refractive index, μ_{He}	Specimen thickness, t_0	Pressure, P	Thickness change, $-\Delta t/t_0 \times 10^3$	Observed change in fringe number $-\Delta N$	Change in refractive index $\Delta\mu \times 10^5$
		Cm	*bars*			
KBr	1.560	0.3210	966.1	2.118	11.3	227
NaCl	1.546	.6748	963.6	1.348	21.6	115
LiF	1.392	.5683	963.2	.349	9.4	0
MgO	1.738	.3970	960.8	.190	6.8	−17
Diamond	2.418	.4037	960.0	.058	3.4	−11
Quartz:						
ω	1.545	.5112	957.6	.944	6.9	103
ϵ	1.554	.5112	961.5	.948	7.5	107
Sapphire:						
ω	1.769	.5659	961.8	.105	6.2	−14
ϵ	1.760	.5659	957.6	.105	6.2	−14
Fused SiO_2	1.458	.4967	960.2	.865	7.3	83
BSC 517/645	1.517	.6693	960.1	.717	8.9	70
Plate glass	1.518	.6566	960.0	.723	9.1	69

4. Results

The results of the experiments are given in table 1 which shows the decrease in fringe number, the dimensional change calculated at the pressure given and the change in index of refraction. It should be noted that the change in fringe number is presented in table 1 for the direction of increasing pressure, i.e., between 1 bar and the pressure listed. For birefracting crystals the index for the ordinary and extraordinary rays are noted by ω and ϵ respectively. Each value of ΔN given in table 1 represents the average of at least two determinations. Such duplicate measurements agreed to within ± 0.1 interference fringe. From the agreement between the duplicate measurements the limit of reproducibility in the index change is estimated to be $\pm 2 \times 10^{-5}$.

From the tabular data it is seen that the change in index is quite small in absolute value, varying from 0 in LiF to 2 in the third decimal place for KBr. The index decreases for MgO, diamond and both the ordinary and extraordinary rays of Al_2O_3. All other materials show an increase in index except LiF for which the index is unchanged. For the birefracting crystals both the ordinary and extraordinary rays show the same change in index within the limits of error.

5. Discussion

5.1. Cubic Crystals

From the experimental values of $\triangle\mu$ and the compressibility data, it was possible to calculate values for the quantity $\rho \frac{\Delta\mu}{\Delta\rho}$ where ρ is the density. For the cubic crystals studied here the theory of photoelasticity shows that $\rho \frac{d\mu}{d\rho}$ can be calculated by the relationship

$$\rho \frac{d\mu}{d\rho} = \frac{\mu^3}{6}(p_{11}+2p_{12}), \qquad (3)$$

where the elasto-optic constants, p_{ij}, relate the state of strain existing in a crystal to changes in the index ellipsoid [6]. These values have been calculated for a number of crystals [1, 6] and the results are given in table 2 together with the initial indices, μ. [The values of these experiments obtained over the pressure interval of approximately 1 kb are denoted $\rho\Delta\mu/\Delta\rho$ while those from the photoelastic theory are designated as $\rho d\mu/d\rho$]. It may be noted that the photoelastic data were obtained for the sodium D line and the present data for the helium yellow line. However, because of the small frequency difference between the two lines and the known low-frequency dependence of $\rho d\mu/d\rho$ [29, 30] any differences arising from dispersion effects can be taken to be negligible.

TABLE 2. *Values of* $\rho\left(\frac{\Delta\mu}{\Delta\rho}\right)$ *and* λ_0 *for cubic crystals and glasses*

Material	Refractivo index, μ_0	$\rho\left(\frac{\Delta\mu}{\Delta\rho}\right)$	$\rho\left(\frac{d\mu}{d\rho}\right)$	L-L equation		Drude equation	
				$\rho\left(\frac{\partial\mu}{\partial\rho}\right)$	λ_0	$\rho\left(\frac{\partial\mu}{\partial\rho}\right)$	λ_0
KBr	1.559	0.35	0.35	0.68	0.48	0.46	0.24
NaCl	1.544	.28	.24	.66	.58	.45	.38
LiF	1.392	.00	.1	.44	1.00	.34	1.00
MgO	1.736	−.31	−.40	.97	1.32	.52	1.60
Diamond	2.417	−1.58	−.28	2.62	1.60	1.00	2.58
Fused SiO_2	1.458	.32	.29	.53	.40	.39	.17
BSC 517/645	1.517	.33		.62	.47	.43	.24
Plate glass	1.518	.32		.62	.49	.43	.26

A comparison of the figures in table 2 shows that there is essential agreement between the predictions of photoelastic theory developed by Pockels [31] and the actual measured values. Considering the errors inherent in the measurements of the photoelastic constants [6], i.e., small fringe shifts, nonuniformity of stress distribution in uniaxial compression, etc., it is considered that the agreement in table 2 is satisfactory except in the case of diamond. For this material, however, the errors in the photoelastic experiments and in the present work are likely to be particularly large because of the small specimens available and the small changes that are observed. There is no question as to the negative sign of $\rho\Delta\mu/\Delta\rho$ for diamond, and it is interesting to note that the results of the present experiments bear out the theoretical predictions of negative values of $\rho\Delta\mu/\Delta\rho$ for both MgO and diamond.

The data for diamond obtained in this study are in much better agreement with the results of Gibbs and Hill [32] on the change of the dielectric constant of diamond with pressure. Gibbs and Hill noted that the dielectric constant of diamond, $\epsilon=5.66$ at 27 °C at frequencies as high as 10 MHz is equal to the square of the optical index of refraction at infinite wavelength, μ_∞, reported by Peter [14]. Since ρ is proportional to $1/V$, $(\partial \ln \rho/\partial P)_T=-(\partial \ln V/\partial P)_T=\beta$, where P is the pressure and β the compressibility. It follows that

$$\rho\left(\frac{\partial\mu_\infty}{\partial\rho}\right)_T\equiv\left(\frac{\partial\mu_\infty}{\partial P}\frac{\partial P}{\partial \ln \rho}\right)_T=\frac{1}{\beta}\left(\frac{\partial\mu_\infty}{\partial P}\right)_T.^2 \quad (4)$$

If $\epsilon=\mu_\infty^2$, it follows that

$$\frac{1}{\mu_\infty}\left(\frac{\partial\mu_\infty}{\partial P}\right)_T=\frac{1}{2\epsilon}\left(\frac{\partial\epsilon}{\partial P}\right)_T. \quad (5)$$

If the value of $\rho\left(\frac{\Delta\mu}{\Delta\rho}\right)$ in the present study is used for $\rho(\partial\mu_\infty/\partial\rho)_T$ in eq (4), and employing the equivalence shown in eq (5), $(1/\epsilon)(\partial\epsilon/\partial P)_T$ for diamond has been found to have the value -2.28×10^{-7} bar $^{-1}$. The value obtained experimentally by Gibbs and Hill for a type I diamond is -2.40×10^{-7} bar^{-1}.

[2] The derivation leading to eq (4) was suggested by D. D. Wagman of NBS.

5.2. Glasses

Experimental values for $\rho\Delta\mu/\Delta\rho$ for the glasses are given in table 2. The values are very similar for all three glasses and probably reflect the fact that the data are determined largely by the SiO_4 tetrahedra common to all three glasses. Numerical comparisons with photoelastic theory can be made only for vitreous SiO_2 for which Jog [29] has reported the necessary photoelastic data. As seen in the table the value of $\rho d\mu/d\rho$ calculated from the photoelastic constants is in reasonably good agreement with the value obtained here by direct measurement. All three glasses show positive values for $\rho\Delta\mu/\Delta\rho$. Furthermore the photoelastic data on numerous optical glasses reported by Schaeffer and Nassenstein [33], Vedam [34], and Mueller [3] invariably indicate positive values for $\rho d\mu/d\rho$. This fact will be discussed later.

5.3. Noncubic Crystals

It is interesting to note that the increase in index for both rays in quartz is greater than observed in fused silica. The result can be explained only by the conclusion that interatomic interactions play an effective role in the index change with density and that those are different in the crystal and the glass because of the crystal structure and the higher initial density of the crystal. From the data of table 1 it appears that the extraordinary ray suffers the larger change in index but the difference is probably within the experimental error. Pockels [35] has measured the photoelastic constants of quartz and his results can be extrapolated linearly to the present pressure of 957 bars to permit calculation of $\Delta\mu$. Under these circumstances the photoelastic data yield calculated values for the change of index of 100×10^{-5} for the ordinary ray and 104×10^{-5} for the extraordinary ray. These values are in good agreement with the results of the present studies. In contrast to quartz both rays of sapphire show the same change in index but both are negative. There do not appear to be enough photoelastic data available to permit calculations of the index change for Al_2O_3 [7].

5.4 Photoelastic Constants of Glass

From the present data on $\Delta\mu/\Delta P$ and auxiliary data on the amount of double refraction introduced into glass by uniaxial compression it is possible to compute the constants q_{11} and q_{12}, where the piezooptic constants, q_{ij}, relate the state of stress which exists in a solid to changes in the index ellipsoid [6]. The relationships used [34] are as follows:

$$\Delta\mu/\Delta P=\mu^3[q_{11}+2q_{12}]/2 \quad (6)$$

and

$$(\Delta\mu_2-\Delta\mu_1)/\Delta P=\mu^3[q_{12}-q_{11}]/2, \quad (7)$$

where, in addition to the terms previously defined, $\Delta\mu_2$ and $\Delta\mu_1$ represent changes in index for light polarized respectively perpendicular and parallel to

the stress direction. (In eq (7), ΔP denotes a change in a compression applied in one direction only.) The quantity $(\Delta\mu_2-\Delta\mu_1)/\Delta P$ is defined as the relative stress optical coefficient and has been reported to be 2.81 brewsters for BSC 517/645 [36] and 2.62 brewsters for plate glass [37]. These data plus values of $\Delta\mu/\Delta P$ serve to determine values for q_{11} and q_{12} from eq (6) and eq (7). These values are given in table 3. Values of Neumann's photoelastic constants, p and q, are also given in table 3; the relationships used in evaluating them are given by Vedam [34].

The elasto-optic constants, p_{11} and p_{12} can also be calculated from the relationships given by Vedam [34]. Values of the elastic moduli needed in the calculations are given by Spinner for BSC 517/645 [27]. From measurements made on commercial plate glass [38], values for Young's modulus, the rigidity modulus and Poisson's ratio have been found to be 723.9 kbar, 297.9 kbar, and 0.215 respectively. These values were used to calculate the elasto-optic constants for plate glass shown in table 3.

TABLE 3. *The photoelastic constants of BSC 517/645 and plate glass*

Glass	p	q	p_{11}	p_{12}	$q_{11}\times10^{13}$	$q_{12}\times10^{13}$
					cm²/dyne	*cm²/dyne*
BSC 517/645	0.168	0.087	0.115	0.221	0.315	1.92
Plate glass	.154	.087	.114	.203	.369	1.87

5.5. Polarizability and Density

It seems self-evident that density changes must be accompanied by some change in molecular or atomic polarizability and it is useful to analyze the present data to ascertain what information can be obtained in this connection. The most comprehensive theory appears to be due to Mueller [2, 3] who considers the following factors involved in change of refractive index of a solid under applied stress: (1) change of density; (2) change of the Lorentz-Lorenz field; (3) change of the coulomb field; and (4) change in the intrinsic polarizabilities of the atoms. Mueller appears to be the first investigator to recognize the importance of the last factor. The theory has been applied so far only to cubic crystals and isotropic solids because in these cases the symmetry is such that the coulomb field can be taken to be zero. Under hydrostatic pressure the symmetry is unchanged and the coulomb field remains zero but the Lorentz-Lorenz field may change because of the change in density.

There are two principal theoretical relationships relating the index and density, the Lorentz-Lorenz,

$$\frac{\mu^2-1}{\mu^2+1}\left(\frac{1}{\rho}\right)=\alpha_1 \tag{8}$$

and the Drude,

$$(\mu^2-1)/\rho=\alpha_2 \tag{9}$$

where α_1 and α_2 are constants. In deriving the Lorentz-Lorenz equation the local field has been accounted for through the theoretical value $4/\pi 3\ \overline{P}$ where $\overline{P}$ is the polarization of the medium and the change in local field with density is accounted for in the equation. The change in the local field with density has been ignored in the Drude equation.[3]

If α_1 and α_2 are rigorously constants eq (8) and eq (9) may be differentiated to obtain the change of index with density. This was apparently first done by Pockels [39] who obtained the two expressions

$$\rho\frac{\partial\mu}{\partial\rho}=(\mu^2-1)(\mu^2+2)/6\mu \tag{10}$$

$$\rho\frac{\partial\mu}{\partial\rho}=(\mu^2-1)/2\mu \tag{11}$$

corresponding to the Lorentz-Lorenz and Drude formulations respectively. In eq (10) and eq (11) partial derivatives are used to indicate that there may be a change in atomic polarizability which is neglected. Mueller [2, 3] introduced the strain polarizability parameter, λ_0, to evaluate the change in polarizability by means of the relationship,

$$\rho d\mu/d\rho=(1-\lambda_0)\rho\partial\mu/\partial\rho. \tag{12}$$

This relationship follows from considering α_1 and α_2 to be dependent on the polarizability. By using eq (12) in conjunction with eqs (10) and (11) and the experimental values of $\rho\Delta\mu/\Delta\rho$ values of λ_0 have been calculated for both the Lorentz-Lorenz and the Drude formulations. These data are shown in table 2.

Burstein and Smith [1] have proposed that λ_0 is a measure of the degree of homopolar binding. According to Burstein and Smith λ_0 should attain a maximum value for intermediate bonding, i.e., bonding containing appreciable amounts of both ionic and homopolar bonding, and decrease as the bonding becomes either more ionic or more homopolar. The present studies do not agree with this conclusion, the value of λ_0 being largest for the most covalent crystal-diamond. It appears that the conclusions of Burstein and Smith were based upon erroneous data for diamond.

For the glasses, calculated values of λ_0 are also given in table 2. These values show little variation for the three glasses but it is noteworthy that λ_0 data calculated by both the Lorentz-Lorenz and the Drude equations are positive. Previous data on liquids show that λ_0 is positive when computed from the L–L equation but negative when computed from the Drude equation [30]. This behavior in liquids was explained as arising from the change in the local field brought about by compression of the liquids

[3] Krishan and Roy [40] have shown the equivalence of the Lorentz-Lorenz and Drude equations in expressing the frequency dependence of the refractive index and have shown that the Drude expression accounts for the local field by incorporating its effect into the value of the fundamental absorption frequency. However, eq (9) is a simplified version of the generalized Drude equation and is valid only at a single frequency. When written in the form of eq (9) it is implied that the change in the local field with density has been ignored.

518-743 O - 73 - 6

and incorporated into a change in the fundamental absorption frequency in the Drude model. This greater change in the local field of the liquids arises from their greater compressibilities.

Ramachandran and Radhakrishnan [41, 6] have shown that the thermal properties of the index of refraction may be understood by considering the change of index with temperature to consist of three independent contributions which they label P, Q, and R. P represents the change arising only from a change in the number of scattering centers and is a pure density effect, Q represents the change in polarizability resulting from the density change, and R consists of the effect of temperature alone on the polarizability. Recognizing that $d\mu/dT=P+Q+R$, the values they derived using the Drude equation are as follows:

$$P=-\gamma\left(\frac{\mu^2-1}{2\mu}\right) \tag{13}$$

$$Q=-\gamma\left(\rho\frac{d\mu}{d\rho}-\frac{\mu^2-1}{2\mu}\right) \tag{14}$$

$$R=\frac{d\mu}{dT}+\gamma\rho\frac{d\mu}{d\rho} \tag{15}$$

where $\gamma=-\frac{1}{\rho}\left(\frac{d\rho}{dT}\right)_P$. Values of $\rho d\mu/d\rho$ they used [41, 6] were evaluated from the elasto-optic constants (eq (3)). The quantities P, Q, and R have been re-evaluated using the same data that Krishnan and Roy used except that experimental values of $\rho\Delta\mu/\Delta\rho$ have been substituted for $\rho d\mu/d\rho$. [There is one further change; it appears that the earlier results were calculated using an erroneous value for the volume expansivity of diamond, 1.3×10^{-5}/°C [41, 6] compared to the correct value, 0.31×10^{-5}/°C [42, 9].] The results are given in table 4. The present data agree reasonably well with the earlier results in most instances, there being slight changes in Q and R resulting from the present values of $\Delta\mu/\Delta\rho$. There is a large change, however, in the results for diamond.

It is also possible to calculate R, which is defined as $(d\mu/dT)_V$, from constant volume data. For a given material the calculations involve finding a temperature at 1 atm where the specific volume is the same as that at 25 °C and an elevated pressure; the refractive indices which pertain for the two different situations may then be used to find R. Such calculations have been made using values of γ and $(d\mu/dT)_P$ as a function of temperature. The original sources for the values of γ and $(d\mu/dT)_P$ for the different solids have been identified in the list of references.[4] Much of the necessary information may be found in the compilation made by Krishnan [6]. The values are given in the column headed R (experimental) in table 4. It can be seen that there is reasonably good agreement between corresponding values of R in the two columns.

[4] BSC 517/645 is known to have the same nominal composition as the glass designated as BSC-2 in reference [8].

TABLE 4. *Analysis of change in refractive index and polarizability for cubic crystals and glasses*

Material	μ_D	$\gamma\times10^5$	$d\mu/dT\times10^5$	$P\times10^5$	$Q\times10^5$	$R\times10^5$	$R\times10^5$ (experimental)	$\left(\frac{V}{\alpha}\frac{\partial\alpha}{\partial V}\right)_T$	$\left(\frac{V}{\alpha}\frac{\partial\alpha}{\partial V}-1\right)_T$	$\left(\frac{V}{\alpha}\frac{\partial\alpha}{\partial V}-1\right)_T\gamma\times10^5$	$\left(\frac{1}{\alpha}\frac{\partial\alpha}{\partial T}\right)_V\times10^5$	$\left(\frac{1}{\alpha}\frac{\partial\alpha}{\partial T}\right)_V\times10^5$ (experimental)
		°C^{-1}	°C^{-1}	°C^{-1}	°C^{-1}	°C^{-1}	°C^{-1}			°C^{-1}	°C^{-1}	°C^{-1}
KBr	1.559	12.0	−3.6	−5.5	1.3	0.6	0.4	0.2	−0.8	−9.4	1.6	0.9
NaCl	1.544	12.0	−3.8	−5.4	2.0	−0.4	−.3	.4	−.7	−7.8	−.7	−.7
LiF	1.392	10.2	−1.3	−3.4	3.4	−1.3	−1.3	1.0	0.0	0.0	−3.8	−3.9
MgO	1.736	3.9	1.6	−2.3	3.5	0.4	0.6	1.5	.5	2.6	0.7	1.0
Diamond	2.417	0.31	1.0	−0.3	0.8	.5	.6	1.7	.7	0.2	.8	0.6
Fused SiO_2	1.458	.15	0.92	−.06	.01	.97	1.0	0.2	−.8	−.1	2.5	2.6
BSC 517/645	1.517	2.0	.17	−.9	.2	.8	.7	.2	−.8	−1.5	1.9	1.5
Plate glass	1.518	1.9	.27	−.8	.2	.9		.3	−.7	−1.4	2.0	

For the present purposes where volume and temperature may both be considered as independent variables it appears more desirable to change the analysis slightly as follows: Starting with the Drude relationship (eq (9)) and considering the polarizability α_2 to be a function of V and T, it can be shown that

$$d\mu/\mu=\frac{(\mu^2-1)}{2\mu^2}\left[\frac{1}{\alpha}\left(\frac{\partial\alpha}{\partial T}\right)_V dT+\left(\frac{V}{\alpha}\left(\frac{\partial\alpha}{\partial V}\right)_T-1\right)\frac{dV}{V}\right]. \tag{16}$$

(In eq (16) and the treatment that follows α_2 is written simply as α.) In this expression $\frac{1}{\alpha}\left(\frac{\partial\alpha}{\partial T}\right)_V$ is the pure temperature coefficient of polarizability, $\frac{V}{\alpha}\left(\frac{\partial\alpha}{\partial V}\right)_T$ is the pure volume coefficient of polarizability and the term $\left(\frac{V}{\alpha}\left(\frac{\partial\alpha}{\partial V}\right)_T-1\right)$ represents the total contribution of volume to the change in index. At present the term $\frac{V}{\alpha}\left(\frac{\partial\alpha}{\partial V}\right)_T$ can be evaluated from the constant temperature experiments. To obtain $\frac{1}{\alpha}\left(\frac{\partial\alpha}{\partial T}\right)_V$ the volume is taken to be a function of P and T. Then $dV=\left(\frac{\partial V}{\partial T}\right)_p dT+\left(\frac{\partial V}{\partial P}\right)_T dP$ and using

$\frac{1}{V}\left(\frac{dV}{dT}\right)_P=\gamma$ and $-\frac{1}{V}\left(\frac{dV}{dP}\right)_T=\beta$ eq (16) can be recast in the form

$$\frac{d\mu}{\mu}=\frac{\mu^2-1}{2\mu^2}\left[\frac{1}{\alpha}\left(\frac{\partial\alpha}{\partial T}\right)_V dT+(\gamma dT-\beta dP)\left(\frac{V}{\alpha}\left(\frac{\partial\alpha}{\partial V}\right)_T-1\right)\right]. \quad (17)$$

The value of $V/\alpha(\partial\alpha/\partial V)_T$ determined from eq (16) can be combined with available data on γ and on the temperature variation of index at constant pressure to derive values for $\frac{1}{\alpha}\left(\frac{\partial\alpha}{\partial T}\right)_V$. If the Lorentz-Lorenz equation is used the same expression will be obtained except that the factor $\frac{(\mu^2-1)(\mu^2+2)}{6\mu^2}$ replaces $\frac{\mu^2-1}{2\mu^2}$. This represents only a different scaling factor and in no way affects the relative magnitudes of the terms in the brackets which are of primary interest.

The calculated values of the terms of interest in eq (17) are given in table 4. The volume coefficient of polarizability is seen to be positive in all materials as expected. The positive sign means a decreased polarizability on compression and an increased polarizability on heating arising from the thermal expansion. In the alkali halides the numerical values of the coefficient increase as the anion packing increases. Values for MgO and diamond are the largest in the table, a result consistent with the large interatomic interactions presumed to occur in these materials. The small values for the glasses are to be expected from the smaller interatomic interactions expected for such open structures. The smallest value in the table applies to fused SiO_2 which has the most "open" structure. (The figures in table 4 have been rounded off and this conclusion is not obvious from the data given.) The total volume contribution is given in the next column and these values are generally negative except for diamond and MgO. It is noteworthy that in every case the volume coefficient of polarizability represents an appreciable portion of the total volume effect. It must be concluded that any relationship between index and density which ignores the effect of this contribution to the index can, at best, be only approximately correct. The next column tabulates the total volume term multiplied by the expansivity to permit comparison on an equal basis with values for the temperature coefficients of polarizability given in the next to last column. The temperature coefficients vary in both sign and size with the value apparently decreasing with packing in the alkali halides and being rather large positively in the glasses. There are insufficient data, however, to verify that these conclusions are typical. $\left(\frac{1}{\alpha}\frac{\partial\alpha}{\partial T}\right)_v$ may also be found by substituting values of $(d\mu/dT)_v$ in eq (16). These values are shown in the last column of table 4. Again it may be seen that there is fairly good agreement between corresponding values in the two columns of figures.

With the data given in table 4 it is instructive to consider the change in index resulting from two separate experiments, first an isothermal compression and second an isobaric temperature change. An isothermal compression produces an increase in index in KBr, NaCl, and all the glasses because the predominant effect is an increase in the number of scattering centers per unit volume. The decrease in polarizability with decreasing volume is less than the density effect in these materials. In LiF the index undergoes no change (to 1 kbar) because the density increase and the polarizability decrease exactly balance. In both MgO and diamond the index decreases on compression, the polarizability decrease overbalancing the density increase.

It is of interest to consider the effect of pressures higher than the 1 kbar used here. It appears likely that the volume coefficient of polarizability will increase as the volume decreases because the interatomic interactions should increase in magnitude. In the absence of other effects, this requires that at some elevated pressure the indices of the solids which increased with pressure at low pressures will pass through a maximum value and then decrease at still higher pressures. Further the decrease in index should accelerate at still higher pressures. The pressures required to produce these changes are not obvious but plans are being made to extend these measurements to 10 kb. In this range it appears very probable that a decrease might be expected for LiF. For the other alkali halides a reversal or a decrease in the rate of change of index with density might be expected.

An isobaric temperature increase results in a decrease in indices for KBr, NaCl, and LiF but an increase in indices for MgO, diamond and all the glasses. Although the resultant effects are similar for members in each of the two groups, there are various causes, and it is necessary to consider the details of the process. In KBr the volume expansion causes a decrease in index because the number of scattering centers per unit volume decreases. This effect is great enough to overbalance the increased polarizabilities arising from the increased volume and temperature. In NaCl the temperature coefficient acts in conjunction with the density effect to decrease the index. In LiF there is no total volume contribution and the decrease in index arises solely from the temperature coefficient of polarizability. In both MgO and diamond the index increases because the positive temperature and volume coefficients of polarizability overbalance the negative contribution from the decrease in the number of scatterers with the volume coefficient playing the major role. All the glasses show an increase in index because the comparatively large temperature coefficient of polarizability outweighs the negative effect of the total contribution of change in volume which is nearly as large.

The glasses require further consideration because it is known [8] that the index-temperature relationships of most optical glasses exhibit a minimum in index. The glass BSC–517/645 has such a minimum near −80 °C and although data on plate glass do not appear to be available a minimum may reasonably be expected at some temperature below 25 °C. On the other hand the index of refraction of fused SiO_2 is known to decrease monotonically with decreasing temperature [51,16] despite the reversal of its coefficient of thermal expansion. A reasonably sound explanation of these effects requires data on the two terms $\frac{v}{\alpha}\frac{\partial^2\alpha}{\partial V\partial T}$ and $\frac{1}{\alpha}\frac{\partial^2\alpha}{\partial T^2}$ which are apparently not available at this time but certain conclusions may be of interest until such data are obtained. It might be expected that for normal materials the polarizability arising from pure thermal agitation should decrease at higher temperatures because of increased interaction. However, it is not apparent whether $\frac{\partial\alpha}{\partial T}$ increases or decreases. Data by Radhakrishnan on LiF [49] show that $d\mu/dT$ becomes increasingly negative at higher temperatures. At room temperature this crystal shows a zero volume contribution, and, if it is assumed that this contribution is negligibly small at elevated temperatures, it may be concluded that $\frac{1}{\alpha}\left(\frac{\partial^2\alpha}{\partial T^2}\right)$ is negative. The glasses, however, are not normal solids so that it is not certain that they exhibit a similar behavior. It is known that $\frac{1}{V_0}\left(\frac{\partial^2 V}{\partial T^2}\right)_p$ is positive for the glasses and it is to be expected that $\frac{V}{\alpha}\left(\frac{\partial\alpha}{\partial V}\right)_T$ would increase somewhat at low temperatures because of the closer packing.

In table 4 it is observed that for BSC 517/645 and plate glass that the total volume contributions and the temperature coefficients of polarizability are of opposite signs but nearly equal in absolute values. If either one or both of those quantities changes slightly in the proper direction on lowering the temperature it is clear that the sign of $d\mu/dT$ will reverse. Since this reversal occurs below room temperature in BSC 517/645 and is expected in a similar temperature range for plate glass, it is probable that the temperature coefficient of polarizability decreases with decreasing temperature in these glasses in contrast with the opposite behavior observed in LiF. The situation in fused SiO_2 is somewhat different in that $\frac{1}{\alpha}\left(\frac{\partial\alpha}{\partial T}\right)_v$ is at least ten times as great as the total volume effect although they are of opposite sign. In this situation a small change in either quantity will be insufficient to reverse the sign of $d\mu/dT$. In addition when the expansivity of fused SiO_2 becomes negative both the volume term and $\frac{1}{\alpha}\frac{\partial\alpha}{\partial T}$ are positive and cannot cancel. It may be noted that the interpretation here is similar to that advanced by Prod'homme [52] for the anomalous temperature-index behavior in optical glasses but is somewhat more detailed. It appears that the positive value of $d\mu/dT$ in optical glasses arises from the comparatively large temperature coefficient of polarizability coupled with the low coefficient of thermal expansion. Since these quantities are not common to all glasses [52] it appears that they must be attributed to the silica network.

The authors are indebted to Harry B. Williams of the National Bureau of Standards for his great help in the preparation of specimens. The authors also thank Edward N. Farabaugh and William S. Brower of the National Bureau of Standards for their help in the orientation of the uniaxial crystals.

6. References

[1] E. Burstein and P. L. Smith, Proc. Ind. Acad. Sci **A28,** 377 (1948).
[2] H. Mueller, Phys. Rev. **47,** 947 (1935).
[3] H. Mueller, Physics **47,** No. 6, 179 (1935).
[4] H. N. Ritland, J. Am. Ceram. Soc. **38,** 86 (1955).
[5] C. Weir, S. Spinner, I. Malitson, and W. Rodney, J. Res. NBS **58** No. 4, 189 (1957) RP2751.
[6] R. S. Krishnan, Progress in Crystal Physics, Volume I (Interscience Publishers, New York, London, 1958).
[7] M. A. Jeppensen, J. Opt. Soc. Am. **48** No. 9, 629 (1958).
[8] (Thermal expansion and du/dT data on BSC 517/645)—F. A. Molby, J. Opt. Soc. Am. **39,** No. 7, 600 (1949).
[9] (du/dT data on diamond)—G. N. Ramachandran, Proc. Ind. Acad. Sci. **A25,** 266 (1947).
[10] R. M. Waxler and C. E. Weir, J. Res. NBS **67A** (Phys. and Chem.) No. 2, 163 (1963).
[11] (Refractive index and du/dT data on KBr)—R. J. Spindler and W. S. Rodney, J. Res. NBS **49** No. 4, 253 (1952) RP2301.
[12] (Refractive index data on NaCl)—F. Kohlrausch, Praktische Physik **II** B. G. Teubner, Leipzig, 528 (1943).
[13] (Refractive index data on LiF)—L. W. Tilton and E. K. Plyler, J. Res. NBS **47** No. 1, 25 (1951) RP2223.
[14] (Refractive index data on diamond)—F. Peter, Z. Phys. **15,** 358 (1923).
[15] (Refractive index data on MgO)—R. E. Stephens and I. H. Malitson, J. Res. NBS **49** No. 4, 249 (1952) RP2360.
[16] (Refractive index data on quartz, and thermal expansion and du/dT data on fused silica)—R. B. Sosman, The Properties of Silica, (Chem. Cat. Co. N.Y., 1927).
[17] (Refractive index data on Al_2O_3)—R. W. Kebler, Optical Properties of Synthetic Sapphire (Linde Co., N.Y.).
[18] (Refractive index data on fused silica)—J. B. Austin, Physics **3** 240 (1932).
[19] (Refractive index data on BSC 517/645)—R. E. Stephens and W. S. Rodney, J. Res. NBS, **52** No. 6, 303 (1954) RP2504.
[20] (Refractive index data on plate glass)—E. Hamilton, R. Waxler and J. M. Nivert, Jr., Optical Glass Technology Development, NBS Report No. 5440 (1957).
[21] (Compression data on KBr, NaCl and LiF)—P. W. Bridgman, The Physics of High Pressure (G. Bell & Sons, Ltd., London 1949).
[22] (Compression data on quartz)—P. W. Bridgman, Am. J. Sci. Fifth series **15** No. 88, 287 (1928).
[23] (Compression data on Al_2O_3)—P. W. Bridgman, Proc. Am. Acad. Arts Sci. **68,** 27 (1932).
[24] (Compression data on fused silica)—P. W. Bridgman, Am. J. Sci. **10,** 359 (1925).
[25] (Compression data on diamond and plate glass)—L. H. Adams and E. D. Williamson, J. Frank Inst. **195,** 475 (1923).

[26] (Compression data on MgO)—C. E. Weir, J. Res. NBS **56** No. 4, 187 (1956) RP2666.
[27] S. Spinner, J. Am. Cer. Soe. **37** [5], 229 (1954).
[28] F. A. Jenkins and H. E. White, Fundamentals of Opties (MeGraw-Hill Book Company, Ine., New York, Toronto, London, 1950).
[29] E. S. Jog, Jour. Ind. Inst. Sei. **39A,** 101 (1957).
[30] R. M. Waxler, C. E. Weir, and H. W. Sehamp, Jr., J. Res. NBS **68A** (Phys. and Chem.) No. 5, 489 (1964).
[31] F. Poekels, Ann. Phys. **37,** 151 (1889).
[32] D. F. Gibbs and G. J. Hill, Phil. Mag. **8,** 367 (1963).
[33] C. Sehaefer and H. Nassenstein, Z. Naturforseh. [A] **8,** 90 (1953).
[34] K. Vedam, Proe. Ind. Aead. Sci. [A] **31,** 450 (1950).
[35] F. Pockels, Ann. Phys. Chem. **37** (4), 269 (1889).
[36] R. M. Waxler and A. Napolitano, J. Res. NBS **59** No. 2, 121 (1957) RP2779.
[37] R. M. Waxler, Glass Ind. **34,** 258 (1953).
[38] S. Spinner, personal eommunication.
[39] F. Poekels, Ann. d. Phys. **7,** 745 (1902).
[40] K. S. Krishnan and S. K. Roy, Phil. Mag. **43,** 1000 (1952).
[41] G. N. Ramaehandran and T. Radhakrishnan, Phil. Mag. **43,** 317 (1952).
[42] (Thermal expansion data on diamond)—R. S. Krishnan, Proc. Indian Aead. Sei. **A24,** 33 (1946).
[43] (Thermal expansion data on KBr)—R. Srinivasan, J. Indian Inst. Sei. **A37,** 232 (1955).
[44] (Thermal expansion data on NaCl)—R. M. Buffington and W. M. Latimer, J. Amer. Chem. Soe. **48,** 2305 (1926).
[45] (Thermal expansion data on LiF)—A. Eueken and W. Dannöhl, Z. Eleetroehem. **40,** 814 (1934).
[46] (Thermal expansion data on MgO)—M. Durand, Physies **7,** 297 (1936).
[47] (Thermal expansion data on fused siliea)—K. Seheel and W. Heuse, Verh. Deutsch. Physik. Ges. **16** 1 (1914).
[48] (du/dT data on NaCl)—F. J. Mieheli, Areh. Se. Phys. Nat. **13,** 217 (1902).
[49] (du/dT data on LiF)—T. Radhakrishnan, Proe. Indian Aead. Sci. **A31,** 224 (1950).
[50] (du/dT data on MgO)—T. Radhakrishnan, Proe. Indian Aead. Sei. **A33,** 22 (1951).
[51] (du/dT data on fused siliea)—F. Rinne, Neues Jahrb. Mineral. Beil. Bd. **39,** 388 (1914).
[52] L. Prod'homme, Phys. and Chem. of Glasses **1** No. 4, 119 (1960)

(Paper 69A4–352)

Reprinted from APPLIED OPTICS, Vol. 4, page 21, January 1965

The Performance of Lenses Made from Inhomogeneous Glasses

Fred W. Rosberry

The performance of lenses made from glass elements containing defects such as bubbles, streaks, striae, etc., has been measured and an attempt made to compare the effect of the various defects. Four defective lenses and a good or control lens were used. The performance in the image plane was determined by four different evaluation procedures. A summary evaluation showed that the lens containing striae, although the most difficult defect to detect, produced the lowest quality image.

I. Introduction

In the interest of improving lens imagery three areas for development are available: lens design, optical materials, and lens fabrication. From the available literature, it is evident that major emphasis is placed upon lens design as the reports of investigations into this subject are more numerous than those dealing with either of the other two areas. It is probable that the difficulty of making measurements that permit the establishment of definite meaningful criteria for standards of quality for transparent optical media has inhibited research in this area. This paper presents an approach to the problem that might prove of use in the evaluation of the effect of inhomogeneities in glass upon the quality of images formed by lenses made from such glass. In view of the ambiguousness of such terms as "defective glasses" and "fuzzy images", the best that can be done in an investigation of this kind is to evaluate the imagery formed by lenses made from a few samples of poor glass. Several methods of analysis are employed. These include consideration of interferograms, sine wave response data, limit of resolution measurements, and viewer reaction to photographs of a given subject scene for each of the selected lenses.

II. Imperfections in Material

Defects in glasses are identified by various terms depending upon the type, but all are characterized by areas containing a rapid change in the index of refraction of the optical material. Small areas of large index differences such as seeds, streaks, and feathers are usually caused during the melting and molding of the glass by improper mixing, stirring, impurities in the ingredients, or by contamination from the surfaces of the pot. Large areas of small index differences are usually caused during the annealing schedule where small temperature gradients exist throughout the sample. The material defects in four glasses used in this investigation were striae, bubbles, streaks and feathers, and scratches. These glasses were each made into a lens with two components, and the described defects appeared in one or both components. The focal length was 30 cm and aperture 2.5 cm. A fifth lens of the same design and good quality glass was used as a control. It is hoped that this constitutes a set for investigation with glass quality and surface defects (scratches) as the dominating variable. It is fully realized that it is difficult to describe the severity of the defects in quantitative terms that can be understood without ambiguity. Also while lenses made from defective glass are listed as containing only one defect they may contain other defects in minor amounts.

Lens A (bubbles) had one bubble 0.5 mm in diameter and several others of smaller diameter. It also contained an inclusion of opaque foreign material about 3 mm long and 0.5 mm wide. A slight irregularity in the fringe pattern (Fig. 1) suggests that there is a variation in index of refraction adjacent to this inclusion.

Lens B (scratches) was deliberately scratched with a diamond or other hard point on two of the four surfaces. The maximum width of the scratch is about 0.8 mm. The scratch pattern is clearly shown in Fig. 1.

Lens C (control) had no defect in the glass that could be observed by the test methods used.

Lens D (striae) contained a heavy band of striae about 5 mm wide running almost diametrically across the lens. A second narrow band of striae is visible at an angle of about 40° to the main band. The glass used in this lens would be judged to be comparable in

The author is with the National Bureau of Standards, Washington, D. C.

Received 3 March 1964.

This work was supported in part by the Advanced Research Project Agency of the Department of Defense.

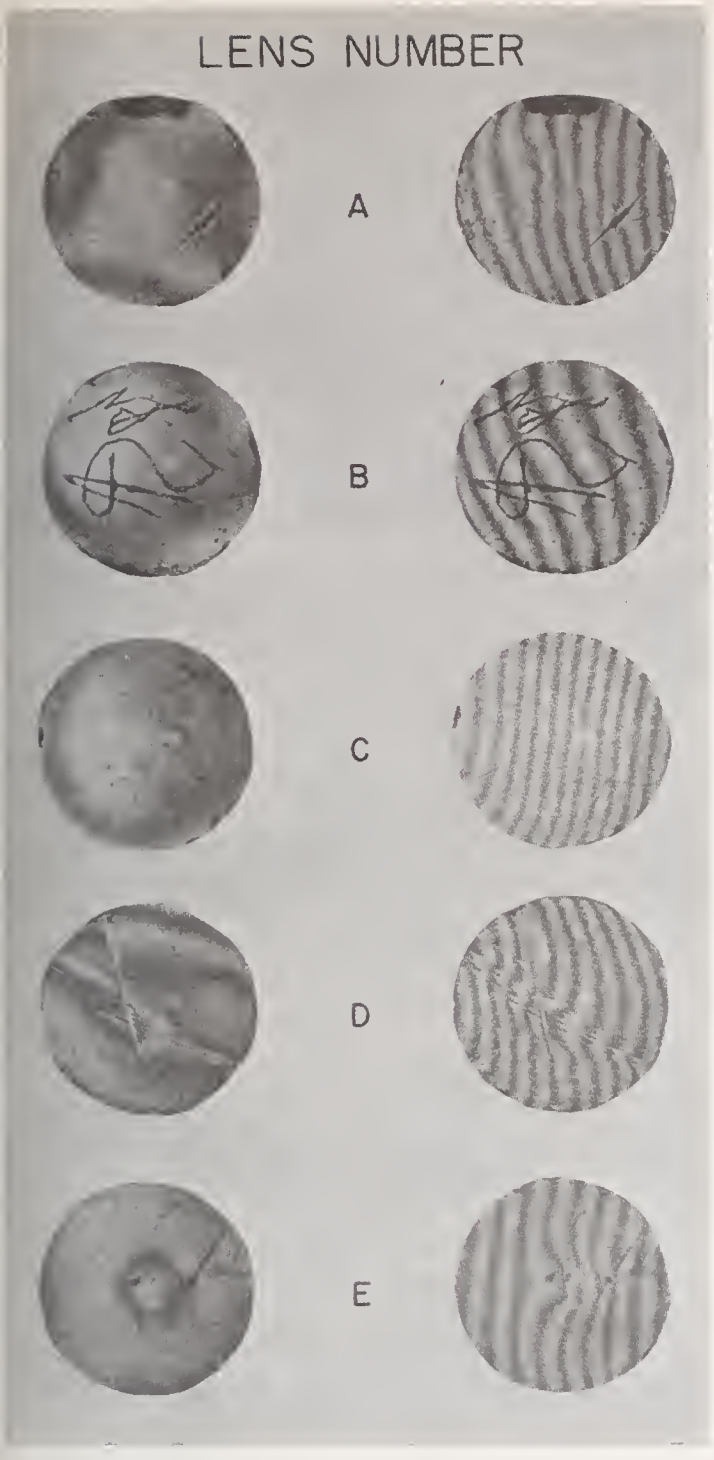


Fig. 1. Interferograms of the five lenses investigated, photographed on a Twyman-Green interferometer. Vertical column at left illustrates the minimum fringe patterns. Right-hand column shows fringe patterns with a small amount of tilt introduced by the interferometer.

quality to D grade optical glass as defined in Joint Army-Navy Specification JAN-G-174, 30 January 1954, Glass, Optical.

Lens E (streaks and feathers) contained a defect that was apparently caused by folding a surface layer into the glass when the lens was molded. The defect included opaque material and a layer of fine bubbles as well as one bubble 0.5-mm diam and one 0.2-mm diam.

III. Methods of Measurement

A. Wavefront Determination

A perfect lens placed in a collimated beam will have a spherical wavefront on the imaging side of the lens. The amount a wavefront departs from the sphere is a measure of the defects in the lens. The wavefront can best be examined interferometrically where the entire front can be observed or photographed at once. Departures from sphericity are indicated by fringes in the field of the interferometer. Any area of the front deviating from a sphere by a radial distance of one-half a wavelength is shown by a dark fringe. The Twyman-Green type of interferometer was found to be quite satisfactory for this task. The light inside the interferometer is parallel. A positive lens introduced into the beam makes it convergent and to return it to a parallel beam one of the plane mirrors is replaced by a convex spherical one. When the center of the reflecting sphere is coincident with the focal point of the lens, the converging beam is reflected back upon itself and reemerges from the test lens again as a reasonably parallel beam traveling in the reverse direction. The fringe pattern produced when this beam interferes with the reference beam indicates the position of minimum fringes and any adjustment will put more fringes into the field. Changes in the fringe pattern can be accomplished by changing the angle of the reference mirror in the comparison leg of the interferometer or by separating the center of curvature of the sphere and the focal point of the lens under investigation. Conversion of this type of fringe pattern to wavefront is described by Twyman.[1]

Two wavefront presentations will suffice for this reporting: one of minimum fringes and one with a small amount of inclination between the interfering beams put in by the reference mirror. The minimum fringe pictures used as a criterion (Fig. 1) show the workmanship to be quite uniform on all except lens E. This lens contains streaks and feathers and there also seems to be a dimple about one fringe deep in the center of the wavefront caused by defects other than the faults in the glass.

B. Sine Wave Response

Sine wave response represents one of the numerous methods of assaying the image-producing qualities of a lens. The sine wave response or contrast transfer function of the lenses was determined on equipment described earlier.[2] Data were taken only on axial positions, and the focal positions were determined by using the plane yielding the highest resolution. This choice may give rise to argument[3] but it was made only as a means of using the same criterion every time to determine a focal plane. The sine wave response curve gives not only an approximation of the limit of resolution, if carried to near zero response, but also shows the degree of reproduction of objects in the image plane. Objects do not reproduce perfectly in the image out to a certain value of lines per mm and then stop with a sharp cutoff. As an example, a lens having

a limit of resolution of 90 lines/mm does not resolve 80 lines as clearly as it does 30 lines. The rate of degradation is gradual and varies from one lens to another. It is common to observe response-function curves indicating similar values for limit of resolution and greatly differing values of response for intervening spatial frequencies.

Each of the lenses in the set was examined on the sine wave bench. A series of signals of progressively higher spatial frequency was projected through the lens, and the resultant luminous intensities of the lines and spaces in the image were recorded by a photocell on a strip chart recorder. As the projected line space frequencies increased the luminosity differences in the image reduced until near the limit of resolution the signal was reduced to the noise level of the system. A curve was plotted for each lens showing the response as a function of increasing spatial frequency. The curves were not plotted beyond a spatial frequency of 75 lines/mm because this represented the highest value on the target.

C. Limit of Resolution

The limit of resolution for the five lenses was determined by a conventional photographic method using a three-line square wave target. Targets are designed with different rates of progression from one pattern to another. The longest ranges of spatial frequency values are covered by targets with large steps between patterns. This type of target also gives the best repeatability because of the large differences between the steps. The target used in this investigation was the U.S.A.F. 1951 target using a length to width ratio equal to 5 to 1 with rather small differences ($\sqrt{2}$) between steps in order to emphasize the resolution differences between these lenses. The lenses were set on a lens bench and focused on the target visually. Photographs were made at this position and at two nearby positions on each side in order to bracket the best focal position. The highest resolution that could be read with a microscope from the five focal positions was called the limit of resolution for that particular lens on axis. The maximum off-axis position for this test arrangement was only 3.5°. Photographs were recorded on Kodak High Resolution Plate.

D. Scene Analysis

Another method of observing the image-producing qualities of these lenses was to examine pictorial scenes made with each. This is undoubtedly the least precise and most subjective of the test methods employed. However, it is the method by which many lenses are rated.

A 35-mm single lens reflex camera was fitted with an adapter to accommodate the long focal length test lenses. The focal plane shutter in the camera body was used to determine exposures with the various lenses. The camera with back removed was mounted in the nodal slide of the sine wave response bench. Repeated scans of a high-spatial frequency pattern were made at various focal settings and the best focus was located by the setting giving the highest response. The film plane

Fig. 2. Photograph of the scene used for subjective rating of lens performance. Four areas on the print were used for examination of detail.

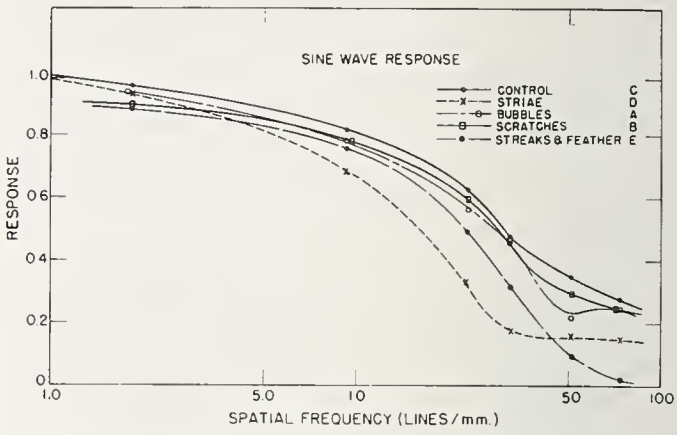


Fig. 3. Sine wave response curves determined from the five lenses investigated.

was established from the camera body by the guiding strips which control the film location. The best aerial image was focused to fall in the same plane. The camera was then removed from the nodal slide, the film threaded, and the back replaced. Photos of the scene (Fig. 2) were made from the tripod-mounted camera. This process was repeated for each test lens. Approximately one hour was required to complete this process for each lens, thus making the lighting angle of the sun different in the successive scenes which was a handicap when the photographs were evaluated. The distance between camera and principal subject in the test scene was about 1950 m. This was virtually equal to a focus on infinity. The developed negatives were printed at a constant enlargement and then examined by several observers who were asked to list the prints in order from best to worst.

IV. Summary of Results

Analysis of the interferograms of the wavefronts shown in Fig. 1 indicates that defects such as bubbles (A), scratches (B), and streaks and feathers (E) are best observed by shadowgraphing methods. The

Table I. Summary of Results as Determined by the Five Methods of Image Analysis

Relative rating	Wavefront		Sine wave response	Limit of resolution	Scene analysis	Averages
	Minimum fringes	Tilt fringes				
Best	C	C	C	B or C or E	C	C (control)
Better	B	A	B	B or C or E	A	B (scratches)
Good	A	B	A	B or C or E	E	A (bubbles)
Fair	D	D	D	A or D	D	D (striae)
Poor	E	E	E	A or D	B	E (streaks & feathers)

change in index between defect and the bulk of the glass is so abrupt that the change in fringe order cannot be counted; also the areas of the defects are too small for a fringe pattern to be seen within the defect. The interferogram of the lens with striae (D) shows quite well deformations in the wavefront caused by the striae in the glass. The irregularities of the fringes in the center of lens E are not caused by the glass defects but by surface curves on the elements and will not be discussed here. Examination of the ten interferograms, with allowances made for the aforementioned irregularities of lens E, shows lens C to be the best and lens D or E the lowest quality.

The sine wave response curves shown in Fig. 3 are the results of measurements of the amount an optical sine wave signal in luminous intensity is attenuated by a single pass through each of the lenses in this investigation. The area under each curve represents useful information; therefore, the curve indicating the highest response for the greatest amount of recordable frequencies represents the lens reproducing the most information. Examination of the curves with this as a criterion shows lens C to be the best and, out to spatial frequency of 45 lines/mm, lens D the worst. Higher spatial frequencies indicate lens E as giving the lowest response.

The limit of resolution results as read from the photographic plate divided the lenses into two groups—one group of 3 lenses resolving about 123 lines/mm and the second group of 2 resolving 111 lines/mm. The control lens C was in the better group and lens D with striae in the lower group. This agreed with results indicated by the sine wave response curves except for the unpredictable resolution of lens E.

The prints of the pictorial scene were rated by four observers who were asked to judge each print by the quality in each of four specified areas on the print, making a total of 16 votes for each print. A typical example of the voting gave lens A 4 votes for first place, 8 for second, 3 for third, none for fourth, and 1 for fifth, thus putting it in second place. The over-all tally of this method put lenses C and A as best, and D and B as the lowest quality.

The results of all the evaluation methods are shown in Table I where another average is taken of the various methods and a final rating shown in the last column. Very little weight was put on the results obtained with lens E because of defects in the surfaces as well as the glass.

V. Conclusions

It would have been an improvement in the investigation if it had been possible to have made some observations of the homogeneity of the glass used for lens D, before it was made into the lens, containing striae. The other samples with bubbles and streaks in the glass are best observed by a shadowgraph technique or are amply well seen in the interferograms. The fringe patterns show that in spite of the manufacturers' efforts to finish the lenses the same except for the material differences there still exist variations due to manufacture. The fringe photographs also show a set of ghost fringes probably caused by the air space between the elements. It was not the intention of this paper to compare the various methods of image analysis, but only to use all of them hoping that the average would be nearer a true rating than any one method would be. It is apparent from these data that striae content, although more difficult to detect, is probably one of the more influential of the glass faults.

References

1. F. Twyman, *Prism and Lens Making* (Hilger & Watts, London, 2nd ed., 1952), Chap. 12, p. 434.
2. Fred W. Rosberry, Photogrammetric Eng. **27**, 155 (1961).
3. F. C. Higgins and L. A. Jones, J. Soc. Motion Picture Engrs. **58**, 277 (1952).

Reprinted from APPLIED OPTICS, Vol. 5, page 961, June 1966

The Measurement of Homogeneity of Optical Materials in the Visible and Near Infrared

F. W. Rosberry

A procedure is described for determining the inhomogeneities of a sample of optical material in terms of small changes in index of refraction at discrete points over the measured surface. The method is used in the visible- and near-ir regions of the spectrum. An He–Ne gas laser source was used for the ir measurements. The variation in index was mapped with contour lines enclosing areas of similar index variations. The largest index change in 1-cm distance was noted and recorded as the maximum variation per centimeter. Over twenty-five different samples of six different materials were examined. The results are presented in a chart indicating the range of maximum refractive index variations of the samples observed. Data are shown to illustrate the possibility of determining index values to higher precision than the material justifies.

I. Introduction

The availability of new materials for optical applications is constantly increasing. A survey of optical material catalogs yields a wealth of information concerning transmission and absorption vs wavelength, refractive index and dispersion vs wavelength, density, melting point, solubility in water, and other properties. One quality that is usually only lightly touched on or not even mentioned is uniformity of structure or homogeneity. When mentioned it is often referred to as a *high degree of homogeneity* or *extremely homogeneous.* These unquantitative terms are justified by the apparent difficulty of putting a numerical value on this characteristic.

Homogeneity is a quality that depends on the purity of a material as well as the residual stress within.[1] Design requirements are putting an increased demand on the precision of measurement of other optical qualities such as transmittance and refractive index. As more precise values for these properties are obtained, the uniformity of structure or homogeneity of the material becomes of increasing importance. The increase in precision of measurement of these qualities is justified by the fact that it represents an integration of the variations over the area observed by the measuring instruments. By interferometric means, the homogeneity of a material can be measured in terms of index of refraction; relative changes in the sixth decimal place are readily detected for extremely small areas. The interferometer is an ideal tool for measuring small changes in refractive index when a nominal index value to about four places beyond the decimal is available. This paper describes a method introduced by Twyman,[2] improved by Tilton *et al.*,[3] and herein advanced by establishing a criterion for applying a numerical value to homogeneity. It also extends the useful range of measurement from the visible into the near ir[4] by utilization of an ir He–Ne gas laser as a source, and image converter equipment to make observations.

II. The Measurement of Homogeneity

The degree of homogeneity in a sample can vary greatly. Low quality material with large areas of striae can be observed with the unaided eye. Smaller areas can be seen only by using a shadowgraph or schlieren technique and graded by subjective comparison with shadowgraphs of the standards for striae in optical glass.* A shadowgraph of one set of standards is shown in Fig. 1. Strain can be detected by observing the sample through a polariscope. These defects represent large changes in refractive index occurring over short distances. Higher quality material will have more gradual index gradients that are not detectable by the above methods.

Figure 2 displays the order of preliminary test procedures used in this laboratory. Four samples, two of glass and two fused silica, are labeled A, B, C, and D.

The author is with the National Bureau of Standards, Washington, D.C. 20234.

Received 22 September 1965.

This work was supported, in part, by the Advanced Research Projects Agency of the Department of Defense.

* See Military Specification of Optical Glass MIL-G-174A.

Fig. 1. Shadowgraph of the standards for optical glass striae grade. Grade D is at left and Grade A at right. Grade B has a small surface scratch near center. Striae can be seen in Grade A on original photograph. These standards are 10.16 cm thick.

Sample A failed the visual test because of the gross lack of resolution in the upper center of the sample. The other samples did not show such pronounced degradation of resolution. The second test is made with the sample in polarized light. Sample B showed a strong strain pattern and was therefore eliminated from further investigation. Sample C showed a slight pattern in this test that was barely visible on the photographs. The third test, the shadowgraph test, showed the location of several areas of striae in Sample C which could be subjectively compared with the standards shown in Fig. 1. The shadowgraph of Sample D showed a fairly high density of inclusions of opaque foreign material which may or may not exclude this sample, also, from further testing. The fourth and last preliminary test is interferometric observation of the sample. A countable number of reasonably straight fringes were observed. It was found that, if the dimensional tolerances conform to those set forth in Fig. 3, the sample was suitable for complete examination.

Small refractive index gradients can be located by the interferometric procedure. A Twyman–Green interferometer is set up with a grid and clamping arrangement to hold the sample perpendicular to the beam in one leg. The grid is used to establish reference points at which measurements can be made. A grid wire spacing of 1 cm has been found convenient. Samples of optical material to be observed on the interferometer have to be finished to closer surface tolerances than do shadowgraph samples. It has been found that surfaces flat to within two fringes of green mercury light ($\lambda = 0.5461\ \mu$) are suitable, and a slight wedge between the faces of about 3 μ to 6 μ is adequate for a sample 50 mm in diam. The exact wedge is dependent on the nominal index of refraction of the material and on the wavelength to be used. The wedge is a deliberately generated departure of the appropriate magnitude for obtaining a convenient number of fringes in one mode of observation. The shape of the sample is not important. Some of our circular disks had two flat surfaces cut and polished on their edges, with an angle of 60° between them in order to form a thin prism. These windows made it possible to obtain a spectrometer determination of the nominal index. If the interferometer is adjusted for zero field fringes and the sample is inserted in the holder with one face close to the grid, the number of fringes that can be observed in transmitted light between any two local points 1 and 2 on the sample is derived from Twyman's expression:

$$\tfrac{1}{2}\, m_t\lambda = (n - 1)\Delta t + t\Delta n. \tag{1}$$

Solving for m_t and assuming that $\Delta n = \Delta n_1 - \Delta n_2$ and $\Delta t = \Delta t_1 - \Delta t_2$

$$m_t = 2/\lambda[(n_0 - 1)(\Delta t_1 - \Delta t_2) + t_0(\Delta n_1 - \Delta n_2)], \tag{2}$$

where m_t is the fringe count, λ is the source wavelength, t_0 is the nominal thickness, n_0 is the nominal refractive index, and Δt_1 and Δn_1 are local variations in thickness and refractive index averaged through the sample thickness at these points. If the end mirrors in each arm of the interferometer are covered, a pattern of fringes will be seen within the sample only, as a result of reflections from the two faces; the number of fringes is given by the expression:

$$\tfrac{1}{2} m_r\lambda = n\Delta t + t\Delta n. \tag{3}$$

Solving for m_r as above

$$m_r = 2/\lambda[t_0(\Delta n_1 - \Delta n_2) + n_0(\Delta t_1 - \Delta t_2)]. \tag{4}$$

Solving Eqs. (1) and (3) simultaneously for Δn: $\Delta n = \lambda/2t[n_0 m_t - (n_0 - 1)m_r]$. The same two equations solved for Δt are $\Delta t = \lambda/2(m_r - m_t)$.

The direction of increasing order of the fringe pattern must be determined during setup. It was found advantageous to photograph the fringe patterns. The photographs could then be enlarged, making fringe counting easier and also providing a permanent record. It was also noticed that during the transmission mode of observation, with interferometer field around the sample clear of fringes, there was a tendency for one fringe to be drifting around in the field. The drift was caused by slight air currents or vibrations. This unstable condition was greatly improved by adjusting an air wedge in the field by a small inclination of one instrument mirror. However, the direction of increasing order of the fringe pattern of this wedge must also be known, and it was found convenient to align the air wedge in the same direction as the sample wedge. In the reflection mode of observation, the number of fringes and their orientation are characteristic of the sample. They are determined by the inclination of the faces to each other and the degree of inhomogeneity of the material. The only adjustment that can be made of this pattern is by rotation of the sample in its holder, and a corresponding rotation of the pattern will result. It was found advisable to make the photographs in rapid succession in order to minimize any chance for

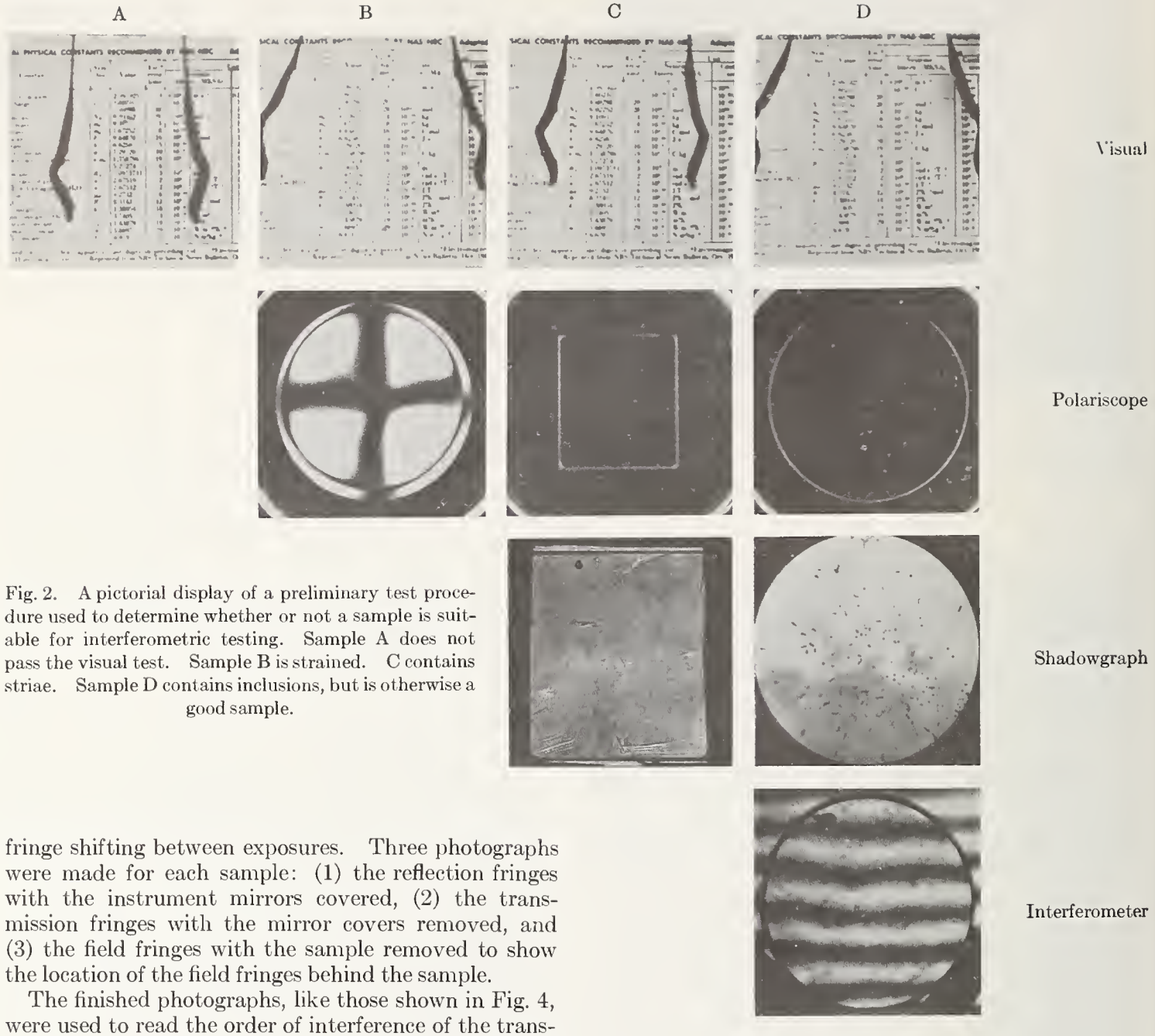


Fig. 2. A pictorial display of a preliminary test procedure used to determine whether or not a sample is suitable for interferometric testing. Sample A does not pass the visual test. Sample B is strained. C contains striae. Sample D contains inclusions, but is otherwise a good sample.

fringe shifting between exposures. Three photographs were made for each sample: (1) the reflection fringes with the instrument mirrors covered, (2) the transmission fringes with the mirror covers removed, and (3) the field fringes with the sample removed to show the location of the field fringes behind the sample.

The finished photographs, like those shown in Fig. 4, were used to read the order of interference of the transmission fringes, reflection fringes, and field fringes at each of the grid reference points. The grid intersections were numbered to avoid confusion. The fringes of the air wedge that were added in the transmission mode of observation must be subtracted when fringe values for transmission corresponding to zero field fringes are desired. The accuracy of readings at the grid points have a definite bearing on the precision of Δn values. Accurate results in the sixth and seventh decimal place required that grid points be read from the negatives on a comparator, while less accuracy and greater speed are attained by reading by eye from the enlarged photographic prints. The order of interference at each grid point and the nominal values for index and thickness are now applied in the Δn formula to give Δn values at each of these points over the face of the sample. A scale drawing of the sample is made, and values for Δn are written in at their proper locations. Contour lines are drawn locating areas of similar index, as shown in Fig. 5. In an effort to assign a number to the homogeneity measurement, the contour lines were used to choose the area in which the greatest change in Δn values occurs between two adjacent grid points; this difference was recorded as the maximum variation per centimeter (maximum variation/centimeter). A second value was also recorded to indicate maximum difference between any two points in the entire sample from the lowest to the highest Δn values. The variation per centimeter is not dependent on sample size, as is the second value. The maximum variation-per-centimeter criterion for determining homogeneity rates of the samples by their areas of greatest change in index thus gives the largest variation possible, with the average variation throughout the sample

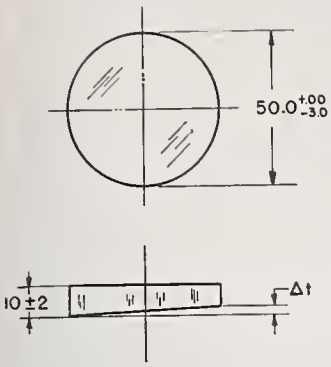


Fig. 3. A drawing showing the dimensional requirements for a sample to be used in an interferometric test. The small wedge between faces provides the necessary number of reflection fringes. The measurements are in millimeters. Both surfaces are polished and flat to within two fringes. If index refraction is below 1.5, Δt is 5 μ to 6 μ. If index of refraction is between 1.5 and 2.0, Δt is 4 μ to 5 μ. If index of refraction is between 2.0 and 2.5, Δt is 3.5 μ to 4.5 μ. If index of refraction is 3.4 (silicon), Δt is 3.0 μ.

lower or better than this value. Poor material has a high variation value, good material a low variation, and a perfectly homogeneous material no refractive index change throughout the sample. A value nearer to average might be obtained by averaging the best and worst variations within a sample. Then one gets into frequency of occurrence of higher- or lower-than-average variations, and this, in turn, leads into a problem in statistics. Therefore, in the interest of expediency, the maximum variation-per-centimeter figure was chosen and a note mentioned that all other variation values are less than this.

III. Infrared = Transmitting Materials

The examination of samples that do not transmit in the visible region of the spectrum presents a few added problems to this procedure. A monochromatic source of ir energy is necessary. Fringe patterns have been produced with the 1.06-μ line of mercury and also with the 1.08-μ line of helium, but the intensity was quite low. Best results were obtained by using a He–Ne gas laser that emits energy at 1.15 μ wavelength. With these ir sources an image converter must be used to adjust visually the interferometer and record the three fringe photographs. Photographs can be made of the visible fluorescent image on the screen of the image converter, or direct ir photography of the ir image can be accomplished by using hypersensitized Class 1-Z spectroscopic plates. The use of image converter and camera with film (such as Plus-X, Tri-X, or Royal-X) for use in the visible range made possible shorter exposures, but some distortion and lower resolution was noticed. The direct ir photography, although requiring longer exposures, gave better end results.

The use of a longer source wavelength will force the use of electronic image conversion equipment as the present laser line is at the wavelength limit of available photographic emulsions. In use, the laser source provided some advantages and some disadvantages. One advantage was that little attention was needed in

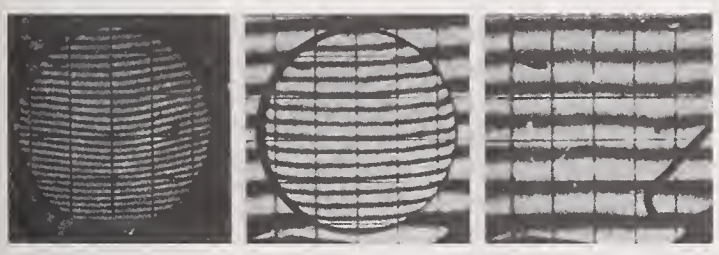


Fig. 4. A sample set of fringe pictures from a disk of fused silica. The picture at left shows the reflection fringes, center picture shows the transmission and field fringes, and picture at right shows only field fringes.

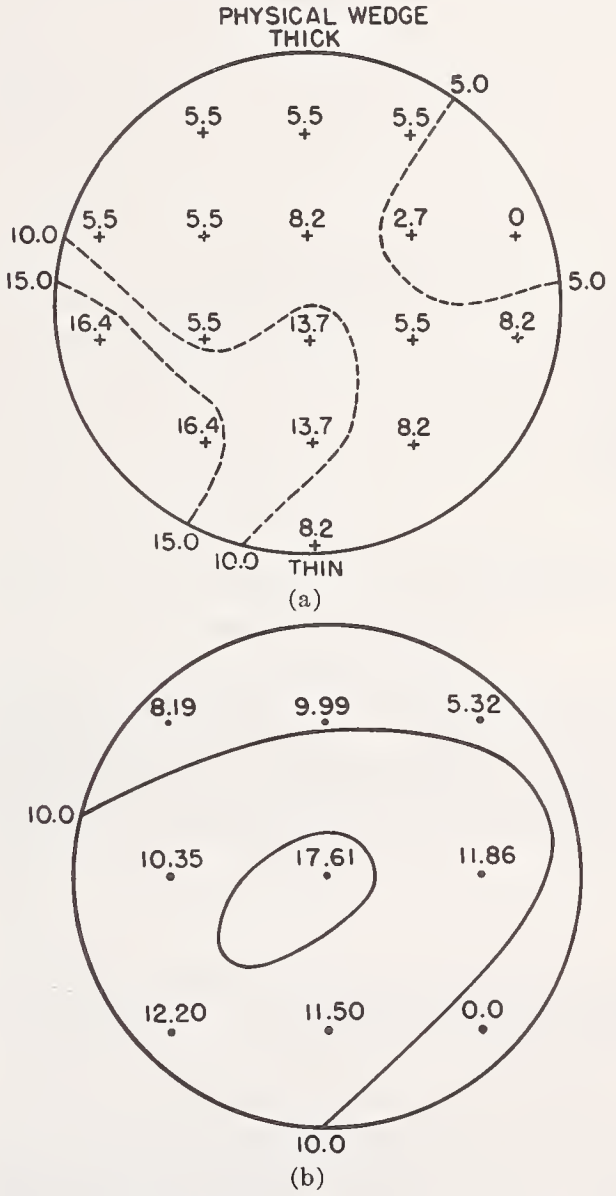


Fig. 5. Contour maps of two samples showing index variation values at points corresponding to the intersections of the grid wires. The upper sample is fused silica; the lower is silicon. The contour lines are drawn along lines of equal index variation values. (a) Value for variation of index of refraction $\Delta n \times 10^{-6}$. Maximum variation per centimeter is 10.9×10^{-6}; total variation is 16.4×10^{-6}. (b) Silicon single crystal 1:1:1 P type $\Delta n \times 10^{-5}$. Maximum variation per centimeter is 12.4×10^{-5}; total variation is 17.6×10^{-5}.

maintaining equal path lengths in the interferometer. The reflection mode of observation necessitates a path difference equal to twice the nominal sample thickness that cannot be easily compensated for. The conventional lamps limited sample thickness to about 5 cm or less while the laser has easily produced fringe patterns at a 15 cm sample thickness. The most troublesome disadvantage with the laser source was in the mode pattern that it projected; this was superimposed on all the interferograms and occasionally made fringe reading at grid points difficult.

The measurement of homogeneity of silicon presented a marginal problem with the equipment described in this paper. The threshold of transmission of silicon is close to the cutoff wavelength of the image converter equipment and to the laser source wavelength. This

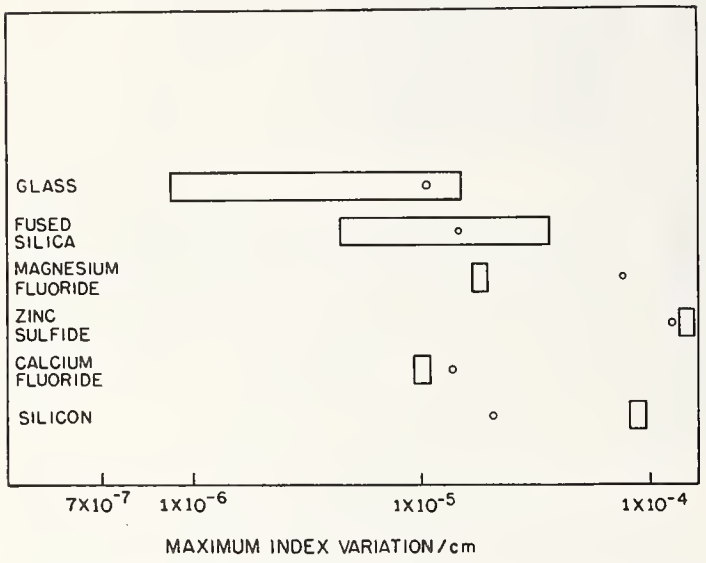


Fig. 6. Chart showing a method of intercomparing samples of different materials. Bottom four materials are represented by only one sample of each kind. The sample of zinc sulfide was the thinnest sample tested and therefore benefited from the less-smoothing effect that added thickness gives. The small circles indicate values of (maximum variation/centimeter)/$n-1$ from Table I.

probably contributed toward some signal attenuation. However, Δn determinations of one sample of silicon were made and are shown in Fig. 5.

IV. Results

If a number of similar-thickness samples of a given material from different sources are examined, a fairly reliable average maximum variation-per-centimeter value can be obtained for this material. Several materials of different thickness were measured; a chart indicating the range of their maximum variations per centimeter is shown in Fig. 6. The values for glass and fused silica are represented by ten different samples of each, while the values for the other materials are for single samples, each of which differed in size and thickness from the others. This type of chart which compares Δn only raises the question of whether a Δn comparison is fair or whether Δn relative to n should be used. The latter method tends to improve the apparent homogeneity of high index materials relative to those of lower index. Table I shows the change of the average values of the materials if the relationship (maximum variation/centimeter)/$(n-1)$ is used. The open dots locate the values of (maximum variation/centimeter)/$(n-1)$ in Fig. 6. The calcium fluoride sample had a low Δn gradient but contained an excessive number of small voids or bubbles distributed evenly throughout the sample, as shown in the shadowgraph in Fig. 7. This sample is a good example of a condition that occasionally occurs in which a noticeable concen-

Fig. 7. A shadowgraph of a sample of polycrystalline calcium fluoride showing the high concentration of bubbles or voids in this sample.

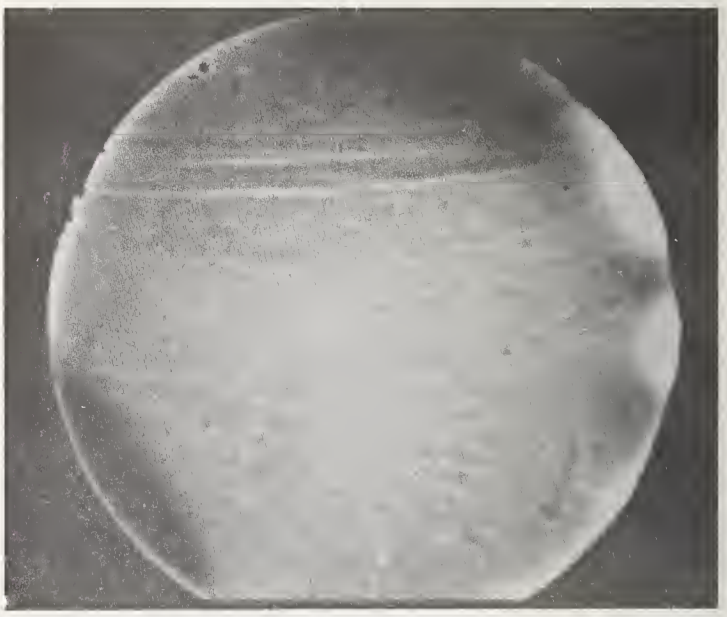


Fig. 8. A shadowgraph of silicon made with an image converter. Horizontal streak in upper half of sample indicates an area of striae.

Table I. Average Maximum Values of Index of Refraction Variation/Centimeter Expressed Relative to n

Material	Nominal n^a	Average maximum variation/ centimeter	Maximum variation/ centimeter $n-1$
optical glass	1.5174	0.6×10^{-5}	1.2×10^{-5}
fused silica	1.4582	1.2×10^{-5}	2.6×10^{-5}
magnesium fluoride	1.3768	3.7×10^{-5}	9.8×10^{-5}
zinc sulfide	2.2867	25.7×10^{-5}	20.0×10^{-5}
calcium fluoride	1.4282	1.0×10^{-5}	2.3×10^{-5}
silicon	3.5120	10.4×10^{-5}	4.1×10^{-5}

[a] Nominal index of refraction at the wavelength used for the interferometric test.

Fig. 9. Shadowgraph through a glass prism on the same path used by the spectrometer when a nominal value for index was determined.

tration of inert material or voids are present in an otherwise homogeneous optical medium. This condition is sometimes not detectable on an interferogram because the inclusions are so small that they do not affect the fringes. The shadowgraph usually indicates the presence of this condition and therefore justifies the time spent in making it. It is probable that this condition will increase scattering and therefore reduce transmittance more than that for a clear crystal; thus, the shadowgraph can give a clue to the amount of scattering and transmittance that can be expected.

Figure 8 shows a shadowgraph of a sample of silicon containing a noticeable area of striae. This picture was made through the image converter; the light streaks around the edge were in the image converter, not the sample. The limiting aperture in this photograph is the image tube. The sample was slightly larger.

One sample examined in this investigation was a 60° glass prism with the ends polished flat, according to the specifications of Fig. 3. The shadowgraph shown in Fig. 9 includes the same path used by the spectrometer for index measurements reported to the fifth decimal place. The index value was undoubtedly correct as an average value, but subsequent inhomogeneity measurements showed the index of this material to be varying by as much as eight in the fourth decimal place. The sample was below average optical quality, but the same condition can exist in better material if the limits of measurement are continuously extended without consideration of the quality of the material.

V. Conclusions

Equipment and methods have been described to put a numerical value on homogeneity of high-quality optical materials. This method is, at best, cumbersome and time consuming. On a sample of the recommended size, 50 mm diam by 10 mm thick, determinations of Δn can easily be made at twenty-two different points, with a total fringe reading and calculation time of about 2 h. However, such work is necessary if sufficient data are to be collected on enough samples of different materials in order to obtain some idea of the index differences that may be expected on many samples of nominally the same material. It has been found by this method that variations in the annealing schedule have made significant changes in optical glass homogeneity.[3] It seems reasonable that other materials could benefit from similar work. This method integrates the variations through the thickness of a sample. Therefore, to minimize this variable, all samples should be of the same thickness. If enough samples of dimensions shown in Fig. 3 can be made available, further homogeneity measurements and a more meaningful chart like Fig. 6 can be made.

A new interferometer has been designed in this laboratory for operation at longer wavelength than the present mode. It is planned to make measurements on germanium with the new equipment.

The author gratefully acknowledges the assistance of E. C. Watts for preparing the illustrations and A. Linder and E. P. Muth of our optical shop for the fine sample preparation. Thanks are also extended to the following companies for their cooperation in supplying samples: General Electric Company, Corning Glass Works, Dynasil Corporation, Engelhard Industries, and Eastman Kodak Company.

References

1. *The Manufacture of Optical Glass and of Optical Systems*, Ordnance Dept. Doc. No. 2037, (Government Printing Office, Washington 1921), p. 29.
2. F. Twyman, *Prism and Lens Making* (Hilger and Watts, London, 1952), D. C., p. 508.
3. L. W. Tilton, F. W. Rosberry, and F. T. Badger, J. Res. Natl. Bur. Std. **49**, 21 (1952).
4. E. W. Deeg, Am. Ceram. Soc. Bull. **39**, 674 (1960).

Reprinted from **APPLIED OPTICS**, Vol. 8, page 1703, August 1969

A Special Method for Precise Refractive Index Measurement of Uniaxial Optical Media

Marilyn J. Dodge, I. H. Malitson, and A. I. Mahan

A method was developed for measuring the refractive index of optical glasses and uniaxial crystalline solids when established refractometric methods are not feasible. A synthetic ruby cuboid was contacted to a prism of known refractive index and a spectrometer was used to measure the angles describing the optical path through the ruby–glass combination. Ray tracing equations were derived to compute the refractive index accurate within 3×10^{-5}. Index values for both polarizations of ruby are given at selected wavelengths from 0.4358 μm to 0.7065 μm.

Introduction

Research reported in a previous paper[1] required a knowledge of the refractive index at selected wavelengths for both polarizations of a sample of ruby, a uniaxial crystal in the form of a cuboid with three polished surfaces. Attempts to measure the indexes using critical angle and V-block refractometers were unsuccessful because of instrumental limitations. The crystal was on loan to the authors,* and its geometrical form could not be altered. Consequently, the classical prism method of minimum deviation for precise refractometry could not be used. The technique employed was somewhat novel, yet took advantage of an established goniometric method of refractive index measurement.

A ray of light that is incident at the boundary between two media of different optical densities will suffer total internal reflection if it is incident on the higher density side and the angle of incidence exceeds the critical angle. This is illustrated in Fig. 1, where the optic axis of the ruby crystal is parallel to the intersection of faces *cf* and *cd*. When a beam of collimated, unpolarized light is incident on the side *cf* of the crystal at an angle i_1, the transmitted light is divided into two rays forming different angles of refraction, where i_2 and all subsequent angles refer to the ordinary ray. The rays diverge and become incident on the side *cd* of the crystal at an angle i_3.

Ruby has a refractive index of 1.77 which yields a critical angle of 34° 24′ at the ruby–air interface. Because angle C is approximately 90° and therefore exceeds twice the critical angle, there cannot be an incident angle i_1 that will yield an i_3 less than the critical angle. Therefore, both rays are internally reflected at the ruby–air interface and, after several such reflections, emerge from the opposite side *de* of the crystal. To permit emergence of the two wavefronts from side *cd* of the cuboid, it is necessary to increase the critical angle at *cd* by substituting for air a medium of higher density than ruby.

Measurement Technique

The procedure devised for measuring the refractive index of this crystal is shown in Fig. 2. The cuboid was contacted to a lanthanum dense flint glass prism of higher index than the ruby with a thin layer of liquid of intermediate refractive index. The light which is incident at the boundary *cd* is refracted by the prism at an angle i_4 and emerges from the glass–air interface at an angle i_6.

Before contacting the ruby to the glass, a precision spectrometer was used to measure angle A and the refractive indexes at the desired wavelengths of the glass and angles C and D of the ruby. (For this study, it was necessary to know the value of D in order to calculate i_1.) Techniques for precise measurement discussed by Tilton,[2] Werner,[3] and Fishter[4] were followed. The ruby–glass combination was then mounted on the table of the spectrometer so that the rays incident on the cuboid were perpendicular to the crystal axis. A position was found where deviated light was observed emerging from the glass–air interface, and the prism table was locked at this position. The relative angular positions of the undeviated beam, the normal to the surfaces *dc* and *ab*, and the emergent beams for both polarizations at the desired wavelengths were measured.

The first two authors are with the Image Optics and Photography Section, National Bureau of Standards, Washington, D.C. 20234; A. I. Mahan is with the Johns Hopkins University, Applied Physics Laboratory, Silver Spring, Maryland 20910.

Received 8 January 1969.

* This ruby crystal was loaned to the authors through the courtesy of R. L. Barns of the Bell Telephone Laboratories.

Table I. Refractive Index of LaDF Glass Relative to Air at 22°C and Angles of Emergence, i_6 for the Ruby-Glass Combination

Wavelength			i_6, ordinary			i_6, extraordinary		
(μm)	Source	Index	Deg	Min	Sec	Deg	Min	Sec
0.4358	Hg	1.910183	3	24	34.33	2	20	38.41
0.5461	Hg	1.887970	4	24	54.28	3	21	34.43
0.5876	He	1.882856	4	38	11.30	3	34	44.76
0.6678	He	1.875566	4	55	32.82	3	52	23.39
0.7065	He	1.872884	5	1	39.07	3	58	30.17

Refractive Index Calculations

Ray tracing equations of the conventional Snell form for the ruby–glass combination are:

$$\sin i_1 = n \sin i_2 \quad (1)$$

$$n \sin i_3 = n_1 \sin i_4 \quad (2)$$

$$n_1 \sin i_5 = \sin i_6, \quad (3)$$

where

$$i_3 = C - i_2, \quad (4)$$

$$i_4 = A + i_5, \quad (5)$$

and n and n_1 refer to the refractive index, relative to air, of ruby and the glass, respectively. The angle of incidence, i_1 was calculated to be 35° 37′ 8.67″. Angle A is 59° 59′ 45.60″ and angle C is 89° 59′ 24.21″. The refractive indexes of the glass accurate within 5×10^{-6}, and the calculated values for i_6, for both polarizations and for all selected wavelengths, are given in Table I. Angles i_5 and i_4 can be determined from the tabulated values of i_6 by the use of Eqs. (3) and (5). Introducing Eq. (4) into Eq. (2), and then using Eq. (1), we obtain

$$n = (1/\sin C)[n_1^2 \sin^2 i_4 + 2n_1 \sin i_1 \cos C \sin i_4 + \sin^2 i_1]^{\frac{1}{2}} \quad (6)$$

from which the refractive index, n, of ruby can be computed. The values of n, relative to air for both polarizations of ruby are given in Table II and refer to a controlled laboratory temperature of 22°C and barometric pressure of 99,992 N/m² (750 mm of Hg). These refractive indexes are considered accurate within 3×10^{-5}.

To obtain indexes in terms of vacuum wavelengths, an equation derived by Edlen[6] for the dependence of refractivity of air on temperature and pressure can be used. The absolute refractive index of ruby is then determined by multiplying the values given in Table II by the refractive index of air for the stated environmental conditions at each wavelength.

Discussion

Synthetic sapphire is the host material commonly used in the manufacture of ruby. Cr_2O_3 is added to α-Al_2O_3 and the concentration of chromium ions present in the lattice influences the strengths and sharpness of two broad visible region absorption bands in ruby.[1,5] The ordinary ray has a higher absorption coefficient for the 0.38–0.44 μm band, whereas optical absorption is predominant in the 0.52–0.6 μm band of the extraordinary ray. The observed relative intensities of the spectral lines showed the influence of these absorptions, but did not seriously impair index measurement within these bands.

Since absorption and dispersion are interrelated, the refractive index also varies with chromium content. The synthetic ruby crystal used in this study contained 0.062 ± 0.002% Cr_2O_3 by weight.[1,7] Figure 3 shows the data of Table II compared graphically with previously reported results of Mandarino,[8] Houston,[9] and Dowell[10] for ruby crystals of different chromium con-

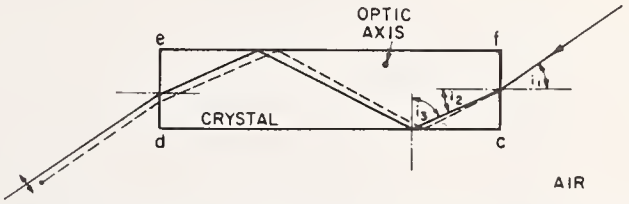


Fig. 1. Total internal reflection in a ruby cuboid. The solid and dashed lines indicate the ordinary and extraordinary rays, respectively.

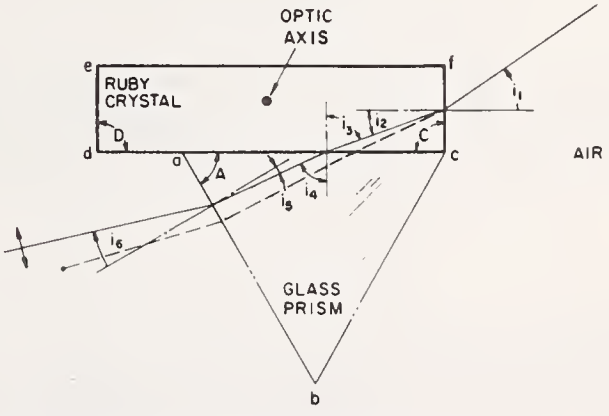


Fig. 2. Ruby-glass combination showing the optical paths for both polarizations, and the angles for the ordinary ray.

Table II. Refractive Index of Ruby Relative to Air at 22°C

Wavelength (μm)	Index	
	Ordinary	Extraordinary
0.4358	1.78115	1.77276
0.5461	1.77071	1.76258
0.5876	1.76822	1.76010
0.6678	1.76445	1.75641
0.7065	1.76302	1.75501

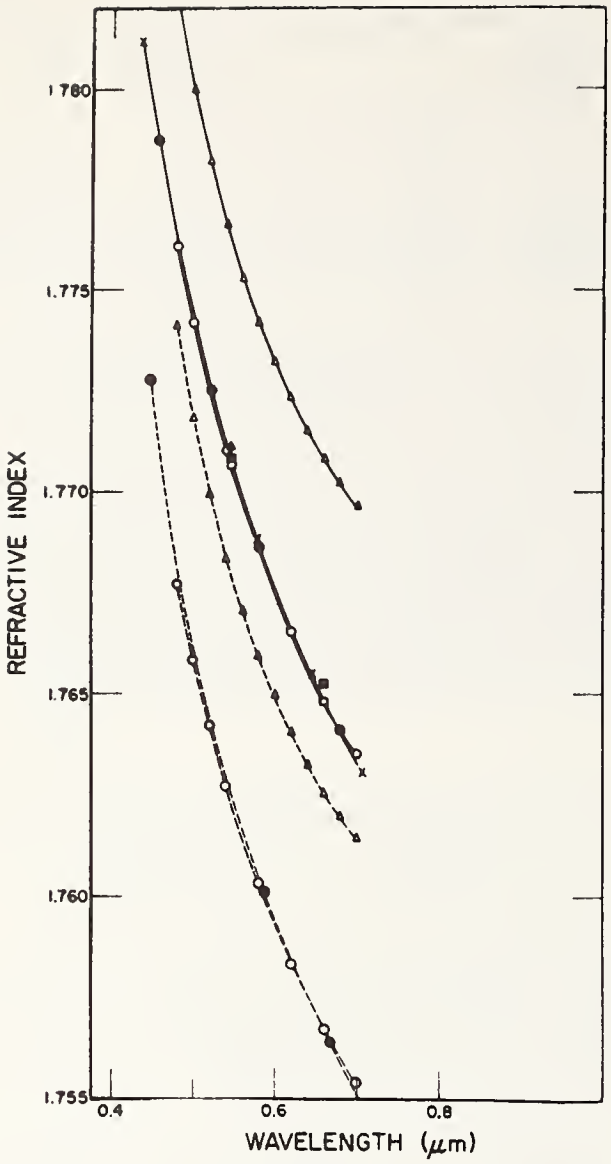


Fig. 3. Comparison of the refractive index of sapphire and ruby as a function of wavelength. The solid and dashed lines indicate the ordinary and extraordinary rays, respectively. X, Sapphire, Malitson; □, Sapphire, Dowell; ▲, 0.01% Cr_2O_3, Houston; ●, 0.062% Cr_2O_3, NBS; ■, 0.077% Cr_2O_3, Dowell; O, 0.11% Cr_2O_3, Mandarino; △, 1.4% Cr_2O_3, Mandarino.

centration. Also included are the index data of Dowell[10] and Malitson[11] for synthetic sapphire. The data of Houston and Dowell are for 24°C, but the temperature was not specified for Mandarino's measurements. No allowance for the effect of temperature has been made for the data plotted in Fig. 3, but assuming that the temperature coefficient of index of ruby is similar to that of sapphire, 1.3×10^{-5}/°C,[11] differences caused by temperature are probably less than 5×10^{-5}. It is evident from these curves, that there is general agreement in refractive index between crystals that contain approximately the same percent of Cr_2O_3, and that the refractive index increases with increased chromium content.

Summary

The procedure of measuring these refractive indexes was adapted to the geometrical shape of the ruby specimen and utilized the techniques already developed for measuring the refractive index of prisms. This technique should not be regarded as a substitute for the minimum deviation method, since one must determine more quantities and the equations for computation are more complex than the simple law of refraction. However, this special method is very useful for determining the refractive index of uniaxial optical materials not available in prismatic form.

This paper was presented in part at the 1967 Annual Meeting (October) of the Optical Society of America, in Detroit.

References

1. A. I. Mahan, C. Bitterli, S. M. Cannon, and D. G. Grant, J. Opt. Soc. Amer. **59**, 49 (1969).
2. L. W. Tilton and J. K. Taylor, in *Physical Methods in Chemical Analysis* (Academic Press Inc., New York, 1960), Vol. 1, p. 451.
3. A. J. Werner, Appl. Opt. **7**, 837 (1968).
4. G. E. Fishter, in *Applied Optics and Optical Engineering*, Rudolf Kingslake, Ed. (Academic Press, Inc., New York, 1967), Vol. 4, Chap. 10.
5. T. H. Maiman, R. H. Hoskins, I. J. D'Haenens, C. K. Asawa, and V. Evtuhov, Phys. Rev. **123**, 1151 (1961).
6. B. Edlen, Metrologia **2**, 71 (1966).
7. D. M. Dodd, D. L. Wood, and R. L. Barns, J. Appl. Phys. **35**, 1183 (1964).
8. J. A. Mandarino, Amer. Mineralogist **44**, 961 (1959).
9. T. W. Houston, L. F. Johnson, P. Kisliuk, and D. J. Walsh, J. Opt. Soc. Amer. **53**, 1286 (1963).
10. Private communication. F. R. Charvat, Union Carbide Corporation, Linde Division, Indianapolis, Ind. (1964 Refractive Index data of H. G. Dowell.)
11. I. H. Malitson, J. Opt. Soc. Amer. **52**, 1377 (1962).

2. Interferometry in Image Optics

Foreword

Most of the papers in this section are the work of J. B. Saunders, whose ingeneous applications of interferometry to the evaluation of optical instruments have lately taken on renewed importance as workers, such as R. E. Swing and D. N. Grimes, have used these techniques in their approach to image evaluation from the point of view of statistical optics. It appears likely that these interferometric techniques will continue to find broad application. The paper by Funkhouser and Mielenz on high-speed holographic interferometry is applicable to the evaluation of optical components subject to vibrational distortion. The Grimes method of evaluating partial coherence is more convenient and more generally applicable than prior methods. Saunders' papers on interferometry applied to alignment were not considered appropriate to this volume.

U. S. Department of Commerce National Bureau of Standards

RESEARCH PAPER RP1668

Part of Journal of Research of the National Bureau of Standards, Volume 35, September 1945

AN APPARATUS FOR PHOTOGRAPHING INTERFERENCE PHENOMENA

By James B. Saunders

ABSTRACT

A photographic instrument is described that was designed for recording the changes produced in the order of interference fringes over long periods and for recording large changes in the order of interference. When this instrument is used with interferometric systems for studying the changes caused in transparent solids by heating, it can be made to yield a continuous record of the simultaneous changes in temperature, time, index, strain, and density. For example, to record those changes that are caused by the annealing of an optical glass at an annealing temperature near the lower part of its annealing range requires several weeks of continuous recording. The total amount of 35-millimeter film needed by the instrument for such a record (covering, say 3 months) does not necessarily exceed 10 feet. To record the expansion of a material that is being heated from room temperature to 500° C, at a rate of 3 degrees centigrade per minute, requires approximately 1 foot of 35-millimeter film. Thus, even very long tests require such a small amount of film that the operator is able to process it in any small dark room. The fringes are photographed at their natural size, and the fringe shifts can be determined from the photographic record without the use of enlarging or projection equipment.

CONTENTS

I. INTRODUCTION

The increasing application of interferometry to the measurement of physical changes in material has created a demand for an automatic recorder. Visual-observation methods have proved unsatisfactory for extended tests because they demand the continuous and undivided attention of an observer and because of the frequent uncertainty resulting from errors in fringe count. The motion-picture camera, as applied by Nix and MacNair [1][1] to the photography of interferometer data, eliminates many of the objections inherent in the visual method. However, this method has certain disadvantages that have prevented its general use. Each individual picture requires an exposure of several seconds, even when the brightest of monochromatic sources of light is used. Consequently, if 20 pictures are required (as was proposed by Nix and MacNair) to record the passing of each fringe, it follows that 1 fringe per 20 times the minimum exposure time is the fastest rate of change in the order of interference that can be recorded. Many tests require a faster means of recording than is possible with a motion picture camera under such conditions. The motion picture method is also impractical for slow changes extending over long periods because the number of pictures per hour cannot be materially reduced without danger of failing to record changes that may occur between exposures. Because of these objections and the cost and difficulties of processing the large amount of film required, this method has not been generally accepted.

Trowbridge [2] proposed a photographic method for recording interference phenomena that permits the recording of much faster changes in the order of interference. Trowbridge's method has the further advantage of continuity over the motion-picture method and requires a relatively small amount of film. Arnulf [17], and later Sinden [18], modified Trowbridge's method but apparently failed to produce a practical instrument. By the addition of several new features, a practical means for recording all types of interference fringes has now been produced.

To fulfill the needs of this laboratory, where the interferometer is used in a large number of tests, the design and construction of an automatic photographic recorder was undertaken. Such an instrument must satisfy certain requirements if it is to be practical. That is, the operation of the apparatus must be reasonably simple; the photographic record (in the form of 35-mm film) must be of such length that the operator can conveniently process it in his own laboratory; the monochromatic light source must be constant, enduring,

[1] Figures in brackets indicate the literature references at the end of this paper.

and sufficiently intense to permit continuous recording of the fringes for long periods; the accuracy of the results must be at least equal to that attainable by visual observation; and the operator must be able, during the recording period, to observe visually and to make necessary adjustments without interfering with the continuity and quality of the record. The instrument described in this paper fulfills all the above requirements and extensive use has proved it to be an exceptionally reliable and easily operated instrument. In the following paragraphs this apparatus will be designated an "interferograph" [3] and the resultant records "interferograms."

II. DESCRIPTION OF THE METHOD

If a fringe system, such as is shown in figure 1, *A*, is focused on a strip of photographic film, a latent image is formed, which, when properly processed, yields a reproduction of the fringe system. However, if the image is partially intercepted by a screen that has a narrow slit perpendicular to the strip of film and which is placed just inside the focal plane of the fringe system, an exposure will yield a reproduction of only the narrow part of the image, which is transmitted through the slit. Figure 1, *B*, shows that part of the image that is intercepted by the screen, whereas figures 1, *C* and 1, *D* show the part transmitted through the slit to the film. If the fringes remain fixed and the film is caused to move perpendicularly to the slit and at a rate that is suitable for proper exposure, the resultant effect is that shown in figure 1, *E*. That is, the part of the fringe system that is transmitted by the slit traces dark and light bands parallel to the direction of motion of the film. On the other hand, if the fringes are caused to move across the plate from bottom to top, in a direction that is parallel to the slit, and the film moves from right to left, the exposed film will yield the reproduction shown in figure 1, *F*. That is, the bands, which were traced parallel to the edge of the film while the fringes were static, will make an angle to the edge if the fringes are caused to move. Movements of the fringes in the opposite direction will produce a mirror image of the reproduction shown in figure 1, *F*.

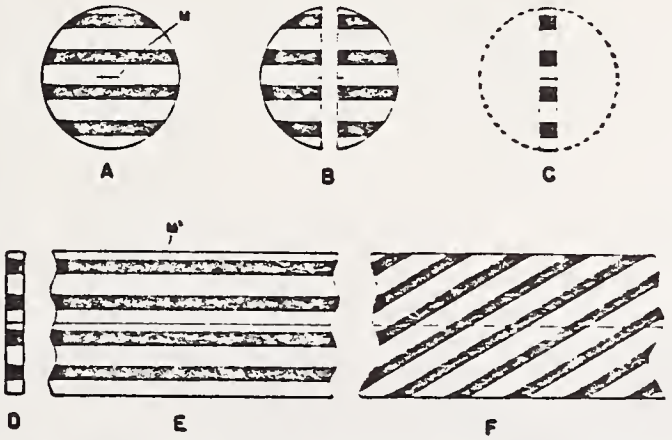


FIGURE 1.—*Diagram to show method of producing interferograms from interference fringes.*

(See text for explanation.)

To determine the fringe movements accurately from the interferograms a reference point is required either in the object plane or in the focal plane of the fringes. Either will serve the same purpose. In most cases a convenient reference point is obtained by scratching a sharp line on the surface of the top interferometer plate, as shown by M in figure 1, or by placing a fine wire (or some other kind of filament) across the top plate. A cross hair in the focal plane of the camera and very close to the film will serve also. In either case, the image or shadow of this fine line should intersect the slit (approximately ¼ mm wide) perpendicularly. This intersection serves as a reference point and traces a sharp line, M' in the latent image on the film. The line appears black on the positive reproduction and its intersection with a photographically reproduced fringe corresponds to the time when that fringe passed the reference point. If the instrument records more than one phenomenon (in addition to time), such as both temperature changes and expansion, then more than one set of fringes is needed and each set requires an individual reference point. A complete datum from such a record should include the order of each set of fringes at a definite time. In order to locate those points that are photographed simultaneously on the reference lines, transverse lines that are equally spaced and parallel to the slit are produced photographically at definite intervals. Such a line is produced either by cutting off the light source for a few seconds (thus producing a narrow, underexposed strip) or preferably by flashing a small lamp suitably located in the camera box. The latter method is used here. That is, a flash of short duration overexposes a narrow strip equal to the width of the slit. These over- (or under-) exposed strips will hereafter be designated "slit-lines." The intersections of the slit-lines with the above-mentioned reference lines serve to locate simultaneously exposed points along the reference lines. The slit-lines are produced automatically and at equal intervals of time by an electrically operated time switch. Since the film moves uniformly, the spacing of the slit-lines affords an accurate measure of time, and simple interpolation is possible for reading points between any two slit-lines.

The data may be read directly from either the negative or a positive reproduction with the aid of a properly designed reading frame, described in section III–5, c, page 175. The reading frame may, however, be dispensed with if the interferogram is prepared for reading by superposing on it straight lines that are parallel to the slit-lines and pass through the desired reading points. This may be done by placing the interferogram on a drawing board so that the slit-lines are parallel to one edge of the board and then by using a T-square to inscribe additional lines parallel to the slit-lines at all selected reading points along either reference line. The author prefers to make such lines at those points along the thermometer reference lines that indicate integral orders of interference. This is preferred because these points correspond to the temperatures and the air corrections (described by Peters and Cragoe) [4] that are given in the calibration sheet of the interference thermometer. The orders of the different sets of fringes can then be read off directly at points where both the temperatures and air corrections are known.

III. DESCRIPTION OF THE APPARATUS

1. GENERAL

A diagram of the assembled apparatus is shown in figure 2 and a photograph of it is presented in figure 15. The several parts are shown in greater detail in additional figures and in all of these drawings a

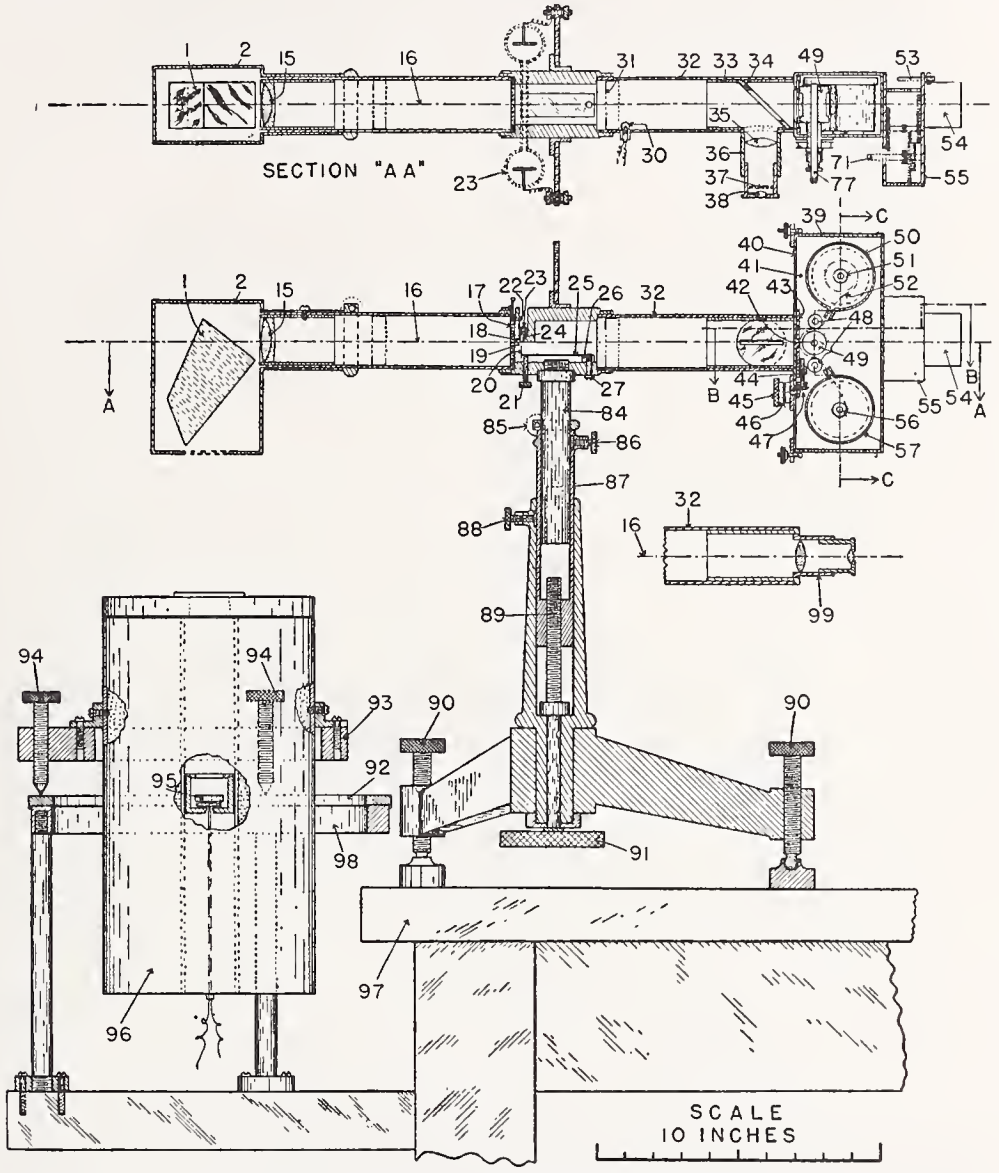


FIGURE 2.—*Interferograph and furnace.*

1, Constant-deviation prism; 2, prism housing; 15, collimating lens; 16, optic axis; 17, iris diaphragm; 18 and 22 are, respectively, the effective source and its semiadjustable jaw; 19, adjustable slit for the isolation of spectral lines; 21, knurled screw for moving adjustable jaw 20; 23, spectrum tube; 24, right-angle reflecting prism; 26 and 27 are, respectively, a block and screw for holding the phosphor bronze strip 25; 30, flash lamp; 31, diaphragm; 32 and 33, telescoping tubes; 34, diagonal mirror; 35, 37, and 38 are, respectively, the collecting lens, shutter, and eye-lens of eyepiece 36; 43, semiadjustable jaw of exposing slit 42; 44, 45, 46, and 47 are, respectively, the screw, knurled head, scale, and right-angle gears for controlling the width of exposing slit 42; 48, two idling rollers; 49, sprocket; 50, cartridge which houses the supply spool 51; 52, photographic film or paper; 53, electrical contact post; 54, motor; 55, gear housing; 57, cartridge which houses the take-up spool 56; 71, gear shaft; 77, sprocket shaft; 84 and 87, telescoping shaft and tube, respectively; 85, 86, and 88, clamping screws; 89, focusing screw with knurled head 91; 90 and 94, leveling screws of the instrument and furnace, respectively; 92, rotatable ring; 93 and 98, supporting rings; 95, interferometer; 96, furnace; 97, table; 99, removable eyepiece. Sections *BB* and *CC* are shown in figures 9 and 8, respectively.

reference number always indicates the same part of the apparatus. In making thermal expansion measurements the complete apparatus, for procuring the interferograms, consists of the interferograph, an electric furnace, an electric control board, and several accessories.

The interferograph is essentially a modification of the viewing instrument, described by Merritt [5], with the addition of a photographic recorder. The principal changes incorporated in the new instrument are as follows: The objective of the telescopic eyepiece and collimating lens are replaced by a single collimating lens; the light filter in the eyepiece and the right-angle reflecting prism at the collimating end of the viewing instrument are replaced by a constant-deviation prism; and a focusing mechanism is installed in the standard or base of the new instrument. Additional changes are also made to improve the collimation, to increase the illumination, and to adapt the viewing instrument to a photographic recorder.

The furnace is quite similar to that described by Peters and Cragoe [4]. Its mounting, however, is changed to permit both rotation and tilting. This change is essential to the efficient operation of the camera because of certain tilting effects. When the temperature of the furnace is changed, differential expansion within the furnace causes small but definite tilting of the interferometer relative to other optical parts. This tilt must be rectified at intervals in order to prevent the fringes from being effaced from the record. For visual observations the type of mounting is of less importance as the rectification may be effected by the leveling screws of the viewing apparatus. In photographic recording, the interferograph should not be tilted in order to effect such a rectification, as this cannot be accomplished without shifting the image of the interferometer relative to the film. A shift of this kind causes a break in the recorded fringes, and although this produces no appreciable error in results, it does increase the difficulty of reading the record. The furnace mounting permits adjustment to compensate for this tilting without introducing appreciable shift in the image on the film.

In making this adjustment it is essential that a means be provided to permit visual observation of the fringe system without interrupting the continuous recording by the camera. This is accomplished by the eyepiece and mirror described in sections III–2, e and f, p. 167–68. It will be noted (fig. 1) that the recorder uses only a narrow strip of the image. The remaining part of the image is reflected to one side, by means of a properly designed mirror, and observed with an accessory eyepiece (fig. 2). Thus the process of observing the image visually, for making necessary adjustments, does not interfere with the continuity of the record.

Results of interferometry with a dispersion prism are superior to those obtainable with a light filter. For interferometer tests in which the observations are made visually, it is generally sufficient to use a good source of monochromatic light and a suitable light filter. For photographic recording, however, the distribution of the spectral sensitivity of the film makes necessary a higher degree of homogeneity in the light, and it is desirable, therefore, to use a dispersion prism. To obtain a large dispersion, a constant deviation prism of the Pellin and Broca [6] type is used to replace the right-angle prism of the ordinary instrument. The light passes through the dispersion prism twice, thus producing sufficient dispersion to isolate any

spectral line in the source if its nearest neighbor differs in wavelength by as much as 50 angstroms. The selection of any line is accomplished by a simple rotation of the prism about an axis which is so chosen that image shifts are negligible.

Visual observations at high temperatures are more difficult with filtered light (used by most of the present-day viewing instruments) than with dispersed light. A filter that will eliminate the incandescent glow of a hot furnace and also isolate a spectral line will usually absorb a large percentage of the light from the transmitted band, thus making both visual and photographic observation difficult. That is, any filter that transmits only a narrow band will also absorb a large portion of the light in this band. If a dispersion prism is used for monochromatization, a filter that absorbs practically all light of long wavelength and that transmits a large percentage of the light of short wavelength may be used to eliminate the incandescent light of the furnace without serious absorption of a selected line of short wavelength. However, a filter is not necessary for photographic work unless panchromatic film is used. By using noncolor-sensitized film (insensitive to the greens, yellows, and reds) with a blue line of the source, only light from this blue line has any effect on the film unless the furnace temperature exceeds 1,100° C. Consequently the result is a clear photograph, free from haze or fog. A filter may still be needed in the eyepiece for visual adjustments.

2. OPTICS OF THE INTERFEROGRAPH

The essential optical parts of the interferograph are shown in figure 3. Light from the source, 23, after being reflected by the right-angle reflecting prism, 24, and collimated by the lens, 15, passes through the constant-deviation prism, 1, to the interferometer, 95. The normal to the center of the interferometer is assumed to form a

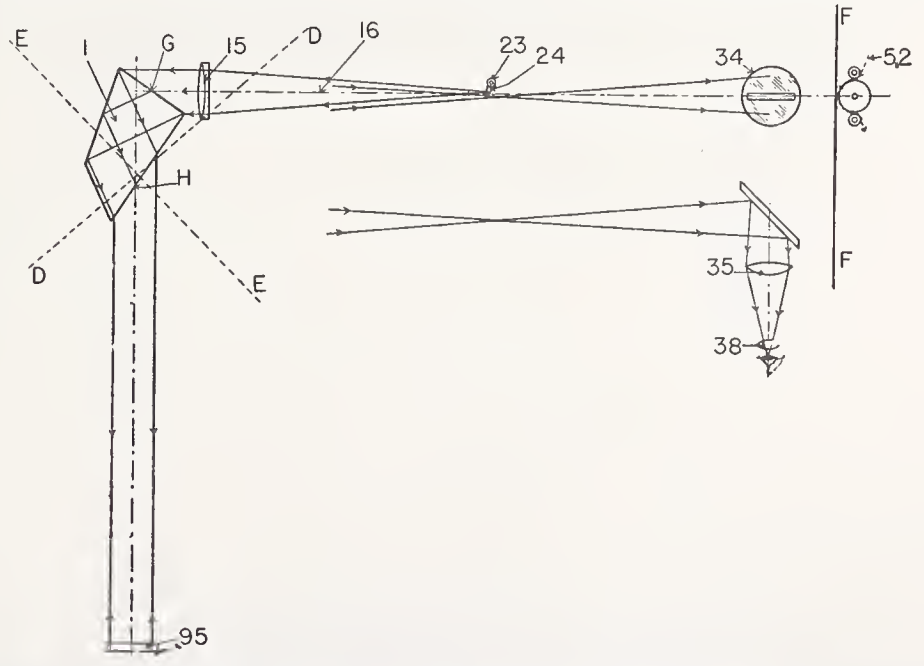


FIGURE 3.—*Optical parts of the interferograph.*

FF, image plane; *G* and *H*, entrance and exit points, respectively, of prism 1; *EE* and *DD*, the loci along which the axis of rotation of prism 1 may be chosen for eliminating, respectively, axial and longitudinal shifts of the image. All other parts are indicated in figure 2.

perpendicular intersection with the optic axis of the collimating lens. The light is dispersed on passing through prism 1. By rotating this prism about an axis perpendicular to the plane of this figure, any one of the refracted beams, corresponding to the different spectral lines of the source, may be caused to fall perpendicularly onto the interferometer plates. On reflection from these plates this beam will retrace its path through the prism to the source. The light beam from neighboring lines will also return through the prism with increased dispersion. From each of these beams, lens 15 will form an image of the source in a vertical plane near the source. By rotating prism 1, any one of these images may be caused to fall on slit 19 (fig. 2), and the light from this line will then pass on to the diagonal mirror, 34. A slit in this mirror transmits a narrow rectangular part of this beam to the film, 52. The remaining part is reflected by the mirror to the collecting lens, 35, which converges the light to a cross section that is sufficiently small for simultaneous observation of all rays from the entire field. The eye lens, 38, enables visual observation of the interference phenomena.

(a) LIGHT SOURCE

The light source ordinarily used with this apparatus is a Pleucker tube of helium gas. This tube (23, fig. 4) rests on a long, slender,

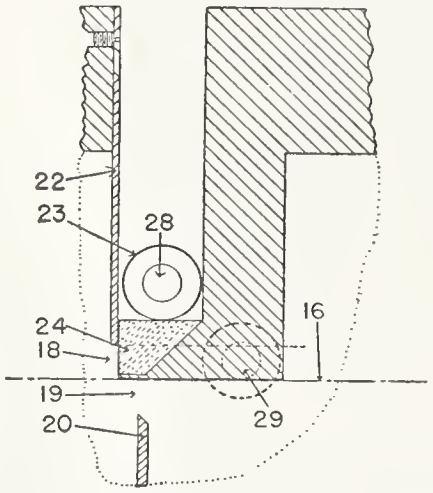


FIGURE 4.—*Enlarged section through light source.*

28 is the capillary of the spectrum tube and 29 its reflected image. All other parts are indicated in figure 2

right-angle reflecting prism,[1] 24, which has one of its edges ground away so that the image, 29, of the capillary, 28, can be located very close to the optic axis of the collimating lens. The width of the effective light source is limited by the slit, 18, that is formed by the lower edge of the reflecting surface of the prism, 24, and the semi-adjustable screen, 22. Placing this slit close to the axis of the lens permits better collimation, less cosine error in the fringe system, and consequently much sharper fringes. The prism, 24, makes these advantages possible.[2]

[1] The instrument described by Merritt (shown in fig. 1, [5]) is designed so that the helium tube (23, fig. 4) lies in the position of the prism, 24. With this arrangement the separation of the returning beam of light from the source is determined by the thickness of the wall of the capillary. Consequently, either the source or its image (produced by light returning to the focal plane of the collimating lens) must be some distance from the optic axis; thus, a larger cosine error is produced in the collimated light. By introducing a slender prism similar to prism 24, it was found that the maximum usable separation of interferometer plates (length of specimen) was doubled. A corresponding improvement in fringe definition was also attained for any given separation of plates.

The effective shape of the light source commonly used for interference measurements with reflected light is that of a half-moon [7]. This shape is considered best for securing maximum intensity with a given cosine error and is entirely satisfactory if the spectral line employed has no close neighbors. However, it is sometimes necessary to use a narrow slit-like source for the isolation of close spectral lines. This can be done without sacrificing intensity if all but a narrow slit along the capillary of the spectral tube is silvered.

In fact, a considerable increase in the directed light from a spectral tube is produced by silvering the back of the capillary. If one-half (180 degrees of arc) of the capillary is silvered, the intensity of the light that is effective in producing the interference fringes is approximately doubled. If two-thirds (or 240 degrees of arc) of the capillary is silvered, the intensity is approximately trebled. Up to a certain point, further narrowing of the unsilvered aperture of the capillary produces a still further increase in the density of the source.[3] This increase in surface density is limited, however, by the increasing absorption resulting from the increasing number of internal reflections before the light emerges from the narrowing slitlike window. Thus, the silvered capillary permits the use of a narrower source and a greater intensity for a given cosine error. A decrease in width of the source and an increase in surface density assist in the isolation and use of close and less intense lines from a multiline source.

In preparing a narrow slit of this kind, the capillary is silvered over its whole circumference for a length of approximately 1 cm near the middle. The silver is then coated with shellac, or any other suitable coating, for protection. After the coating is dry, an axial slit with straight edges and with the desired width is then produced by removing a narrow strip of the silver with a sharp razor blade. In the case of the well-separated helium lines, a width of 2 mm is suitable.

(b) COLLIMATING AND FOCUSING LENS

The collimating lens (15, fig. 3) is a cemented doublet, corrected for chromatic and spherical aberration. Light reflected from its surfaces [4] forms ghost images, but these do not interfere with the photographic record if they are made to fall below the exposing slit. The placement of the optical parts is arranged so that the collimating lens also serves as the focusing lens. Fortunately, this arrangement is convenient both for obtaining the desired magnification and for properly locating the controlling aperture (18, fig. 4). The number of optical parts is thus reduced to a minimum.

(c) DISPERSION AND DEVIATION PRISM

If a right-angle reflecting prism (R, fig. 5, A) is combined with the two parts, $P1$ and $P2$, of a 60-degree dispersion prism (fig. 5, B) in the manner shown by figure 5, C, the result is a prism having both the dispersion properties of the dispersion prism in figure 5, B, and the right-angle deviation properties of the reflecting prism, R. The path of the ray of light shown in the reflecting prism is that for which the prism produces no dispersion. The path of the ray of light shown

[3] This method of increasing the effective illumination has also been applied to mercury vapor lamps. It is recommended for all such sources where increased illumination from spectral lines is desired.

[4] It is desirable to give this lens the treatment described by Cartwright [5] for extinguishing reflection of light from glass.

in the dispersion prism is that for which the prism produces minimum deviation. For the combination, shown in figure 5, *C*, each part of the dispersion prism produces both the same dispersion and deviation as shown in figure 5, *B*, but as the beam is inverted by reflection before entering the second half of the dispersion prism, the deviation of this half cancels the deviation of the other, whereas the dispersions are still additive, as in figure 5, *B*. The dispersion prism in figure 5, *B*, may be set for minimum deviation of a ray of any other wavelength by a simple rotation, and when thus rotated the new ray passes along the same path within the prism as that previously followed by the first-mentioned ray. The same is true for a rotation of the constant-deviation prism, except that no change is produced in the total deviation. The prism used with this interferograph, and shown in figure 5, *D*, has the same properties as that shown in figure 5, *C*, and is a single piece of glass.

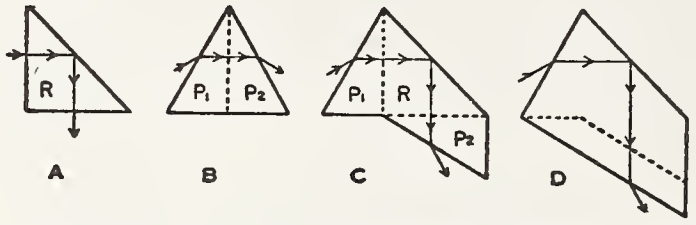


FIGURE 5.—*Properties of the constant-deviation prism.*

A, right-angle reflecting prism; *B*, 60-degree dispersion prism; *C*, constant-deviation prism compounded from parts of *A* and *B*; *D*, constant-deviation prism made from a single piece of glass.

A rotation of the constant-deviation prism (1, fig. 3) in changing from one line to another generally produces an axial shift in the image plane of the interferometer. This is caused by a change in the effective object distance. That is, in general, both the air and glass paths, comprising this distance, change with the rotation of the prism. If, however, the axis of rotation is chosen anywhere along the line *DD*, the change in the air path cancels the change of path in glass to a close approximation. Rotation of this prism also produces a lateral (vertical in this case) displacement of the interferometer image. This displacement is caused by a lateral shift in the rays on emerging from prism 1. For example, when the prism is in the position shown, the axial ray of the collimating lens emerges as the axial ray of the interferometer plate and the image points of these coincide in the image plane, *FF*. If the prism is now rotated clockwise about the entrance point, *G*, of the axial ray, this ray, now of another color, but deviated by 90 degrees, still follows the path traced through the prism to the exit point, *H*, and emerges parallel to the axis of the interferometer plate; but the point *H* has been moved to the left by this rotation. Thus, the axial ray of the collimating lens and that of the interferometer plate will be imaged at two different points, one above the other, in the image plane, *FF*. In general, rotation about any other point also causes a shift in the exit point of the axial ray, 16. It has been found, however, that if the axis of rotation is chosen along the line *EE*, the lateral shift of the image caused by the rotation of the prism may be minimized to the same degree as the axial shift was minimized by the means described above. Fortunately, the intersection of these two lines, *EE* and *DD*, is so located relative to the

prism that it can be chosen as the position of the axis of rotation for the prism. As the result of this choice, both shifts are practically eliminated.

The purpose of eliminating any image shift is to improve the application of the interferograph for measuring the length of samples to a very high degree of accuracy by means of the "theory of coincidence," described by Perot and Fabrey [9]. However, the lateral shift of the image does not affect ordinary visual observations.

(d) INTERFEROMETER

The recording of time is automatic with the interferograph, as described in section II, page 160. Consequently, if the data to be recorded consist only of temporal changes in any single quantity that may be observed with an interferometer, then one set of fringes is sufficient. However, if two or more kinds of such changes are to be recorded, additional sets of fringes are necessary.

Two sets of fringes are required for recording both thermal expansion and temperature by interferometry. In figure 6 the lower plate, T, serves as an interference thermometer [10], and the expansion fringes, showing the expansion of specimen S, are produced in the usual manner by light reflected from the top surface of T and the bottom

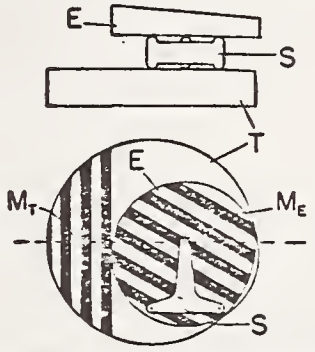


FIGURE 6.—*The interferometer.*

E, An interferometer plate; *S*, test sample; *T*, interference thermometer; M_T and M_E, reference lines of the temperature and expansion fringes, respectively.

surface of plate E. The broken line in figure 6 represents the position of the exposing slit relative to the two reference points, M_T and M_E, of the temperature and expansion fringes, respectively. The thermometer reference point, M_T, is the intersection of the reference mark (M, fig. 1, A) with the indicated position of the exposing slit. The reference mark should be sufficiently long so that an appreciable rotation of the interferometer about M_E may be effected without removing the image of M_T from the exposing slit. Also the reference mark should be parallel to the fringes so that the rotation causes no change in the order of the thermometer fringes relative to their reference point.

(e) DIAGONAL MIRROR

The diagonal mirror (34, fig. 2), used for intercepting a part of the light of the interferometer for visual adjustments or observation, may be made either of silvered glass or stainless steel. The slit, through its center, should be wide enough to permit some latitude in the adjustment of the exposing slit and also for visual adjustment of the

image in the slit before closing the camera to begin a recording. A 1.5-mm slit in the mirror has been used suceessfully.

(f) EYEPIECES

Two eyepieces [5] (36 and 99, fig. 2) of identieal optical design, and corrected for spherieal aberration, are used with the interferograph. Each consists of a collecting lens and an eye lens that are separated by their eommon foeal lengths of 5 cm. The collecting lenses are sufficiently large to cover the whole image, whereas the eye lenses are quite small.

Eyepieee 99 is designed for easy removal and, after the interferometer is assembled, it is replaeed by the photographic recorder and diagonal mirror. When this eyepieee is in its position for visual work, its collecting lens is in the image plane of the interferometer. Eyepiece 36 is permanently attaehed to the side of the interferograph, as shown in figure 2. It is used, during the recording period, for observing the interferometer and for making adjustments when necessary. The collecting lens of this eyepiece is located in the reflected image plane of the interferometer. This eyepiece has a shutter, loeated between its lenses, that is controlled by a meehanism similar to that used in cameras having the between-the-lens type of shutter. The shutter operates exaetly as a camera shutter with "bulb" and "time" settings. It is opened only after the eye is placed in position for observing and closed before its removal. A camera release cable greatly reduces the danger of jolting the instrument when opening and closing the shutter. A small exit pupil and an eyecup help to prevent stray light from getting to the photographic film while observing.

3. MECHANICS OF THE INTERFEROGRAPH

For effieient operation the interferograph must permit elose control of light conditions. This requires the calibration of several seales which indicate the settings of deviees that eontrol the amount and quality of the light producing the reeord. Also it is important that these controls be operative without affecting the continuity of the record.

(a) ROTATION OF THE CONSTANT-DEVIATION PRISM

The constant-deviation prism (1, fig. 7) is fastened between two plates with two layers of felt to serve as padding. Two pressure springs, 6 and 12, tend to rotate the prism clockwise about its bearing shaft, 10. The rotary motion is limited by the lever arm, 8, which bears against pin 7 on plate 3 and is rotatable about screw 5. The rotation of this lever arm is controlled by the motion of the micrometer serew, 13. Pressure springs, 6 and 12, maintain permanent contact between the arm, 8, and both the micrometer screw, 13, and pin, 7. Thus, all baek lash is eliminated and the accurate calibration of the micrometer head, 14, against the wavelength of known spectral lines is possible. The bearing shaft, 10, is held in place by a ring, 9, and set screw, 11.

(b) DIAPHRAGMS AND STOPS FOR LIGHT CONTROL

Figure 4 shows a cross section of the speetrum tube, 23, with the unsilvered slit facing downward. The reflected image of this slit

[5] By modifying the mounting of eyepiece 36 it could serve both purposes, thus eliminating eyepiece 99.

faces the collimating lens which is to the left. Slit 18 is slightly narrower than this reflected slit and is, therefore, the effective source. The area of this slit-like source of fixed width is controlled by means of an iris diaphragm (17, fig. 2). It is important that the size of the iris diaphragm be readily controlled. This is accomplished by means of a lever in combination with a calibrated scale.

The light from the effective source (18, fig. 2) is collimated by lens 15 and dispersed and deviated by prism 1. It then travels vertically to the interferometer where it is reflected back through prism 1 which further increases the dispersion and causes another right-angle deviation. Lens 15 then produces, in a vertical plane through 18, inverted images of the light source, one corresponding to each spectral line in the source. A rotation of prism 1 moves these images vertically. By observing through eyepiece 99, the operator may cause any one of these images to coincide with aperture 19.

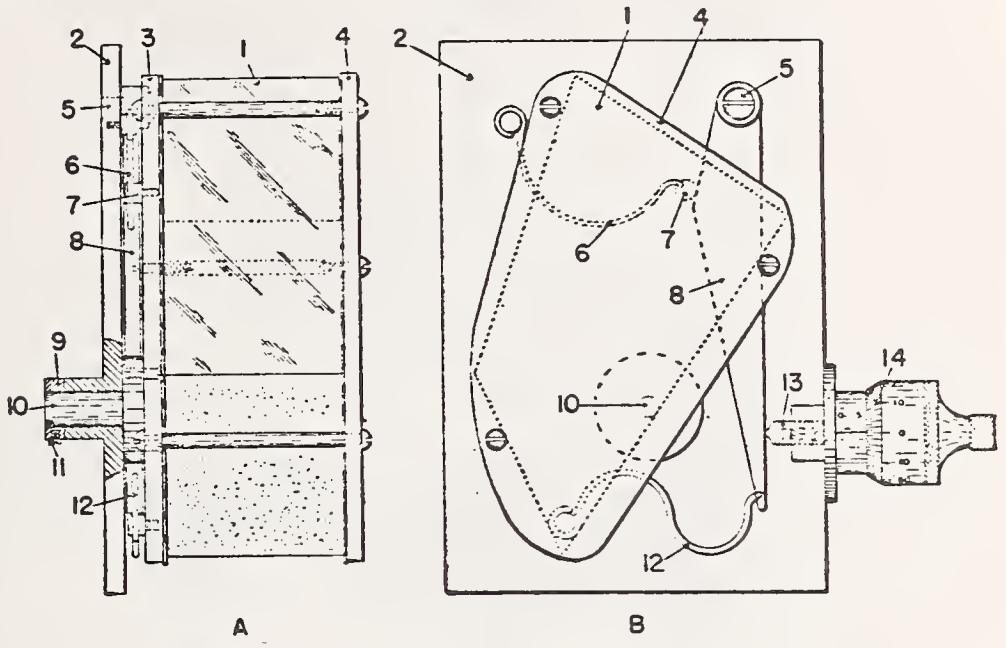


FIGURE 7.—*Mounting of the constant-deviation prism.*

3 and 4, Plates for holding the prism in place; 5, pivot screw; 6 and 12, springs exerting thrusts on pin 7 and lever arm 8, respectively; 9 and 11, ring and set screw, respectively, for preventing axial motion of shaft 10, on which the prism rotates. The rotation of the prism is controlled by means of screw 13 with the aid of a micrometer scale 14.

The width of slit 19 is determined by the position of screen 20 which is controlled by means of knurled headed screw 21 (fig. 2). This screen is attached to a phosphor bronze strip, 25, which maintains contact with screw 21 and is securely fastened to the instrument through block 26 and set screw 27. The length of slit 19, which in operation is approximately equal in width to that of 18, is also controlled by iris diaphragm 17.

Additional diaphragms are placed in suitable parts of the instrument to prevent stray light from reaching the photographic film. One such diaphragm is shown by 31 in figure 2. All inside parts of the instrument are painted a dull black.

518-743 O - 73 - 8

(c) CAMERA

The front part, 40, of the camera box is brazed to tube 33, in which the diagonal mirror is located. Tube 33 telescopes into tube 32 and is replaceable by eyepiece 99 for visual work. Both the camera box and prism housing 2 may be rotated 90 degrees about the optic axis, 16, of the interferograph. This permits the instrument's use for recording with vertical-type interferometers.

The essential elements of the camera box are shown in figures 2 and 8. It is lighttight and made in two sections which may be

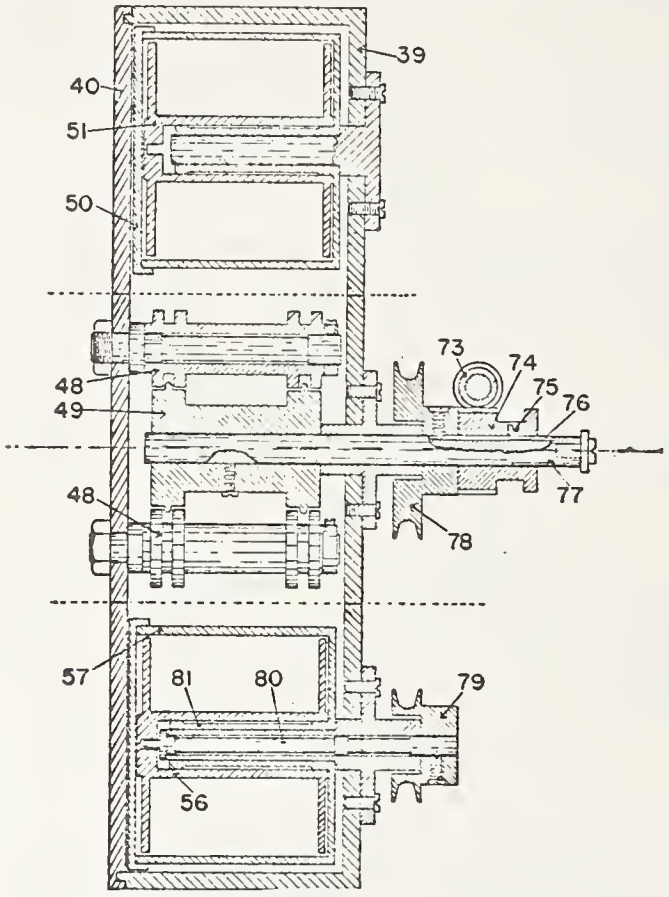


FIGURE 8.—*Camera box (section CC of fig. 2)*

73, Worm; 74, worm gear; 75, key; 76, keyway; 39 and 40, the two separable sections of the camera box; 78 and 79, slip-belt pulleys. All other parts are indicated in figure 2.

separated. In viewing the camera box in the direction indicated for observing section *CC* of figure 2, the top, bottom, back, and right-hand sides form one of these sections, 39, which is referred to as the removable part, and the front and left-hand sides form the other section, 40, which is referred to as the fixed part as it is brazed to tube 33.

The film spools are housed in lighttight cartridges, 50 and 57, which can be removed from the camera box without exposing the film. The slots through which the film is fed into and out of the cartridges are lined with a deep, black plush to prevent light leakage. The capacity of each spool is approximately 35 feet of photographic film. An

extra take-up spool with its cartridge permits the processing of one record while making another.

The mechanism for moving the film is shown in figure 9. The power for driving sprocket 49, which imparts uniform motion to the film, is supplied by a small synchronous motor, 54, which by means of gear 58 drives gear 59, the shaft of which also carries gears 60 and 61. The three gears, 62, 63, and 64 are fastened rigidly to a sleeve, 65, that can be moved along a shaft, 71. This movement permits the alternative engagement of three pairs of gears, 59–62, 60–63, and 61–64, each of which gives a different gear ratio and consequently permits the selection of three different speeds for the movement of the photographic film. Sleeve 65 is constrained to turn with shaft 71 by a set screw, 67.

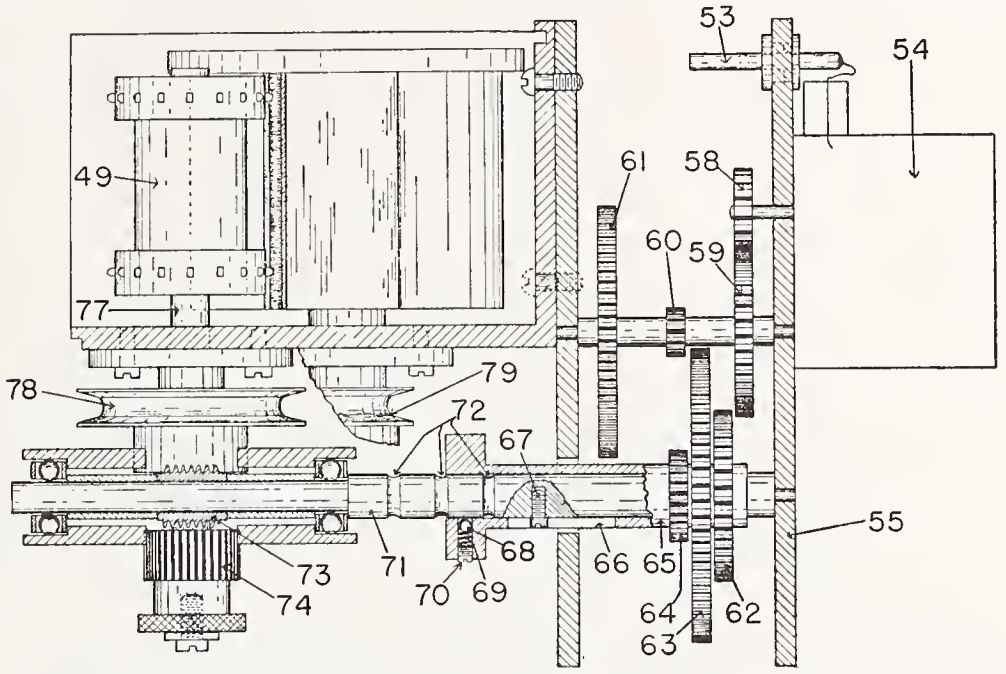


FIGURE 9.—*Film-driving mechanism (section BB of fig. 2).*

58, Transmission gear that drives the set of gears 59, 60, and 61 at a fixed speed; 62, 63, and 64, a second set of gears that rotate together with sleeve 65 and can be made to engage gears 59, 60, and 61, respectively, by moving sleeve 65 along shaft 71; 66, groove; 67, pin; 68, steel ball that is pressed against shaft 71 by pressure spring 69 and set screw 70. All other parts are indicated in figures 2 and 8.

Slot 66, which is parallel to the axis of the sleeve in which the head of set screw 67 is located, permits movement of the sleeve along the shaft without relative rotation. A steel ball, 68, and its spring, 69, holds the sleeve in position when the gears are meshed as its position then coincides with one of the three grooves, 72. The sprocket, 49, and worm gear rotate with shaft 77.

In order to bring unexposed film into position before starting a record and also to wind up that part of the exposed film that is between the exposing slit and the take-up cartridge at the end of a recording, it is necessary to move the film manually, independent of the driving mechanism. To accomplish this the worm gear can be disengaged from the worm in the manner shown in figure 8. That is, the worm gear can slide along shaft 77 while key 75, which fits into the keyway, 76, constrains the two to rotate together in either its

engaged or disengaged positions. Thus, the sprocket may be rotated manually when the worm gear is disengaged. In order to move the film a small known amount, ball 68 may be set half way between any two of the adjacent grooves 72 (as in the position shown in fig. 9), thus disengaging the gears. With the worm gear engaged, sleeve 65 may be rotated manually, causing the film to move. A 360-degree rotation of the sleeve will then move the film exactly one-tenth of an inch.

The take-up spool (56, fig. 8) is powered through the V-shaped pulleys, 78 and 79, which are connected by a spring slip-belt (not shown). These pulleys are constructed so that the plane of the belt falls within the bearing areas of their shafts. This eliminates side thrusts that tend to increase friction. Further reduction in friction is attained by reducing the bearing areas of shaft 80 and sleeve 81 on which spool 56 turns. The sides of the pulley grooves are made very steep and the diameter of the spring slip-belt slightly less than the maximum diameter of the pulley grooves.

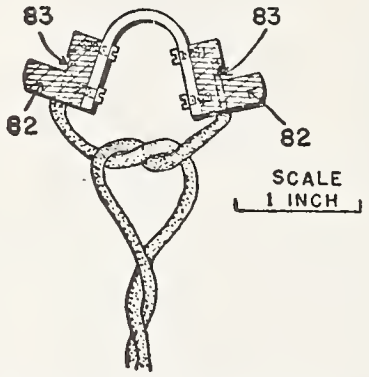


FIGURE 10.—*Electric connector for the camera motor.*

The two nonconducting parts, 82, are separated by and fastened to a steel spring. Electric contact is made by compressing the spring and placing the two terminals, 83, in contact with the terminal posts (53, fig. 2) of the motor. Tension in the spring maintains permanent electric contact.

The film cartridges and sprocket, with its driving mechanism (motor and transmission gears), are attached to the removable part of the camera box (fig. 2), and the two floating rollers and exposing slit with its controls are attached to the fixed part. This arrangement permits both easy access to the slit for its adjustment relative to the image of the interferometer and easy threading of the film between the rollers, 48, and the sprocket wheel.

When loading and unloading the camera, it is desirable to disconnect the electric power. Also it is important to guard against moving the instrument while remaking the connection. The terminals of the motor are connected to two parallel posts (53, fig. 2). The attachment, shown in figure 10, for electrifying these parts is light and allows connections to be made without danger of shaking the instrument. The other end of the electric cord is plugged directly into the 110-volt alternating-current power line.

(d) FOCUSING MECHANISM

Focusing of the interferograph (fig. 2) is accomplished by varying the object distance, which is the optical distance between the interferometer, 95, and the collimating lens, 15. This may be done

either by raising or lowering furnace 96 by means of the three leveling screws, 94, by raising or lowering the interferograph by means of its three leveling screws, 90, by sliding shaft 84 along inside of tube 87, or by operating the elevating screw, 89, by means of wheel 91. Only the latter two means are used in practice, one for coarse adjustment and the other for fine adjustment.

4. FURNACE

Figure 2 shows the furnace, partly in elevation and partly in section. It is supported by a ring, 93, which in turn is supported by three equally spaced leveling screws. These screws rest on ring 92 which fits on and over another ring, 98. Ring 92 is rotatable about ring 98. An overhanging flange constrains the axis of the two to coincide at all times. The rotatable top ring carries the furnace and the bottom ring is fastened rigidly to supporting parts which are firmly attached to the table on which the interferograph is supported. The electric connections (thermocouple and heater leads) to the furnace are made flexible so as to prevent breakage from flexure caused by rotating the furnace. The interferometer, 95, is near the center of the furnace and on a level with the rotatable ring, 92.

5. CONTROL BOARD AND OTHER ACCESSORIES

(a) WIRING DIAGRAM AND TIME SWITCHES

A wiring diagram of the control board is shown in figure 11. The Variac transformer, T_1, is well suited for automatic control of heating rates. The intensity of discharge in the spectrum tube, H, is controlled by means of a slide-wire rheostat, R_2, in series with the primary of a 12,000-volt neon-sign transformer. A control of the intensity of the source is necessary because such control is the best means of adjusting the exposure to the several rates of film speeds. The time-switch transformer, T_3, delivers 6 volts across its secondary. This voltage is divided between the time-lamp, FL, and the resistance, R_3 (a 100-ohm, radio-type, slide-wire rheostat) so that the lamp may be made relatively faint when fast films are used.

The slit lines, defined in section II, page 160, are produced by causing the time-lamp, 30, (AA, fig. 2) to flash on momentarily. Two switches make it possible to control this lamp either automatically or manually. The manually operated switch is used to record the time of incidental events such as a change in heating rate and the automatic switch records the passage of time intervals only. The automatic time switch was constructed from an electric clock. The contact point (C_1, fig. 11) makes contact with the second hand, SH, once each minute. The duration of this contact is approximately 1 second. The contact, C_2, makes contact with the modified minute hand, MH, once in each 30 minutes. The duration of this contact is approximately 1 minute. The C_2 contact is made when the second hand is approximately 180 degrees from its contact position; consequently, the C_1 contact is made near the middle of the period of contact at C_2; thus the possible danger of two C_1 contacts during one of the C_2-contact periods is eliminated. With this adjustment made, the time-light is on for 1 second at accurate intervals of 30 minutes. If more or less frequent intervals are desired, the number (two in the case described) of contacts on the modified minute hand must be

changed accordingly. A disk, having any number of such contact points, would produce that number of flashes during each hour. The manually operated time switch is of the push button or doorbell type.

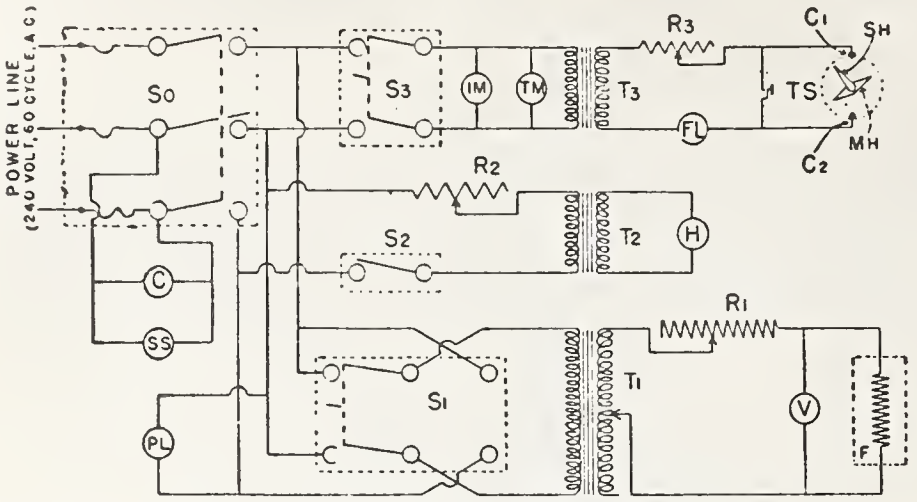


FIGURE 11.—*Wiring diagram of the control board.*

S_0, Triple pole, single throw switch; S_1, double pole, double throw switch; S_2, single pole, single throw switch; S_3, double pole, single throw switch; T_1, variable transformer, Variac; T_2, neon sign transformer; T_3, door bell transformer; R_1, R_2, and R_3, variable rheostats; *C*, electric clock; *SS*, spare socket; *PL*, panel board light; *IM*, interferograph motor; *TM*, time-switch motor; *FL*, flash lamp; *TS*, time switches: automatic and manual; *MH*, minute hand; *SH*, second hand; *H*, spectrum tube (helium); *F*, electric furnace; *V*, voltmeter.

(b) FILM-LOADING KEY

A film-loading key (fig. 12) is convenient for loading the supply cartridge with unexposed film. This key is shaped so that it engages the spool in the same manner as shaft 80 of figure 8. The film is attached to a hook on the spool, which is then inserted into the cartridge, with the film passing out through the plush-lined slot. The

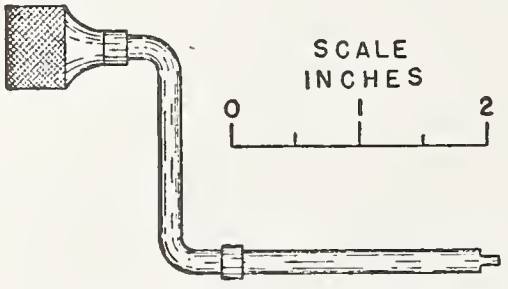


FIGURE 12.—*Film-loading key.*

cartridge is then closed and the film wound onto the supply spool until full.

Notations are made on the emulsion side of the film, after each ten turns, by means of a stylus.[6] This serves to keep the operator informed regarding the amount of film remaining after each successive record is cut off and developed. Similar notations, that identify the individual tests, are written on the protruding end of the film at the end of each record. One loading of the supply cartridge serves for

[6] This stylus may be similar to that used with an autographic camera and the notations develop in the same way as the autographs do from such cameras. A hard lead pencil also serves excellently in making these notations.

many complete records as each of these usually requires only 1 to 2 feet of film.

(c) READING FRAME

In section II, page 160, one method is described for transforming the interferograms into numerical data. The record may also be read directly from the negative with the aid of a reading frame. This consists of two plates of glass that are separated approximately one-half millimeter by two strips of cellulose (or any other suitable material). The separator strips have their adjacent edges parallel, straight, and separated by a distance slightly greater than the width (35 mm) of the photographic film used. The film is moved through the rectangular slot so produced. A straight mark on one of the glass plates, parallel to the slit lines, serves in determining the fringe orders of the several sets of fringes corresponding to any desired intermediate time or any desired order of a set chosen for any particular reason. The usual procedure, in thermal expansion work, is to read the order (whole number and fraction) of the expansion fringes corresponding to chosen integral orders of the interference thermometer fringes.

(d) MOUNTING TRIPOD

A tripod used in mounting and arranging the interferometer is an important adjunct. It consists merely of a flat-topped tripod with adjustable feet. Its height is adjusted so that the object distance will remain unchanged when the interferometer is transferred to the furnace. This tripod is placed near the furnace so that the constant-deviation prism may be swung around to a position directly over it merely by rotating the interferograph on its base about shaft 84. The interferometer system is assembled on this tripod and adjusted while being observed through eyepiece 99.

IV. ADJUSTMENTS OF THE APPARATUS

When the interferograph is used for expansion work only, some of the necessary adjustments are final, whereas others must be repeated for each new sample or test. The interferograph may be leveled by the use of two horizontal mirrors, formed by shallow dishes of clean mercury. One mirror is placed on the tripod and the other replaces the interferometer in the furnace. After the instrument is once levelled the levelling screws, 90, should remain unmolested. By replacing the mercury mirror, on the tripod, with an optical flat and observing the reflected light, from this mirror, the instrument is collimated by adjusting the position of lens 15 until the field, observed in the eyepiece, is uniformly illuminated. The instrument is focused by adjusting the height of the standard, as described in section III–3, d, page 172. A ground glass is placed in the plane of the photographic film and the image of a reference line is observed on this glass with an eyepiece while focusing. The reference line should be at the object distance intended for the interferometer while in the furnace. If the change in this object distance remains small, no further focusing will be necessary.

The adjustments required for a simple interferometer (that is, one having a single set of fringes) have been given in a previous publication [5]. A compound interferometer, such as that shown in figure 6,

imposes additional requirements on the adjustments, the most important of which is the requirement that the two or more sets of fringes shall be equally bright. The orientation of the interferometer image, relative to the exposing slit, is also more critical for the compound interferometer as it is necessary to have the images of all reference points fall in the exposing slit.

Plate E, figure 6, is made wedge-shaped so as to eliminate the light reflected from its top surface. Consequently, the light that is transmitted by this plate, which forms the expansion fringes, does not return parallel to the light which forms the temperature fringes. That is, the two returning beams do not coincide and, consequently, separate images of the source are formed that are relatively displaced in the plane of aperture 19. As aperture 19 is equal, both in size and in shape, to either image of the source, and as these images do not coincide exactly, all the light from the interferometer cannot pass through this aperture.

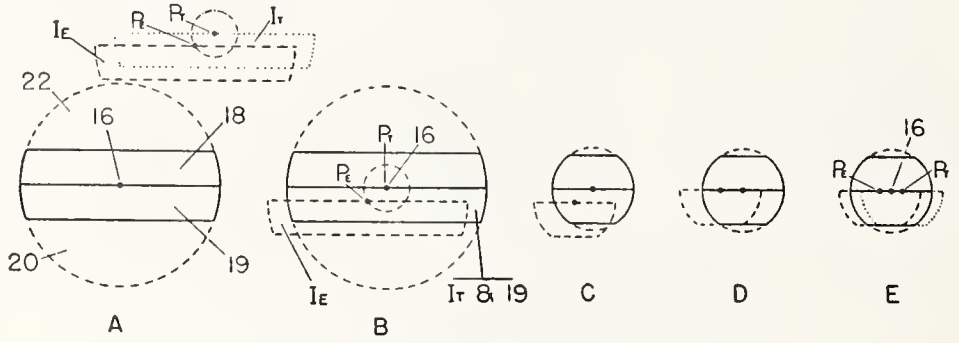


FIGURE 13.—*Adjustments of light beams.*

See text for explanation.

The adjustments for securing maximum brightness of images, consistent with the requirements for equal brightness, are best explained with the aid of figure 13. The parts designated by numbers have already been defined. I_T and I_E are the images of the source, formed in the plane of aperture 19, by light of the thermometer and expansion fringes, respectively, whereas P_T and P_E represent the corresponding images of point 16 which is the intersection of the optic axis with this plane.

Figure 13, A, represents the general condition produced when an interferometer, such as is shown in figure 6, is placed on the mounting tripod, and the interferograph is swung into position for observing it. In general, neither of the two images, I_T and I_E (which are assumed to be equally bright), fall on aperture 19 and, consequently, the interference fringes cannot be seen by an observer looking into the eyepiece. An outline of the interferometer can be seen, however, by diffuse light from an outside source. This light serves to identify the proper location of the interferometer relative to the interferograph. By tilting the tripod the two images, I_T and I_E, are caused to move relative to aperture 19, whereas their positions relative to each other remain fixed. With diaphragm 17 (fig. 2) wide open and by tilting the mounting tripod, the refraction thermometer image, I_T, is easily brought into coincidence with aperture 19, as is shown in figure 13, B.

This condition is attained when the temperature fringes appear brightest. The diaphragm opening is then reduced until the expansion fringes become quite faint. This condition indicates that only a small portion of the image, I_E, is coinciding with aperture 19, as shown in figure 13, *C*. Usually the brightness of the expansion fringes may be increased by rotating plate *E* of figure 6, without shifting it from its proper position, on the specimen, *S*. When plate *E* is thus rotated, the point, P_E, moves on a circle with P_T (which now coincides with 16) as its center. When P_E is on the horizontal line through P_T, the expansion fringes are brighter than for any other angular rotation of the plate. This condition (represented by 13, *D*) is reached when the total amount of light transmitted through aperture 19 becomes a maximum for this size of aperture; however, the two sets of fringes are still unequal in brightness. When the mounting tripod is further tilted in the proper direction, the condition shown in figure 13, *E*, is produced, in which the intensity of the two beams becomes equal without changing the total amount of transmitted light.

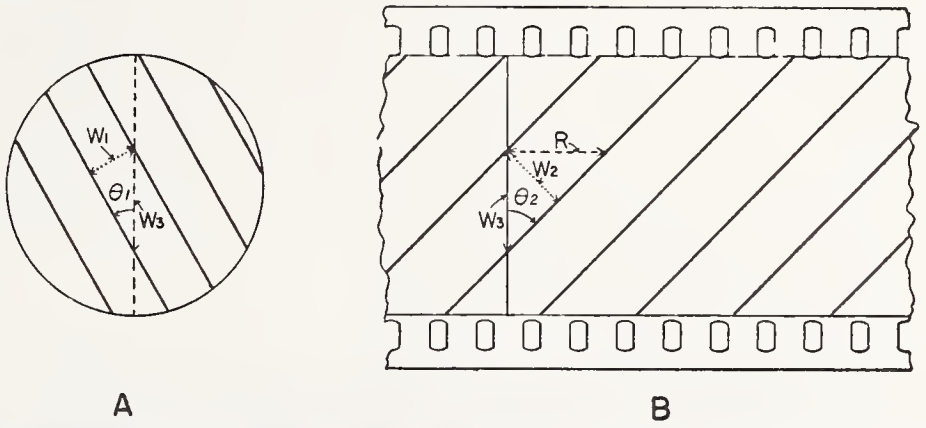


FIGURE 14.—*Relation between interference fringes and interferograms.*

A, Interference fringes; *B*, corresponding interferogram; W_1 and W_2, width of interferometer and interferogram fringes, respectively; W_3, expanse of fringe width along the slit; θ_1 and θ_2, angular orientation of interference and interferogram fringes, respectively; *R*, length of film passing a reference point during the passing of one fringe.

V. ASSEMBLING THE INTERFEROMETER

1. PRELIMINARY CONSIDERATIONS

For accuracy in reading the interferogram, the width of the bands (as in visual readings of the interferometer fringes) should be neither too wide nor too narrow. However, when flat interferometer plates are used, it sometimes happens that the tilting of the plates will cause the interferometer bands to become either much wider than the visible field of the interferometer or too narrow to be counted. These conditions occur when the angle formed by the two reflecting surfaces becomes either too small or too large. Such conditions make it impossible, when taking data visually, to read the fractional order of the fringes. However, this broadening or narrowing of the fringes is usually less troublesome with the interferograph, provided certain precautions are taken while assembling the interferometer. If the width W_1, (fig. 14) and orientation θ_1 of the interferometer fringes

are known, the corresponding width, W_2, and orientation, θ_2 of the interferogram fringes are given by the equations,

$$W_2=(R^{-2}+W_3^{-2})^{-1/2}$$
$$\theta_2=\tan^{-1}(RW_3^{-1}),$$

where $W_3\equiv W_1 \csc \theta_1$ is the expanse of fringe width along the slit, and where $R\equiv V\dot{N}^{-1}$ is the length of film used per fringe. The quantities V and $\dot{N}$ are the speed of the film motion and the rate of change in the order of interference, respectively.

An analysis of the above transformation equations show that W_2 can never exceed either R or W_3. Therefore, if either of these quantities are prevented from becoming large, the width of the interferogram fringes cannot become too large for reading. If the surfaces of the interferometer plate are not flat (as is usually the case) it is quite obvious that the interferometer fringes can never become too wide.

The operator can never be sure that the fringes will not become too narrow during a test. If the width, W_1, of the interferometer fringes become too small, then either W_2 becomes too narrow or contrast on the interferogram will be poor because the condition of narrow fringes approximating parallelism to the slit is approached. If the fringes become both parallel to and as narrow as the slit, there will be no contrast and therefore no fringes on the interferogram. A procedure that has proved successful is to use the T-type sample, shown in figure 6, and to so adjust the height of the contact points supporting the top plate that the order of interference at the weighted point, M_B, is about five greater than at the unweighted contact points. When the three-pin type of sample is used, the fringes will often grow quite narrow during the heating and must be rectified by a previously described procedure [11, p. 188].

The value of W_2 will also be small if R becomes small. The value of V is limited to finite values; that is, to ¾, 3, and 11.4 inches of film per hour, for the instrument described above. Therefore, the only unavoidable condition that will cause R to become too small is for $\dot{N}$ to become so large that the instrument is unable to provide a film speed sufficient to prevent excessive crowding of fringes. The largest value for $\dot{N}$ that permits satisfactory resolution of the fringes is that which corresponds to approximately 100 fringes per inch of film. To aid in preventing the fringes from becoming narrow, the initial value of θ_1 should be small if the fringes are subject to rotation and changes in their width. In a stable fringe system, such as that of an interference thermometer—where the two reflecting surfaces are permanently bound—the value of θ_1 may have any value that permits a reasonable value for W_3.

2. ASSEMBLY PROCEDURE

The preparation of the specimens which serve as spacers between the interferometer plates is fully described by Merritt [5], Saunders [11], and Johnson and Parsons [12].

The steps usually taken to assemble an interferometer similar to that shown in figure 6 are as follows:

With eyepiece 99 in position, the interference thermometer is placed on the mounting tripod so that an image of its reference line, M_T, will

be perpendicular to the exposing slit. The beam of reflected light from the thermometer is adjusted for maximum brightness by adjusting the screws of the mounting tripod. The specimen, S, is placed in a position so that the exposing slit may be made to coincide simultaneously with the expansion and thermometer reference points, M_E and M_T, respectively. The top plate, E, is then placed on the sample. At this stage the expansion fringes may or may not be visible, depending upon whether the image (I_E, fig. 13) falls in or without aperture 19. In either case, by rotating plate E, the expansion fringes may be made almost as bright as the temperature fringes. The mounted interferometer is then carefully placed in the furnace and at approximately the same orientation relative to the base of the interferograph. Prism 1 is now brought into position, over the furnace, by rotating the interferograph about shaft 84. The interferometer is releveled by means of the furnace leveling screws, 94. A rotation of the furnace restores the relative angular orientation of the two reference points relative to the exposing slit. Eyepiece 99 is now replaced by the front part, 40, of the camera box. A slight movement of the interferograph on table 97 while a magnifying glass[7] is used to observe the interferometer through the exposing slit (opened to approximately 1 mm for this adjustment) will then bring the two reference points into coincidence with the exposing slit. Equal brightness of the two sets of fringes, together with maximum illumination of photographic film, is restored by a final adjustment of the furnace leveling screws. The interferometer is then properly mounted and ready for use.

VI. PHOTOGRAPHIC EXPOSURE

Correct exposure for an interferogram requires good contrast and sufficient density for easy reading. The requirements for contrast are specified below but those for density must be ascertained by trial. These two factors depend both on the type of photographic film and the characteristics (light source and adjustments) of the interferograph.

1. SELECTION OF PHOTOGRAPHIC FILM

The type of photographic film to be used is an important feature in any photographic work. In the case of the interferograph, contrast and speed are the essential factors. The speed rating as specified by the manufacturers does not necessarily indicate the sensitivity of the film to any given spectral line. The relative speed factor for interferographic work with any line may nevertheless be determined from the specified speed rating, together with the spectral sensitivity curve of the film.

The speed of the film is important, as the more sensitive the film is to the light used, the narrower the slit may be made, and the faster the film may be moved past the slit. A narrow slit gives high resolution which is very important when the interferometer bands are narrow and nearly parallel to the slit, and also when they are moving rapidly. The ability to record fast-moving fringes is necessary when-

[7] This magnifying glass may be a simple lens of wide aperture, and it may be permanently attached to the inside wall of the camera box. It should be attached to the fixed part (40, fig. 8) and well back of the sprocket so that it does not interfere with the removal of part 39 when opening and closing the camera.

ever expansion tests are made at high heating rates, whenever moderate rates of heating are used with materials that have a high expansivity, and when following the changes in density of crystals during the transformation from one form to another. When yellow light from the helium source is used, any of the fast panchromatic films are satisfactory.

In deciding upon the required spectral sensitivity distribution of the film, the available light source must be considered. If a helium source is used, it is desirable to select panchromatic film for use with the yellow line because this film is highly sensitive to yellow light. This combination gives plenty of speed and also produces good visibility for visual adjustments. However, greater contrast can be obtained by using some of the noncolor-sensitized films. If the latter type of film is used, blue light is required because such films are insensitive to greens, yellows, and reds but are highly sensitive to the blues. Consequently, high contrast may be obtained by using the blue line (λ=4471A) of helium with a noncolor-sensitized film. The blue line of helium is only about one-half as intense as the yellow line, and the product of sensitivity of film by intensity of the line may be less than the corresponding product when the yellow line with panchromatic film is used. The combination of yellow light with panchromatic film is recommended for temperatures up to 675° C and blue light with noncolor-sensitized films for tests that require higher temperatures.

2. CONTRAST

If the area of the effective source is large, the cosine error in the interference phenomena is large and the fringes appear washed out from lack of contrast. The smaller the source the greater will be the contrast, but a compromise must be considered between contrast and amount of light needed for the record. Contrast in the interferogram depends also upon the width of the exposing slit. That is, the interferograph will reproduce the contrast visually observed in the interference phenomena only if the change in the order of interference is small during the exposure of any point on the film. The best results require the use of a source with the largest area that will give reasonable contrast in the interference fringes, the use of the widest exposing slit that will give reasonable reproduction of contrast, and the making of other necessary variations in exposure by changing the intensity of the source and by varying the film speed.

3. PHOTOGRAPHIC DENSITY OF FILM

Assuming that a reasonable amount of contrast is attained in the interference phenomena, the best exposure is that for which the photographic density approaches zero at the center of the light fringes. The only remaining factors available for controlling exposure (after the consideration already given to contrast) are the intensity of the light source and speed of film motion. The speed of the film motion is determined mainly by the rate of change in the order of interference. Consequently, the intensity of the source is the final factor in deciding the photographic density to be obtained. Hence, if the film speed is changed during a test, the intensity of the source must be changed accordingly to compensate for the resultant change in exposure time.

VII. OTHER APPLICATIONS OF THE INTERFEROGRAPH

The interferograph is especially applicable to the measurement of thermal expansions. However, it is equally well adapted to the measurement of many other interference phenomena. It is suggested that this apparatus furnishes a convenient means of obtaining data for those investigations that require a determination of the absolute order of interference, as by the method of coincidence; for the recording of changes in the order of interference that are very large, such as are encountered in the measurement of the standard meter bar; for the recording of changes in interference fringes over long periods of time, as in ether-drift experiments; and for other investigations that present difficulties to visual observation.

A simple arrangement for measuring the rate of disappearance of strain in a plate of optical glass, such as E of figure 6, is to replace sample S by a nonreflecting sheet of material or by removing the sample S, fine grinding the top part of plate T under plate E and then laying plate E directly on plate T. The arrangement is completed by placing a nicol to the left of and close to diaphragm 17 of figure 2, so that all light passing from aperture 18 to lens 15 and from lens 15 to aperture 19 will pass through the nicol. The light after passing through the nicol will be plane polarized. It traverses the strained plate, E, is reflected (in part) at the lower surface of this plate and again traverses the plate from which it returns through the nicol prism to aperture 19 and on to the photographic film. The nicol prism serves both as a polarizer and an analyzer. The resultant interferogram will contain a measure of time, temperature, and changes in double refraction effects. A polaroid screen, placed anywhere between lens 15 and diaphragm 17, serves as well as a nicol. Greater accuracy may be attained by having the specimen shaped so as to cause the light to suffer several internal reflections before returning to the analyzer.

Preliminary experiments indicate the applicability of the interferograph to the determination of visibility curves of spectrum lines [13] that are composed of doublets, triplets, or single bright lines with associated satellites. The intensity measurements for the various orders of interference, on the interferograms, are easily and rapidly made with a suitable photoelectric densitometer.

VIII. RESULTS

To illustrate an application of the interferograph to expansion work, two interferograms of the expansion of quartz in the range between room temperature and 650° C are reproduced. The interferogram (shown near the bottom of fig. 16, A) was produced while the sample was being heated at a constant heating rate of 2.4 degrees per minute and while 3 inches of film per hour was used. Under these conditions it is impossible to follow the fringe count in the neighborhood of the transformation temperature, 573.3° C. [14] [8]. Consequently, in order to record this part of the data legibly, a second interferogram (a part of which is enlarged and shown in fig. 16, B) was produced which differs from the first only in that the heating rate was lower and the film speed greater. Thus, when 573° C was reached, the heating rate was approxi-

[8] Bates and Phelps have found that the start of the inversion of α-quartz to β-quartz occurs at 573.3° C. This temperature is accepted as one of the base points for calibrating the interference thermometer.

mately 1.6 degrees per hour and the speed of the film was 11.4 inches per hour. These conditions allot 7 inches of film to each degree rise in temperature, whereas in the interferogram of figure 16, A, approximately one-fiftieth of an inch per degree was allowed. This combination of changes in film speed and heating rate is equivalent to a magnification of approximately 330 in the longitudinal direction of the film and was sufficient to resolve the fringes even while the samples was passing through the transformation from low to high quartz.

A U-shaped sample was used as this shape permits an image of one of the unweighted contact points (M_2 of fig. 16) to be included, along with the weighted contact point, M_1, in the exposing slit. Thus, including that of the thermometer, there are three reference lines in each of the interferograms. The two end-contact points of the U-shaped sample were alined with the thermometer reference point so that the images of all three could be included in the exposing slit. As usual, one of the contact points of the sample was loaded by shifting the center of the top interferometer plate toward it.

The purpose of including a second reference point in the field of the expansion fringes is to enable the interpreter to follow the fringe count through the transformation at 573.3° C. A small and usually unavoidable temperature gradient between the two parts of the sample that are imaged in the exposing slit causes them to transform at different times. Consequently, the fringes may be moving quite slowly at one of these reference points while they are moving rapidly at the other where the transformation is occurring.

The initial length [9] of the sample was obtained from the approximate length (measured with micrometer calipers) by applying the "theory of coincidence" to the fractional orders of interference produced by each of several known light sources. The application of this principle has been described by Meggers [15] and Perot and Fabrey [9]. A photograph of the interference fringes, produced by each of the several sources, is obtained by taking a series of still pictures with the exposing slit wide open. To prevent double exposure the film is moved forward after each exposure by an amount slightly greater than the width of the slit. The lines [10] 4358, 5461, 5770, 5791A of mercury and 4471, 4713, 4922, 5016, and 5876 of helium were used for this record.

After the length of the sample at room temperature is known, the absolute order of interference, corresponding to any of the fringes crossing the exposing slit, is also known. This enables the interpreter to label the interferogram fringes (at the starting end of the interferogram) with numbers that correspond to the absolute order of interference of the interferometer fringes producing these fringes. All other interferogram fringes can then be labeled by counting from the original order. These numbers are written above the interferogram in figure 16, A and the last two significant figures of these numbers are written on the corresponding fringes of figure 16, B. Obviously the count of the fringes in figure 16, A, cannot be followed through the transformation temperature but it can be followed in figure 16, B. The order of the fringes above the transformation temperature, in figure 16, A, is easily ascertained from figure 16, B, as the order of the expansion fringes in the two interferograms will agree at all temperatures of this range.

[9] A micrometric determination of the specimen length is sufficient for most expansion work.
[10] Usually five well-known lines are sufficient.

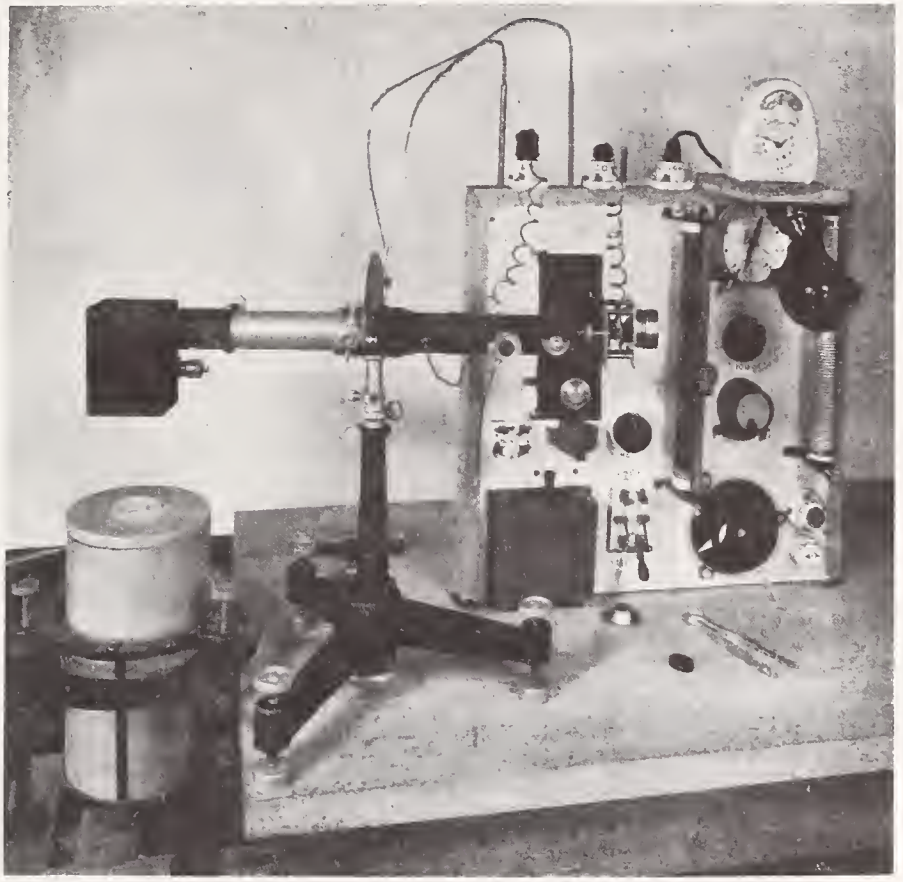


FIGURE 15.—*Assembly of the thermal expansion apparatus.*

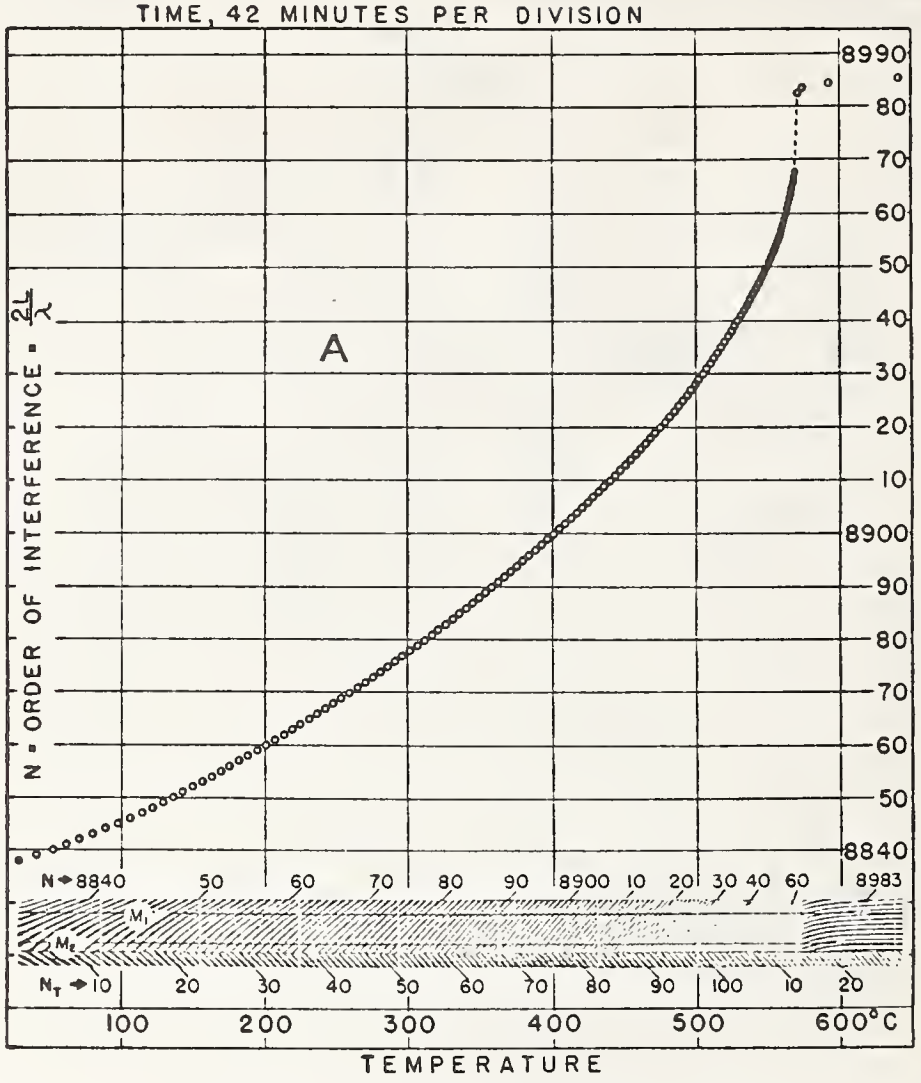


FIGURE 16, A.—*Interferograms of the thermal expansion of quartz with graphical representations of the included data*

N, Absolute order of interference that produced the corresponding expansion fringes; N_T, arbitrarily chosen numbers assigned to temperature fringes for identification and calibration purposes; L, length of sample; λ, wavelength of light used for producing interference; M_1, reference line corresponding to weighted reference point; M_2, reference line corresponding to unweighted reference point. The curve of A and curve 1 of B were obtained from the weighted reference line. Curve 2 of B was obtained from the unweighted reference line. The numbers appearing on the interferogram of B are the last two significant figures of N. The interferograms of A and B were made while approximately 0.02 and 7.0 inches of film was used respectively, for each degree rise in temperature.

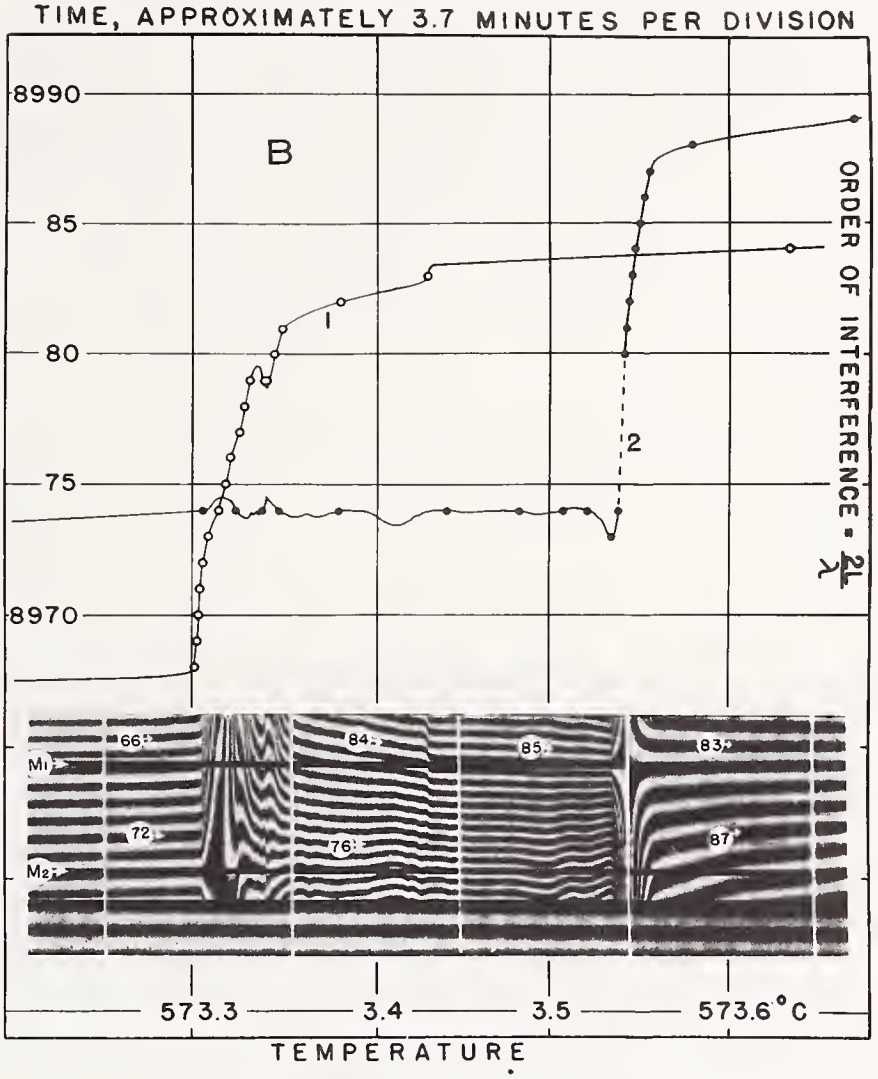


FIGURE 16, B.

(See fig. 16, A, for explanation.)

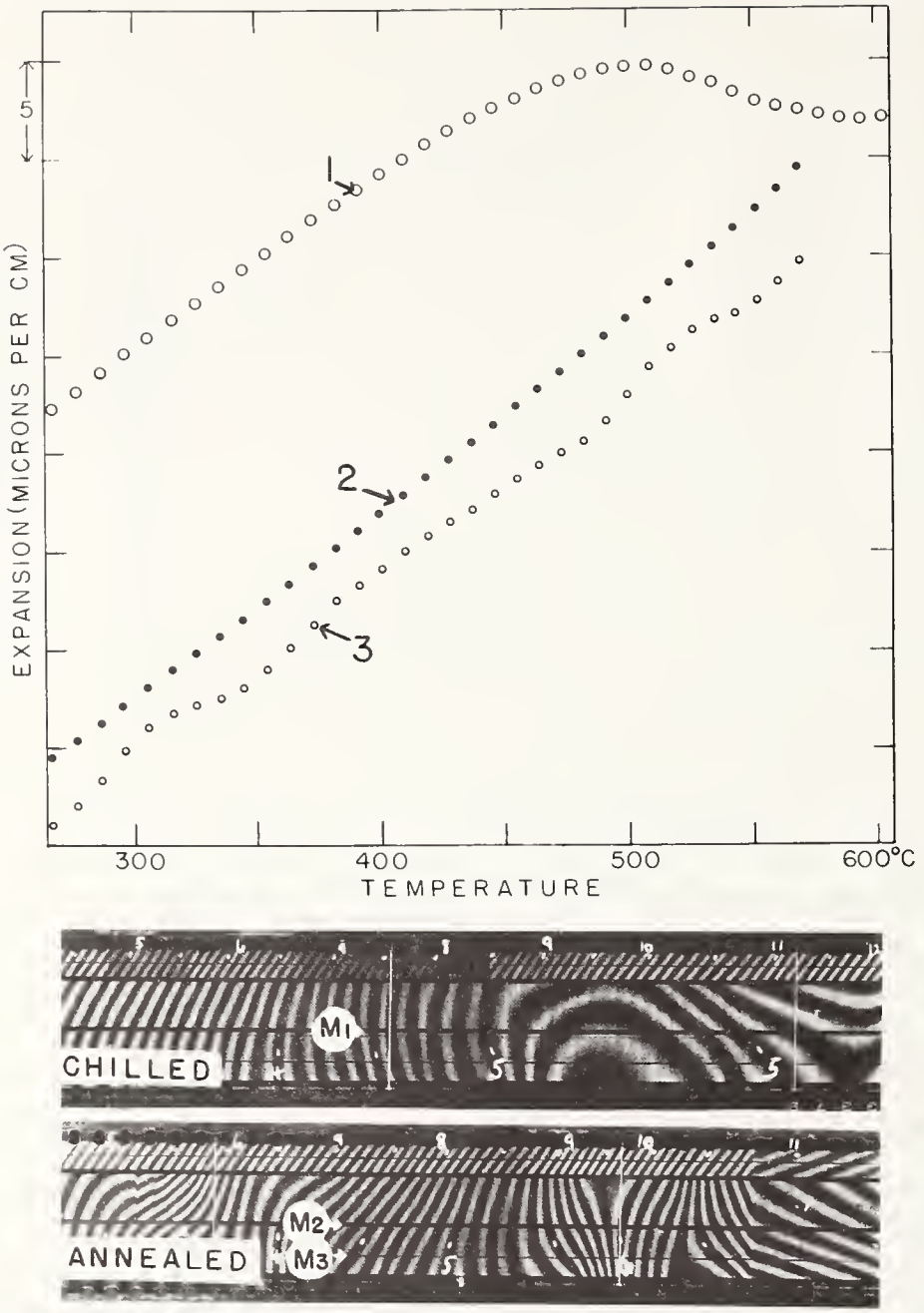


FIGURE 17.—*Interferograms and corresponding expansion curves for a glass in both the chilled and annealed states.*

Curves 1 and 2 were computed from data that were taken along the two parts, M_1 and M_2, of the weighted reference line, whereas curve 3 was computed from the unweighted reference line M_3. The relative position of the curves along the axis of abscissas were arbitrarily chosen to prevent congestion and to save space.

Plots of the absolute orders of interference versus temperature (and time) are given above the interferograms in figure 16. The ordinates represent the absolute length of the specimen in terms of half wavelengths of the light used. The abscissas represent both temperature and time; the time scale is at the top and the temperature scale is at the bottom of the graphs. This simultaneous representation of both temperature and time by the abscissas of a graph is possible because the heating rates were approximately constant. The length of the film exposed was used as a measure of the elapsed time, and the temperatures were determined from the calibration sheet of the interference thermometer. No other computation was necessary in preparing the plotted points, as a drawing board and T-square served to transform the data from the interferogram directly to the graph being plotted.

It will be noted that the slopes (θ, fig. 14) of the expansion fringes on the interferogram and the corresponding parts of the expansion curves, plotted in such a manner, should be closely related if the heating rate and film speed are constant, if there is no tilting of the two reflecting surfaces relative to each other, and if the wavelength is kept constant by performing the test in vacuum. The correlation between the interferograph fringes and the corresponding expansion curves is quite apparant in figure 16 although some tilting and variation in the heating rate, as well as gradients, were present.

Discontinuities such as those in the interferograms of figure 16 are caused by a rapid change in the expanse of fringe width (W_3, fig. 14) along the exposing slit. That is, this expanse increases from a normal value to infinity and then returns to normal after reversing the direction of the fringe count along the slit. This change in the expanse of fringe width along the slit is caused by a tilting of the top interferometer plate relative to the bottom plate.[11] The observed effect (when looking into the eyepiece) may be either a change in width without rotation, a rotation without a change in width, or both a rotation and change in width, the latter being the usual case. The photographed effects resulting from a rotation of fringes are usually not distinguishable from those resulting from a change in the width of the fringes.

If the tilting of the interferometer plate proceeds slowly, as is often the case, the photographed effects include a progressive change in the slope of the interferogram fringes. If the tilt occurs suddenly, which frequently occurs, a discontinuity is produced. Both of these effects are shown in a previously published [3] interferogram. A drawing that represents a part of this interferogram is shown in figure 18. This drawing is included here as an aid in explaining the interpretation of results in the neighborhood of such discontinuities. The numbers occurring in this drawing are inserted to indicate the fringe count from room temperature. The previously published results were taken from the weighted reference line, M_1. Reference line M_2 was produced by a reference point at the median of the triangle formed by the three supporting points of the top interferometer plate. When

[11] In the case of the first discontinuity of figure 16, B, the tilt of the top interferometer plate was caused by a rapid expansion of the weighted part of the spacer that produced M_1. As this part of the spacer increased in length and became equal to that part of the spacer corresponding to M_2 the dark interference fringe, No. 74, which produced the interferogram fringe then coinciding with reference line M_2, became parallel to the slit and this fringe then covered both reference points. After this occurred the movement of the interference fringes, in the slit, was reversed as the numbering of the interferogram fringes indicate. At the second pronounced discontinuity, the 84th fringe was on the weighted reference point M_1 when the unweighted spacer passed through the transformation as it expanded very rapidly. This effect again reversed the direction of motion of the interference fringes along the slit.

the first discontinuity, shown near the center of figure 18, occurred fringe 40 was on the weighted reference point. A sudden tilting of the plate caused a corresponding change in the expanse of fringe width along the exposing slit. It has already been demonstrated [11] that this rotation of the fringes always occurs about an axis that passes through the weighted contact point. Therefore fringe 40 still passed through the weighted reference point after the sudden tilt occurred. It is known that the sample was expanding uniformly from a consideration of other interferograms, which recorded the changes in this material through this same temperature range. Consequently, the next fringe to cross the weighted reference line is known to have been 41. The discontinuity shown near the right end

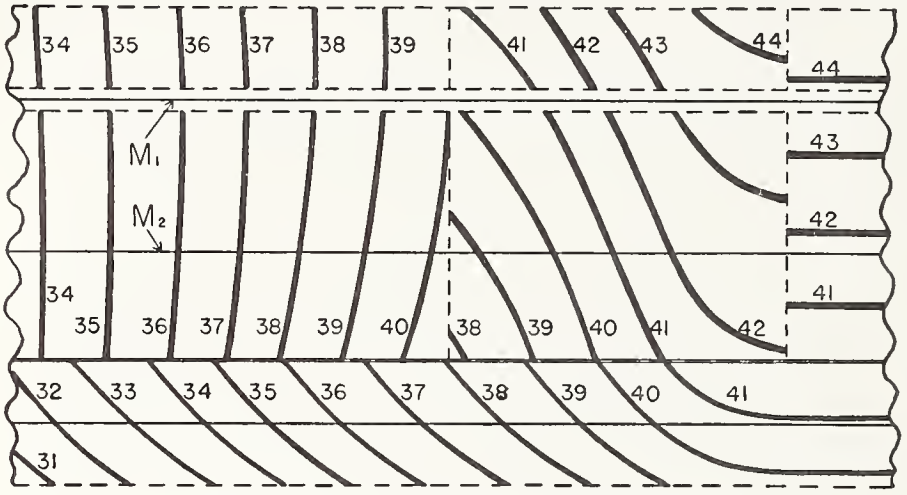


FIGURE 18.—*Discontinuities in interferograms.*

The inserted numbers represent the orders of interference, counted from room temperature; M_1 and M_2 are the weighted and unweighted reference lines, respectively. The upper and lower sets of fringes are expansion and temperature fringes, respectively.

of figure 18 occurred when the temperature of the sample was constant and when the fringe count at the weighted reference point was 43.7, and it is again apparent that no change in the fringe count occurred at the weighted reference point. If the expansion data had been obtained from a fringe count taken at the median reference line, M_2, the magnitude of the sudden change in the fringe order at this point would not have been noted and the results corresponding to points above this temperature would have been in error to the extent of one fringe at least. The correct fringe count can always be followed by accepting the fact that no sudden change occurs in the order of interference at a properly weighted reference point.

For representing expansion data graphically, the usual procedure is to tabulate the data obtained from points along the selected reference lines of the interferograms and from these data to compute the temperatures and corresponding changes in length. The results are then plotted without further reference to the interferograms. Such a graphical representation, along with the interferograms from which the original data were taken, is introduced in figure 17 to show the close correlation between an interferogram and the corresponding computed curve, even when it passes through maximum and minimum points. The location of such points are more readily obtained from the interferogram than from the curves. The same is also true of

critical points in the expansion curves of crystals (fig. 16). The interferograms in figure 17 are parts of the record of a glass that was previously chilled from its molten state. The chilled sample was heated at a constant rate of 3 degrees centigrade per minute to 610 degrees centrigrade and annealed at this temperature. It was then cooled to room temperature and, without dismantling the interferometer, a second expansion test was made on the sample, now in an annealed state. The complete record is not shown here as the excluded parts contained nothing of current interest. Curves 1 and 2, which are computed from data taken along reference lines M_1 and M_2, are typical expansion curves for chilled and annealed glass, respectively [16]. In this figure, reference lines M_1 and M_2 are produced by the weighted reference point, whereas M_3 is typical of accidental marks formed by particles of dust that get into the slit during a test. Curve 3 was computed from data obtained at M_3 and is included to demonstrate the effect of errors when the data are not taken along the weighted reference point. The lower interferogram also shows the effect resulting from the squirming of the interferometer fringes that is caused by changes in the thickness of the air films between the specimen and interferometer plates at the unweighted reference points. It is these air-film effects that produce the irregularities that are obvious errors in the data from which curve 3 was computed. The effect of a change in the film speed, while heating the interferometer at a constant rate, is shown near the right-hand end of this interferogram. The upper set of fringes in each interferogram are the temperature fringes. The white numbers inserted in the interferograms were written on the negatives to aid in the fringe count when the data were translated from the interferograms to numerical values.

IX. PROPOSED REFINEMENTS

As a result of experience gained with the interferograph, a few changes in its design are proposed. Although not affecting the performance, these changes incorporate certain desirable features. For example, the constant-deviation prism described above is quite massive and although no difficulty has been experienced, there is a possibility of injury to it due to heat shocks when measuring expansions to very high temperatures. To reduce this danger a new prism has been designed in which all dimensions except thickness have been reduced to one-half, thus reducing the mass to one-fourth that of the original prism. This reduction in the size of the prism does not reduce the dimensions of the aperture in the directive parallel to the exposing slit. The visible field is limited by the collimating lens and the two horizontal edges of the entrance face of the new prism. Thus the visual field has the general shape of the fringe field shown in figure 1, *C*, with the two straight sides separated by an amount equal to the radius of the collimating lens.

It is desired to call attention to the difficulty sometimes experienced in transferring the interferometer from the mounting tripod to the furnace. This is minimized by first transferring the interferometer from the tripod to a small cup fitted with a bail-like handle, and then lowering it into the furnace by means of a hook. Even so, the operation frequently involves several unsuccessful trials. Therefore, in place of this procedure it is proposed to modify the furnace so that the mounting of the interferometer may be done directly in the posi-

tion it will occupy during the observations. This may be accomplished in a number of ways. The furnace may be constructed so that one-half may be removed while the interferometer is being set up, and then replaced. Or the interferometer may be supported on a pillar independent of the furnace and the furnace elevated and swung out of line while adjustments are made.

X. CONCLUSIONS

The experimental instrument described herein was found to perform excellently. The results from a large number of tests have conclusively demonstrated the adequacy of its mechanical performance. An outstanding advantage of the interferograph over visual instruments is that the recorded interferograms are permanent records that are available for repeated critical reviews of their interpretation. The small amount of attention required for the operation of this instrument allows the operator to prepare new specimens for later tests and to compile the data from previous tests during the photographic recording.

XI. REFERENCES

[1] F. C. Nix and D. MacNair, Interferometric-dilatometer with photographic recording, Rev. Sci. Instr. **12,** 66 (1941).
[2] A. Trowbridge, Photography of moving interference fringes, J. Opt. Soc. Am. **6,** 195 (1922).
[3] J. B. Saunders, Interferometer measurements on the expansion of iron, J. Research NBS **33,** 75 (1944) RP1597.
[4] C. G. Peters and C. H. Cragoe, Measurements on the thermal dilation of glass at high temperatures, Sci. Pap. BS **16,** 457 (1920) S393.
[5] G. E. Merritt, The interference method of measuring thermal expansion, BS J. Research **10,** 59 (1933) RP515.
[6] Ph. Pellin and A. Broca, Spectroscope à déviation fixe, J. phys. **8,** 314 (1899).
[7] A. Pérard, Quelques études particulières au dilatomèter Fizeau, Travaux et Mémoires du Bureau International des Poids et Mesures, XIX, 21 (1932).
[8] C. H. Cartwright, Treatment of camera lenses with low reflecting films, J. Opt. Soc. Am. **30,** 110 (1940).
[9] A. Perot and Ch. Fabry, Méthodes interférentielles pour la mesure des grandes épaisseurs et la comparison des longuers d'onde, Ann. Chim. phys. **16,** 289 (1899).
[10] M. Luckiesh, L. L. Holladay, and R. H. Sinden, An interference thermometer and dilatometer combined, J. Franklin Inst. **194,** 251 (1922).
[11] J. B. Saunders, Improved interferometric procedure with application to expansion measurements, J. Research NBS **23,** 179 (1939) RP1227.
[12] W. H. Johnson and W. H. Parsons, Thermal expansion of concrete aggregate materials, J. Research NBS **32,** 101 (1944) RP1578.
[13] A. A. Michelson, On the application of interference methods to spectroscopic measurements, Phil. Mag. [5] **34,** 280 (1892).
[14] F. J. Bates and F. P. Phelps, A suggested new base point on the thermometric scale and the $\alpha \rightleftarrows \beta$ inversion of quartz, BS Sci. Pap. **22,** 315 (1927) S557.
[15] W. F. Meggers, Notes on comparisons of lengths of light waves by interference methods, and some wave lengths in the spectrum of Neon gas, Bul. BS **12,** 203 (1915) S251.
[16] A. Q. Tool, D. B. Lloyd, and G. E. Merritt, Dimensional changes caused in glass by heating cycles, BS J. Research **5,** 642 (1930) RP219.
[17] M. A. Arnulf, Un appareil enregistreur pour la mesure des dilations des verres, Rev. optique **3,** 270 (1924).
[18] R. H. Sinden, An interferential dilatometer employing automatic photography, J. Opt. Soc. Am. **15,** 171 (1927).

WASHINGTON, May 10, 1945.

Journal of Research of the National Bureau of Standards Vol. 47, No. 3, September 1951 Research Paper 2239

Precise Topography of Optical Surfaces[1]

James B. Saunders

The unit of length usually used in measuring optical surface features, by interference of light, is one-half the wavelength of the monochromatic light that is used. A method is described in which the unit of length is a much smaller fraction of the wavelength. Topographic maps of optical surfaces are made in which the contour interval is less than one-thirtieth the wavelength. The smallness of the unit depends upon the quality of the surfaces. Irregularities, too small to be detected with the Fizeau or two-beam fringes, are made to appear very prominent. Surface markings, caused by the final polishing actions, can be made to stand out in bold relief.

The accuracy of results that may be obtained with any type of interferometer depends upon the accuracy with which one can locate a line in the fringe field, along which the order of interference is constant. The intensity of light in a direction normal to these lines, in a set of two-beam or Fizeau [2] type of interference fringes, is a smooth mathematical function without abrupt changes, whereas for multiple beam or Fabry-Perot [3] type fringes it has sharp maxima and minima for transmitted and reflected light, respectively. The accuracy attainable with the two types of fringes is roughly proportional to the ratio of fringe separation to fringe width, the fringe width being defined here as the width at one-half maximum intensity.

Accurate contour maps of one surface relative to a standard surface of known shape are obtained by photographing the interference fringes, produced by these two surfaces, with monochromatic light. The contour intervals are one-half wavelength. As in geodesy, the smaller the contour interval the more detail is revealed in the contour map and the narrower the lines the more detail is revealed along the line. With broad fringes, of the two-beam type, irregularities of several hundredths of a wavelength can be completely hidden in or between the dark broad lines. An irregularity of several tenths of a wavelength can be completely invisible if it falls within the area between two adjacent Fabry-Perot fringes that appears uniformly illuminated. The one-half wavelength interval is, therefore, inadequate for representing local irrregularities on high-quality optical surfaces.

This paper describes a procedure whereby very narrow fringes are used for producing contour maps of optical surfaces with contour intervals of a small fraction of a wavelength. These intervals can be made as small as the narrowness of the fringes will permit. If m resolved fringes can be equally spaced between two of the original fringes, the separation of which corresponds to one-half wavelength or to one order of interference, then the contour intervals will be $[1/2(m+1)]\lambda$. Values of m as large as 19 have been used, for which the intervals are $\lambda/40$.

The usual procedure in precision measurements on glass optical surfaces is to place the unknown close to and almost parallel to a standard surface of known shape, with a thin air wedge between, and to observe the fringes produced by the interference of monochromatic light reflected normally from these two surfaces. Figure 1 shows a set of two-beam interference fringes, observed by reflection with a Fizeau viewing instrument, which one might have under observation for the purpose of measuring the deviation of an unknown surface from a standard plane surface. The orders of interference corresponding to each fringe, although not usually known, are indicated by the numerals. The exact position of the lines that represent the loci of integral orders of interference, are only known to lie somewhere within the darker portion of the fringes. This locus for fringe number 3 is indicated by the narrow solid line. Usually, as will be shown later, these lines are not as smooth [4] as the much broader fringes appear to be. The assumption that this line lies at the center of the dark fringe is usually false and, consequently, results based on this assumption are in error by the amount of displacement of the line from the center; which can amount to over a tenth of the separation between adjacent fringes. Even if one could locate the lines of integral orders, he would still be unable to measure the change in order of interference along a straight line such as the one shown dotted in figure 1, for the purpose of surface measurement. The change in order of interference along this straight line is an accurate measure of the deviation of the unknown surface from a plane, but if the shape of the surface between the fringes is unknown one has no way of determining the fractional orders of interference at points on this straight

[1] This work was done as part of a research project sponsored by the Air Force. The paper was presented at the March 1950 meeting of the Optical Society of America (Abstract, J. Opt. Soc. Am. **40**, 258 (Apr. 1950)). Since then the author's attention has been called to a paper by Takemara Sakurai and Kôrô Shishido at the Research Institute of Scientific Measurements, on "Test of an optical flat by the Fabry-Perot etalon" and published in the Scientific Report of the Research Institute, Tôhoku University, [A], **I**, No. 1 (May 1949).

[2] Fizeau. Compt. rend. **54**, 1237 (1862).

[3] C. Fabry and A. Perot, Ann. Chim. Phys. **12**, 459 (1897).

[4] The irregular serrations along the edges of the fringes are a good indication of the shape of the lines of constant orders. In fact, the location of lines of constant order (although not integral) are most accurately locatable along the margin of the fringes where the density gradients are a maximum. This is the most accurate means, known to the author for locating lines of integral differences in order of interference for surface measurements.

FIGURE 1. *Two-beam interference fringes.*

line that are not also on the lines of integral orders. Linear interpolation, for points not on a fringe, can be in error by several tenths of an order. The measurement of deviations of a surface from a plane, by measuring the change in order of interference along this line, is in error both from the inability to locate the lines of integral orders of interference and also from interpolation based on the false assumption that the order of interference varies linearly with the displacement of the point from the adjacent fringes. Unless a fringe passes through the center of a surface, the deviation of the surface at this point, relative to others, cannot be accurately measured.

The accuracy of measurements on the deviations of the unknown surface from an arbitrarily chosen reference plane, along any one of the lines of integral orders such as the narrow solid line in fringe 3, is limited only by the accuracy with which the line can be located. The deviation of the surface at any point on this line, from a plane through the indicated straight line (which is coincident with neither of the reflecting surfaces but intersecting the unknown surface at the two ends of fringe 3) and making an angle θ with the standard reference plane, is equal to the product of θ and the perpendicular distance of the point to the straight line where θ equals one-half the wavelength of the light used divided by the fringe separation. The fringe separation is defined as the average distance between the centers of adjacent fringes and is the separation that would be obtained if the unknown was replaced by a true plane, placed in the position of the arbitrarily chosen reference plane. The value for θ should be constant for all computations, although the observed fringes may neither be parallel nor equally spaced. The use of different values for fringe separation is equivalent to measuring deviations from different planes at various angles to the standard.

If both the standard and the unknown are coated with a uniformly thin coat of highly reflecting metal, such as silver or aluminum, the fringes in which the lines of integral orders of interference are located become narrower. The narrowness of these fringes, relative to their separation, increases with the reflectance of the surfaces and also with decreasing angle of incidence. The decrease in width of these narrow line fringes, relative to their separation, increases the accuracy of locating the lines of integral order of interference. With modern methods of metal deposition and with smooth surfaces that are almost parallel to each other, fringes may be obtained whose width, at one-half the maximum intensity, are only one-fiftieth part of their separation.[5] With such narrow fringes, deviations can be measured to better than one-hundredth of a wavelength. These narrow-line or Fabry-Perot type fringes are dark by reflected light and bright with a dark background, with transmitted light. Figure 2, A, shows a set of two-beam fringes produced by reflected light, and B shows a graph of multiple reflection fringes, formed by the same surfaces by transmitted light after applying the thin coat of aluminum. Note the increased accentuation of detail along these narrow fringes over that shown by the broad ones.

When measuring surface features to one-hundredth of a fringe, the mechanical errors in the micrometer (assuming a micrometer eyepiece to be used) and personal errors of settings become large unless the changes in order of interference with movement of the cross hair, are relatively small. This requires that the fringes have considerable separations. At least two fringes must appear on the surface in order to determine the unit of deviation of a fringe from a straight line corresponding to a known unit of deviation of the surface from a plane. This unit is usually one-half the wavelength of the monochromatic light being used. Figure 3, A, shows such a set of fringes, formed by the same surface that was used for figure 2, B. Further enhancement of detail along the fringes in figure 3, A, over those of figure 2, B, is obvious. This is partially due to the decrease in width of the fringes relative to their separation, caused by making the wedge between the plates smaller and, consequently, reducing the angle of incidence that favors the narrowing of these fringes.

In order to further enhance irregularities along a fringe and still be able to measure the magnitude of the corresponding deviations in the surface, a multiple spectral line source, such as the mercury arc for instance, may be used. Each spectral line produces its own set of fringes. If the absolute orders of interference of each and all of the fringes are known, the separation of the surfaces along the fringes is known and, consequently, the change in separation of the surfaces between two adjacent fringes of different colors is also known, although this difference in separation may be only a small

[5] S. Tolansky, Multiple-beam interferometry of surfaces and films (Oxford at the Clarendon Press, 1948).

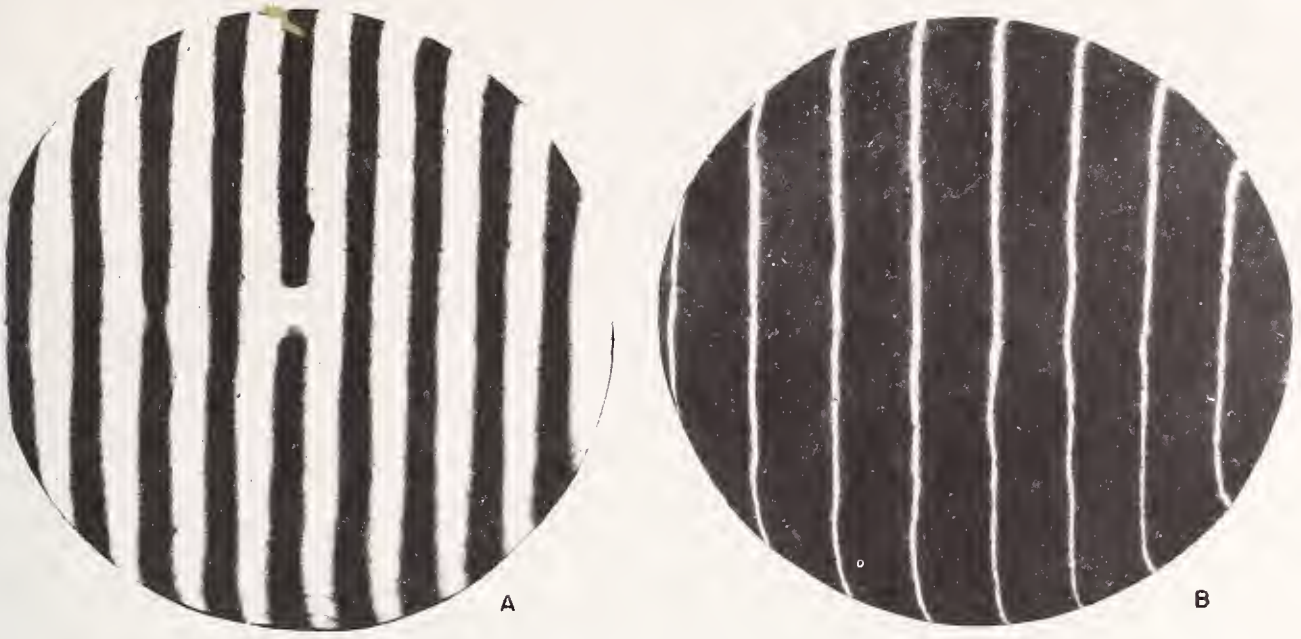


FIGURE 2. *Equal-thickness interference fringes.*

A, Two-beam fringes produced by reflected light; B, Multiple-beam fringes produced by transmitted light.

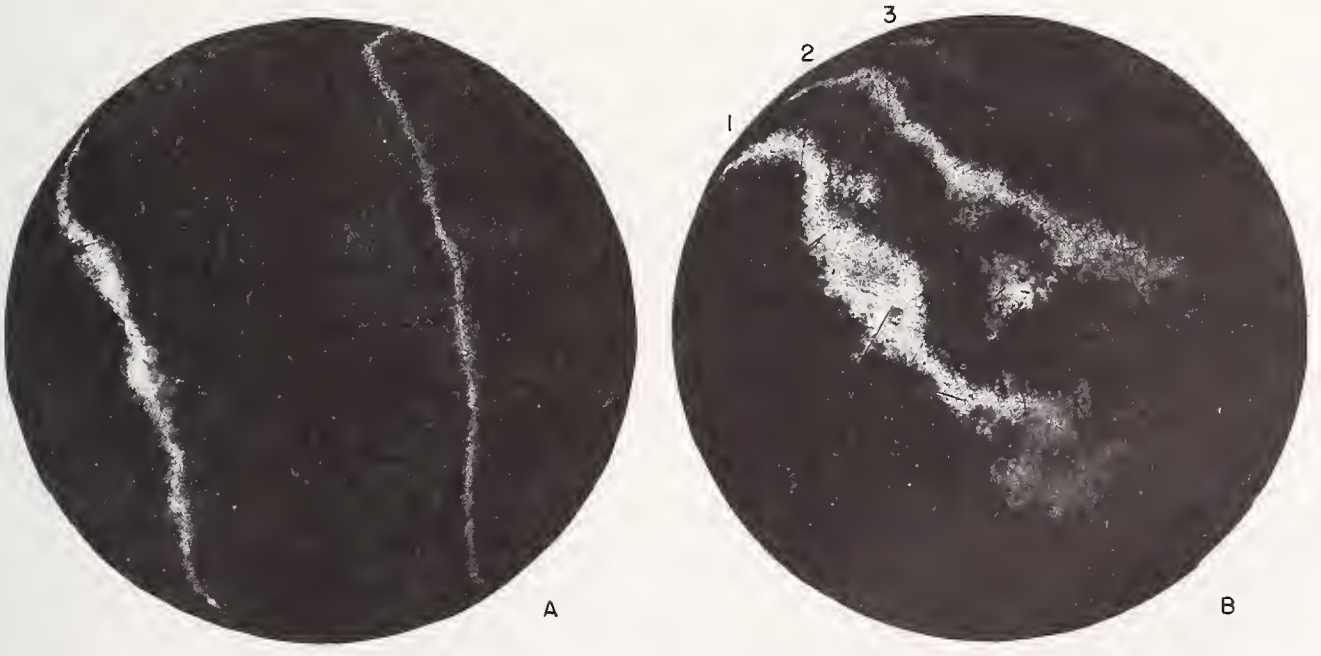


FIGURE 3. *Multiple-beam interference fringes.*

A, Single spectral line. Note increase enhancement of surface detail. B, Multiple spectral line. One order of each of three lines shown on plate. Wedge too small to include other lines. Note further enhancement of surface detail.

fraction of the unit of length, one-half wavelength. The thickness and angle of the air wedge between the surfaces may now be adjusted so that one of two or more closely spaced fringes of different color pass through the center of the surface and one or more of the others just inside the edge. The surfaces are then as near parallel as is practicable, thus favoring the attainment of the narrowest fringe possible, relative to fringe separation, and measurements may be made at all radial distances. Still greater enhancement of the irregularities is thereby attained, and two or more fringes are available for quantitative evaluation of these irregularities. Figure 3, B, shows a set of fringes, produced by a mercury source, in which the total change in order of interference across the plate is less than three-tenths of a unit ($\lambda/2 = 0.27\mu$).

Maximum enhancement of surface irregularities is obtained by adjusting the surfaces to as near parallelism as is possible and the separation near to some integral number of half wavelengths. If one had two plane reflecting surfaces (the existence of which is doubted by the author) these could be placed parallel to each other, which is the condition for fringes of infinite width. If the separation is now adjusted to some value in the neighborhood of where the resultant interference produces maximum variation of intensity with change in separation, then small deviations in phase, caused by irregularities of only a few angstroms, such as may be produced by the individual particles of the polishing compound, become quite visible. Figure 4 shows the above-mentioned surfaces (actually both surfaces contribute equally) in which the parallelism and separation are most favorable to the revelation of small surface detail. Note the fiber-like appearance in the darker regions where maximum enhancement of detail occurs. Quantitative measurements of these finer irregularities have not yet been attempted.

The manner of determining absolute orders of interference may be described by means of figure 5 The large circularly enclosed fringe system represents a set that might be under observation. The colors of the fringes are indicated by R for red, Y for yellow, G for green, B for blue, and V for violet. The wavelengths of the light, forming the fringes, are identified usually by the color of the corresponding fringes. The small circularly enclosed fringe system represents the same fringes as those shown in the large circle. They are straight as if formed by parallel surfaces and are inserted to show the relationship between the fringes in the large circle and the lines, which correspond to fringes, in the chart above. If we assume the surfaces that form this fringe system to be extended in the direction of their decreasing separation to their line of intersection, and one could observe the fringe system in this extended area, one would find and could identify the zero order of interference for each set (color) of fringes, all of which coincide. Extending from the zero order fringe, one would observe equally spaced red fringes (assuming plane surfaces), the order of which could be identified by counting from the zero order.

FIGURE 4. *Multiple-beam interference fringes.*

Surfaces are as near parallel as the surfaces will permit (maximum enhancement of detail). Note fabric-like marks; believed to be due to final polishing action.

Thus the numbers associated with the vertical lines indicates the order of interference for all the visible colors in the source, and their positions correspond to the positions of the fringes. One sees on this chart, in the neighborhood of 3.3-μ separation of the plates, that several fringes fall in a space corresponding to 0.12-μ increase in separation of the plates. This space corresponds to a change of only 0.36 in the order of interference for this red spectral line. If the maximum and minimum separations of the plates are adjusted to 3.2 and 3.4 μ, respectively, the particular set of fringes that correspond to the lines falling in this range on the chart can be seen on the plates. We therefore have a fringe system in which the increase in separation of the surfaces from one fringe of a given color to another fringe of a different color is a small fraction of the conventional unit, $\frac{1}{2}\lambda$, that is usually used for this type of measurement. Furthermore, the width of the bright line fringes, relative to separation of adjacent orders, is made smaller by the two surfaces being made nearer parallel than is practical when using monochromatic light. The narrower the fringes, relative to separation of adjacent orders, the more concentrated a cluster or grouping of fringes can be used, without overlapping. Figure 3, B, shows a constellation or group of fringes whose separations correspond to an increase in separation of the surfaces of 270 A from fringe 1 to 2, and 225 A from 2 to 3.

When an optical flat is placed on an unknown surface with only an air film separator, the weight of the top plate forces the air film to become relativerly thin, but the plates will not come into absolute contact at any point. Usually, small particles of dust will prevent the air film from becoming as

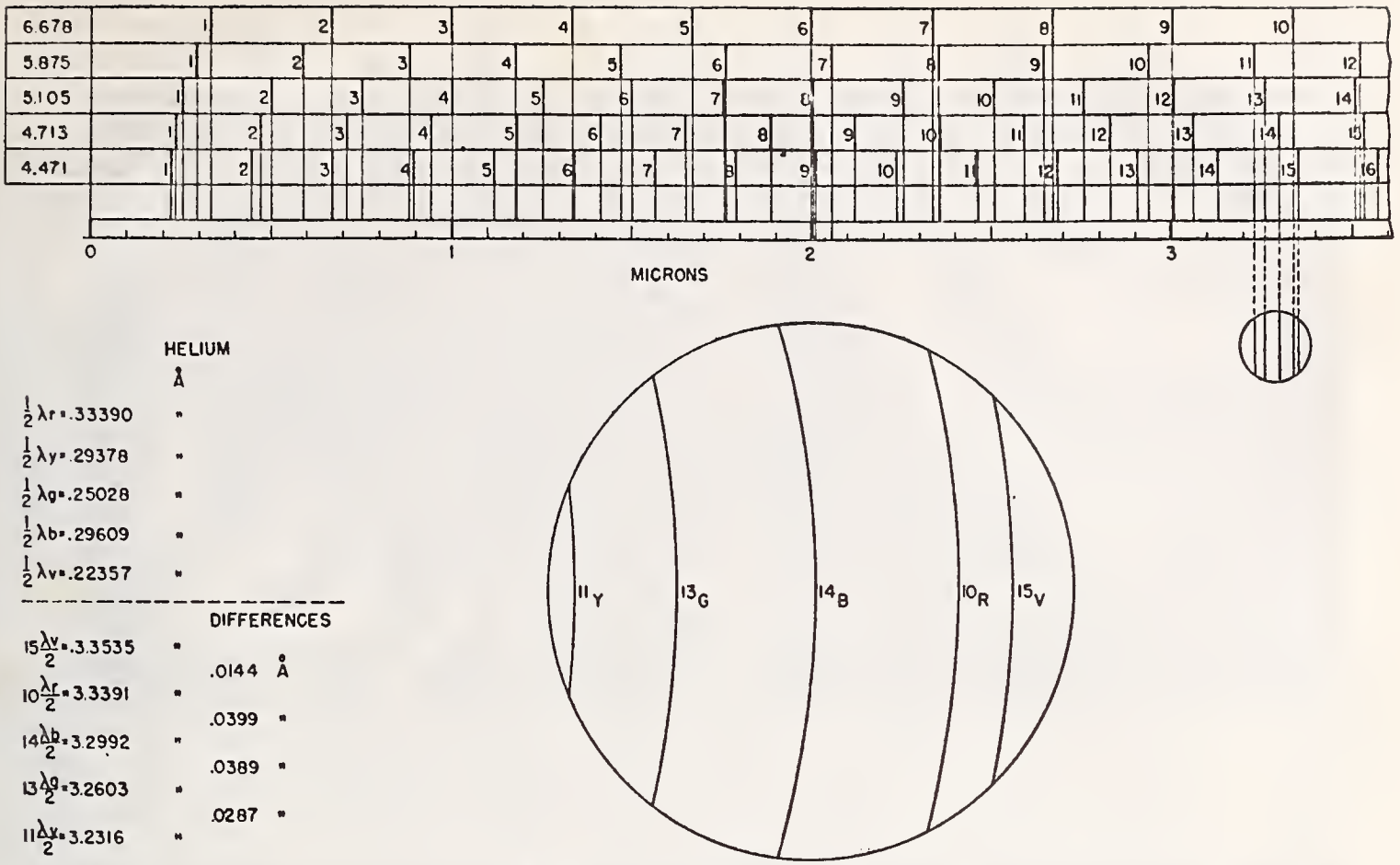


FIGURE 5. *Chart showing how the absolute order of interference is identified from the relative distribution of colored fringes produced by a multiple line source.*

thin as would be attained in their absence The usual range in thickness, for circular surfaces of 4-in. diameter and free from dust, is from 8 to 9μ. For larger surfaces this range may extend to 10μ. The range, 0 to 3.6μ, shown in figure 5 is not practical but was used for explanatory purposes only. The practical range is from 6 to 9μ and is shown in figure 6 for several different light sources. If one has access to different sources he may choose a constellation most suitable for a given film-thickness range.

A more elegant topographic map of an optical surface, relative to a standard, can be obtained by using monochromatic light and photographing the image of a given fringe in several different positions on a single film; the shift in the fringe system being accomplished by changing the optical separation of the surfaces. There are several different ways of changing the optical separation, only one of which will be described in detail. The interferometer plates are enclosed in an airtight chamber (see fig. 7) in which the pressure of the enclosed gas may be controlled at will. The collimator and collector lenses form the windows to the chamber. A gas line, connected to a manometer through a two-way stop cock, permits either compression by a compressor, or evacuation by a vacuum pump. A needle valve in the line, not shown, facilitates control of the pressure in the chamber.

According to the law of Gladstone and Dale, the index of refraction of a gas varies linearly with the density. For a constant temperature of 25° C, which was maintained during this experiment, the density is directly proportional to pressure, to a close approximation. The relation between refractive index and pressure, is $n-1=P(n_0-1)/760$ where n is the refractive index of the gas, P is the pressure in millimeters of mercury, and n_0 is the refractive index of the gas at 760 mm pressure and at 25° C.

If Δn is the change in n for a given change of ΔP in P, it may be shown that the corresponding change in order of interference, ΔN, at a point where the separation of the surfaces is L, is

$$\Delta N=\frac{2L\Delta n}{\lambda}=\frac{2L}{\lambda}\frac{(n_0-1)}{760}\Delta P.$$

From this relation one may compute the separation, L, that is necessary in order to get a chosen change in order of interference for any desired pressure range. For instance, suppose the available pressure range to be 1 atm or 760 mm of mercury and the desired change in order of interference to be one (the maximum practical change for this work), then $\Delta P=760$, and $\Delta N=1$. Consequently,

$$L=\frac{\lambda}{2(n_0-1)},$$

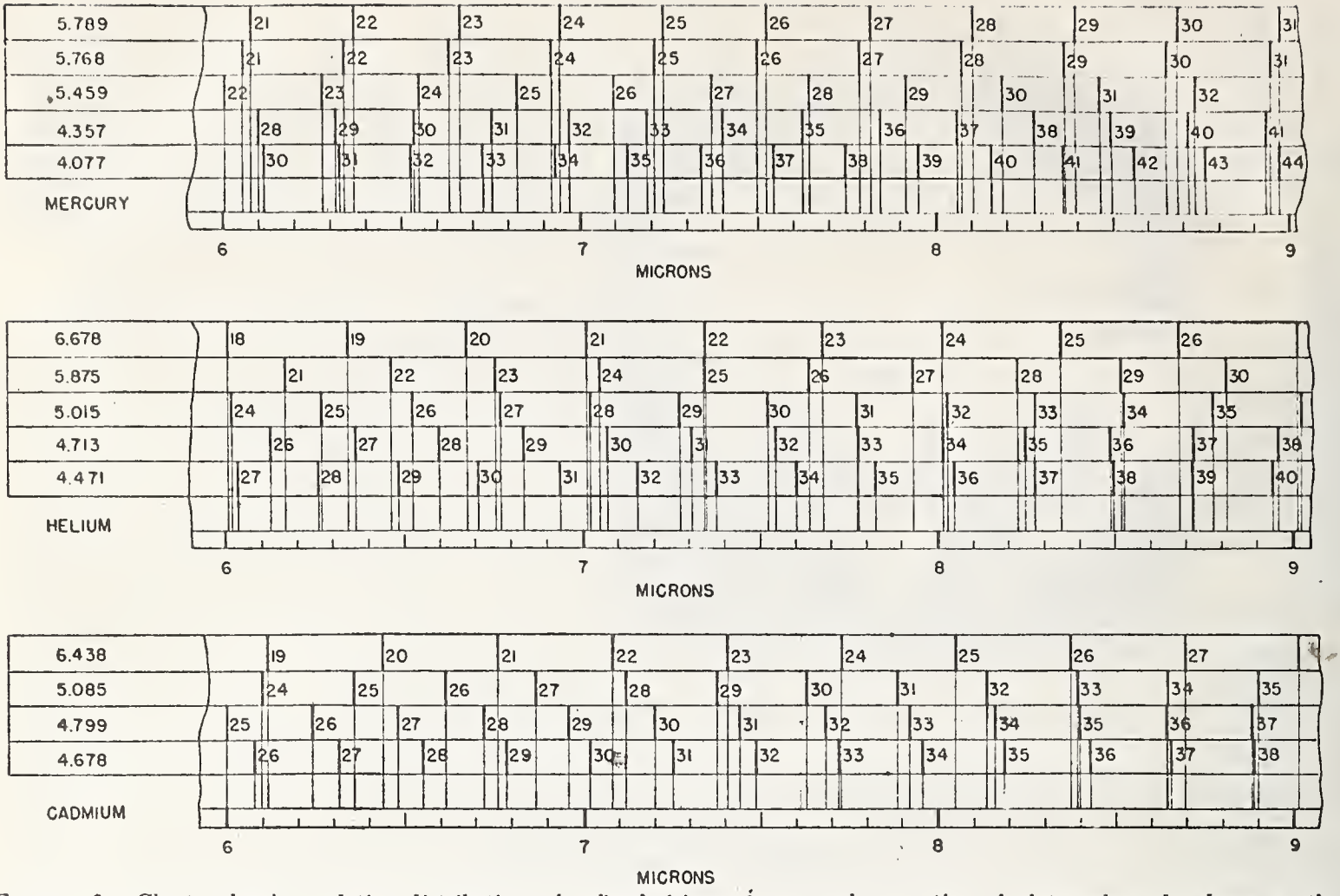


FIGURE 6. *Charts showing relative distribution of colored fringes for normal separation of plates when placed one on the other for test.*

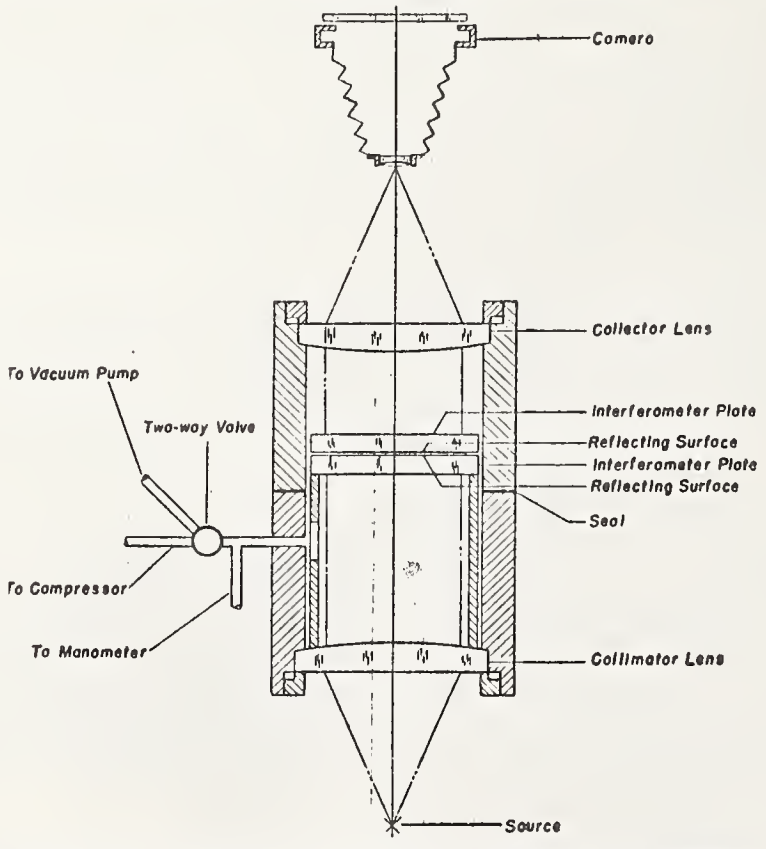


FIGURE 7. *Vacuum chamber used to change optical separation without changing the geometrical separation of two interferometer plates.*

FIGURE 8. *Topographic map of one optical surface, relative to a standard, produced by photographing a given fringe (order of interference) successively, as it was moved across the plate, by controlled evacuation of air wedge between the plates.*

Total movement corresponds to a change of one-half order of interference.

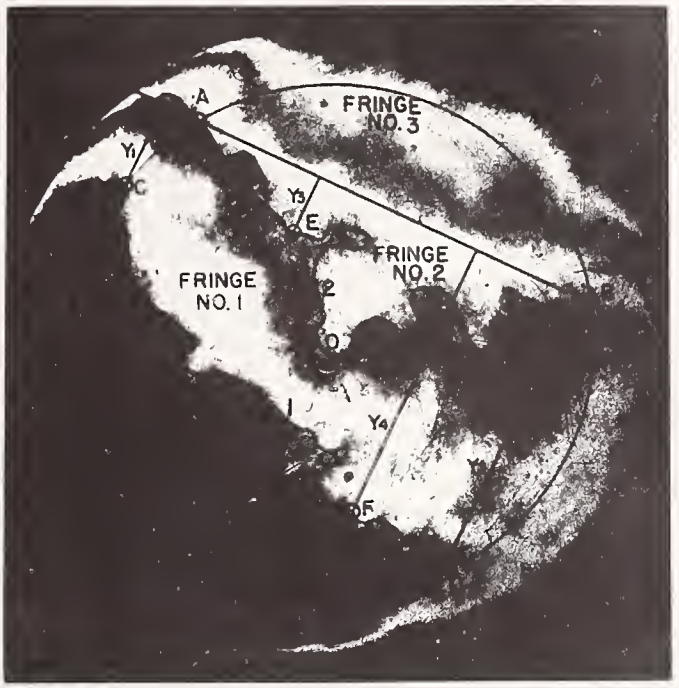


FIGURE 9. *Same photograph as in figure 3, B, with inserted drawing, for surface measurement.*

If, for instance, the light to be used is the green line of mercury, for which $\frac{1}{2}\lambda = 2.7 \times 10^{-5}$ cm and $n_0 - 1$ is 2.8×10^{-4}, then L equal 0.96 mm. Consequently, with this separation, pressure range and light source, a fringe can be made to take on any position between the positions of this fringe and that of one of its adjacent neighbors at the extremes of the pressure range. If there is more than one fringe on the plate and one wishes to interpolate nine additional, equally spaced fringes, between each of the adjacent original fringes, thus completely filling the photographed field with fringes whose separations are one-tenth that of the original set, then it will be necessary to change the pressure in steps of 76 mm for nine times, or a total of nine-tenths of an atmosphere. If the wedge between the surfaces is such as to give a maximum change in order of interference of one-half, then one-half-atm change in pressure is sufficient to move the center of a fringe completely across the surface. If, for this wedge, one wishes to photograph 10 fringes on the plate, the necessary change in pressure will be nine-twentieths of an atmosphere. It is, of course, essential that this pressure range be within that available, which in this particular case is 0 to 760 mm.

Figure 8 represents a topographic map with contour intervals of approximately 170 A, or one-thirtieth of a wavelength of cadmium green light. The separation of the two surfaces was 0.82 mm, and the change in separation, in terms of interference fringes, was approximately one-half fringe (1,300 A).

To illustrate the application of these smaller-than-one-half-wavelength-unit topographic maps to quantitative determination of surface features, measurements have been made on the deviations of one of the surfaces used to produce figure 3, B, relative to the other. For this discussion one surface will be assumed a true plane (called the reference surface) and the other (called the unknown) to have all the irregularities corresponding to the irregularities shown in this fringe pattern. Lines and points have been inserted in figure 3, B, and reproduced as figure 9.

The deviations of the unknown from a true reference plane, for several points along the lines AEB and CFD have been computed and are plotted in figure 10. The chosen reference plane is an imaginary plane that passes through points A and B (consequently including all points on the straight line AB) and making an angle θ with the reference surface approximating the average angle between the two reflecting surfaces. The unknown surface and the chosen reference plane are made as near parallel to each other as the irregularities in the unknown will permit. The value of θ is obtained from circle $ABDC$, which is concentric with respect to the center of the plates, the straight line AB, and the perpendicular projections of points C and D on line AB. Circle $ABDC$ must intersect at least two locatable lines, such as AEB and CFD, along each of which the order of interference is constant, and between which the change in separation of the reflecting surfaces is known. The line AEB is the lower edge of fringe 2, and line CFD is the lower edge of fringe 1. The change in separation of the plates between these two lines is approximately equal to that between the centers of the corresponding fringes

and, as was given previously, equal to 270 A. The magnitude of θ is taken to be $2d/(Y_1+Y_2)$, where d equals the change in separation of the plates between fringes 1 and 2, Y_1 equals the normal distance of the point C from the straight line AB, and Y_2 equals the normal distance of point D from AB. This choice for θ is entirely arbitrary and in no way affects the computed shape of the unknown surface.

To correlate the data shown in figure 10 with the topographic map (fig. 9) from which it was derived, the designated points in figure 9 are labeled likewise in the graph. Thus, the ordinate of point E in figure 10 is the deviation of the unknown at point E of figure 9. The lower branch of solid curve $2R$, figure 10, represents deviations along the right-hand portion of line 2 figure 9, and the upper branch, $2L$ having the circled points, corresponds to the left-hand portion of this line.

These measurements are not limited to points along a single fringe or line, but may be made at any point along any fringe that appears on the photograph. The deviation of the unknown from the chosen reference plane, at a point on any other line, such as point F on line 1, is equal to $Y_4\theta - Nd$, where Y_4 is the normal distance between point F and line AB, and N is the difference in order of interference between F and any point on line 2. Thus, measurements may be made at all points along each and every contour line shown.

The plot shown in figure 10 does not give any better idea of the shape of the unknown surface than the contour map from which it was derived. The values are, however, better defined numerically.

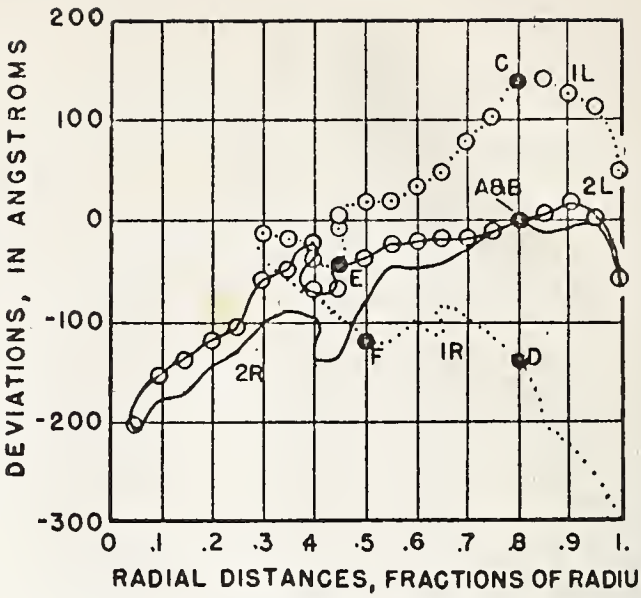


FIGURE 10. *Radial distances, fractions of radius.*
Radius equals one-half diameter of plates.

The failure of all computed points to fall on a single smooth curve indicates that the unknown surface is not a figure of revolution about its center. The appearance of more than two values of a given radial distance and on a single line is due to local depressions or elevations. The irregular shape of the curves also indicates the degree of local irregularities in the surface. These local irregularities represent deviations approximating 50 A, in some cases.

WASHINGTON, January 2, 1951.

U. S. GOVERNMENT PRINTING OFFICE: 1951

For sale by the Superintendent of Documents, U. S. Government Printing Office
Washington 25, D. C. - Price 10 cents

Reprinted from JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, Vol. 44, No. 3, 241–242, March, 1954
Printed in U. S. A.

In-Line Interferometer

JAMES B. SAUNDERS
National Bureau of Standards, Washington, D. C.
(Received October 28, 1953)

An interferometer is described for the measurement of beam dividing surfaces. The apparatus consists essentially of the elements for a Twyman-Green interferometer. The end mirrors are made of optical glass and the reflecting surfaces of these mirrors transmit a part of the beam. The advantages of the instrument are (a) a sensitivity $(n+1)$ times that of the Fizeau viewer, which is usually used for this test, (b) compactness, and (c) ruggedness.

WHEN an optically plane surface is coated with a partially reflecting film and used as a dividing plate in an interferometer of a Michelson or Twyman-Green type, one has a more sensitive test of the planeness of the surface than is available in Fizeau methods that are generally applied in an optical shop for the production of plane surfaces. The reason for this is that, if the beam-divider surface is not flat, one of the resulting wave fronts emerging from the interferometer—after one reflection and one transmission at the dividing surface—will be concave and the other will be convex. With the usual Fizeau method of testing, the curvature of fringes corresponds to two interfering wave fronts differing by only twice the curvature of the surface being tested. The difference in sensitivity is so great that plates tested in an optical shop by the Fizeau method and appearing to be of satisfactory planeness may be quite unsatisfactory when mounted in a Twyman-Green interferometer to serve as a beam divider. Because of this recognized difficulty it was an earlier practice at the National Bureau of Standards to test a dividing plate periodically during its figuring by actually assembling it in a Twyman-Green interferometer system. This method yielded satisfactory results but the space required for a 10-inch Twyman interferometer was unavailable in the shop, necessitating the removal of the plate to another location for test; also the foreshortening of aperture resulting from the large angle of incidence and the thickness of the plate was troublesome.

It became apparent that an interferometer of a compact type with full circular aperture, which could be set up on an optical bench in the shop, would be most useful as a means for critically testing optical surfaces while they are being figured. These requirements lead to the "in-line interferometer," or "series interferometer"[1] as D. Post of the U. S. Naval Research Laboratory has named it. The optical system shown in Fig. 1

[1] D. Post, J. Opt. Soc. Am. 44, 243 (1954).

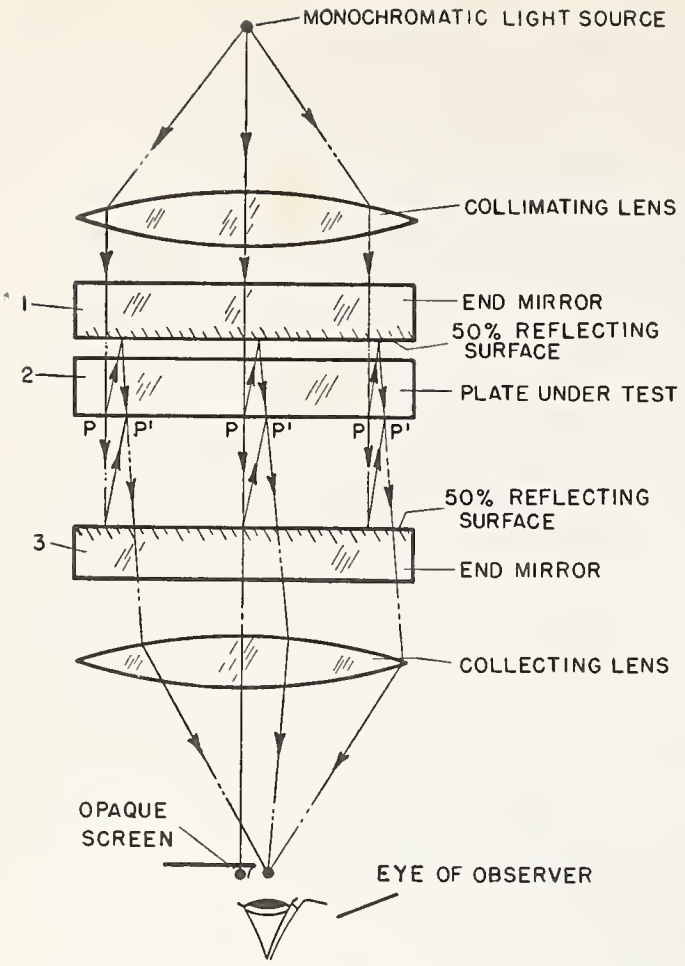


FIG. 1. In-line interferometer for precision testing of beam dividers.

may be roughly described as equivalent to two Fabry-Perot interferometers arranged in tandem with the two adjacent plates combined into one, or a Twyman or Mach-Zehnder with zero angle of incidence at the divider.

The optical arrangement consists of a collimating lens (see Fig. 1); an end mirror, No. 1, with the aluminized or silvered surface plane and the other surface clear but not necessarily plane; the specimen or beam dividing plate, No. 2, having both surfaces plane; a second end mirror, No. 3; a collecting lens and a screen to intercept all beams except the one used. Mirror No. 3 is adjusted normal to the beam so that it returns the reflected light into the source. Plate No. 2 (the beam dividing surface) is adjusted to make a small angle θ with mirror No. 3. This produces a series of images of the source in the focal plane of the collecting lens. Of this series the image by the directly transmitted beam is one member. A second image is the result of two reflections, and the remaining images correspond to multiples of two reflections. In Fig. 1 the first image of this series is shown on the optic axis and the second is indicated as the position of the observer. If mirror No. 1 is adjusted to make an angle 2θ with mirror No. 3, two images of the source become superimposed at the observing position. The paths of the rays forming these two images are shown in Fig. 1. When the two optical paths from P to P'—one reflected from each of the end mirrors—are nearly equal, interference fringes may be seen. Points P and P' are shown here to be displaced relative to each other in order to show the course of the light rays.

For visual testing of a beam-divider plate by this method it is not necessary to apply high-reflecting coats to its surface. However, when maximum intensity is desired, $\frac{2}{3}$ transmitting, $\frac{1}{3}$ reflecting (non-absorbing) films should be applied to the end mirrors and a $\frac{1}{2}$ transmitting, $\frac{1}{2}$ reflecting film should be applied to the beam dividing surface.

If mirror No. 3 is made similar to the beam divider with the reflecting film placed on the back face and the three plates are equally spaced, then white light may be used.

A 10-inch interferometer of the above description was assembled from parts that were available and used for testing divider plates as early as July 1950. Its use was found to show many advantages over the former practice of assembling the plate in an interferometer of the conventional type. Such advantages include the following:

(a) For a Twyman-Green interferometer with a beam-divider plate of refractive index n, the change in order of interference caused by a given departure (δ) from planeness is 2 $(n+1)\delta$, whereas that caused by the same plate when tested with the Fizeau method is 2δ.

(b) The plate is tested at nearly normal incidence. It therefore is not foreshortened and curvature of a fringe is directly interpreted in terms of departure from planeness, measured normal to the surface.

(c) The interferometer is so compact that it is not seriously affected by vibration or distortion of bed arising from imperfect support. This enables the instrument to be conveniently mounted in the workshop where the figuring of the surface is proceeding.

(d) The instrument is more readily and quickly adjusted than the usual interferometer because of the greater rigidity of the very short bed and also because, even for an interferometer of relatively large aperture, the adjusting screws can be brought within reach of the observer.

A more finished instrument of the in-line type is being constructed. It will be mounted on a conventional optical bench and the critical parts will be enclosed in a temperature-controlled space.

Pressure of other work since the completion of the first instrument at the National Bureau of Standards has limited the application of this interferometer. Post has recognized its value for use in photoelastic work and has independently developed a theory of it which is presented in the following paper. I am most grateful to him for having recognized an additional application for this instrument, for giving a development of its theory, and for his courtesy in communicating with me prior to publication of his work.

Journal of Research of the National Bureau of Standards Vol. 53, No. 1, July 1954 Research Paper 2514

Testing of Large Optical Surfaces With Small Test Plates

James B. Saunders

A procedure for testing large optical surfaces with relatively small optical standards is described. Simplified formulas are used to apply a statistical method for obtaining increased precision. A practical example is used to illustrate the procedure for testing surfaces that may be assumed to have revolution symmetry.

1. Introduction

With the increased use of massive optical parts, there is need for a practical method of testing large optical surfaces with relatively small standards, such as are available in the average optical shop. The conventional test of observing fringes between the standard and unknown at various positions on the surface enables the optician to form a rough estimate of the nature of the surface. For more precise values either extensions of the present conventional tests or new methods must be used. Present shop methods of testing optical surfaces are not practical for massive optical elements as standard test plates of the sizes required for conventional test methods are seldom available.

A test for shape of very large optical surfaces, by means of interference of light, may be made with a standard test plate (flat or spherical) that is much smaller than the surface to be tested. The principles involved are not new and may be considered elementary, but those who perform the tests on optical surfaces during the polishing are not as a rule sufficiently versed in mathematics to apply the principles outlined in the present paper to the best advantage. They can, however, apply observed data to a set of simple formulas and compute the ensuing results.

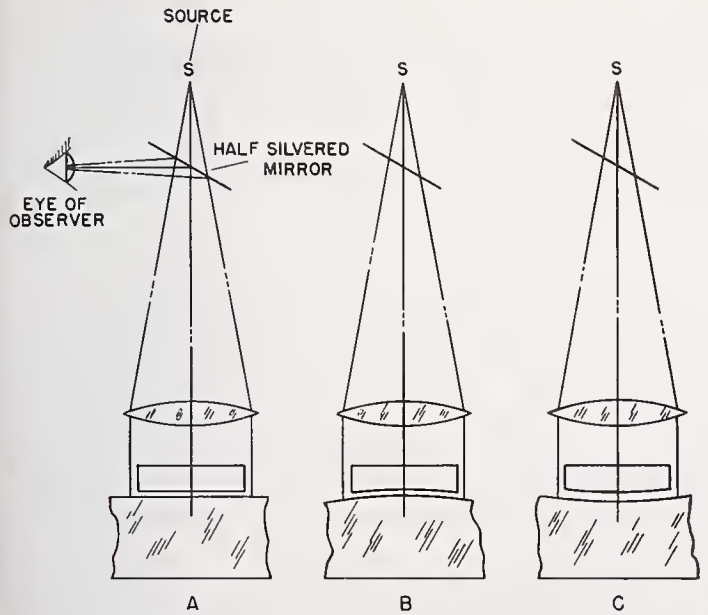


FIGURE 1. *Optical arrangements for testing (A) flat surfaces, (B) convex surfaces, and (C) concave surfaces.*

A procedure is described in which an ordinary-sized standard test plate can be used to measure the shape of surfaces that are much larger than the standard. Use is made of statistical methods for improving the computed results. Simplified formulas will be developed that may be applied by opticians, after a short training period, even though they may not be able to follow the mathematical derivations and theory. The application of the general formulas to a typical set of data will then be made to illustrate the testing of a surface by means of a smaller standard optical flat. The method is not limited to plane surfaces. The sphericity of very large spherical or aspherical surfaces may be measured with standard test plates of approximately the same radius, if the viewing system permits the use of normal incident light. Figure 1, (A, B, C) shows simple optical arrangements for testing plane, convex, and concave surfaces, respectively.

The basic principle may also be applied to the testing of off-axis curved, or nonsymmetrical, surfaces, with modifications of the formulas. However, as these are encountered only rarely, the present paper will be limited to surfaces of revolution.

We will consider a Fizeau viewer (fig. 1) that has an aperture equal to or greater than the area of the standard surface. Spherical surfaces are referred to the spherical master with which they are tested, and plane or approximately plane surfaces are referred to planes. The equations and measurements are identical for plane and spherical surfaces. This discussion will deal with the testing of an approximately plane surface against a standard optical flat. The unknown surface is assumed to be a figure of revolution about an axis normal to and passing through a known point on it. In general, this point will be at or near the center, if the surface is circular.

2. Experimental Procedures

If the standard flat is adjusted normally to the collimated beam of light and the unknown is placed close to and approximately parallel to the standard, interference fringes may be observed with monochromatic light of known wavelength. Measurements made on the fringe pattern permit a computation of the shape of the unknown relative to a plane (sphere for spherical surfaces) over the visible region of interference. If adjustments permitted, the entire area of the unknown could be covered by

518-743 O - 73 - 10

moving the standard step by step without rotation, in a plane parallel to the standard plane. The aggregate of the resultant fringe patterns, when properly assembled to form a composite pattern of the whole surface, would be quite similar to what would have been obtained with a standard surface that covered the whole of the unknown.

As it is virtually impossible to make the above assumed adjustments, corrections for rotation and changes in separation of the unknown, relative to the original surface of the standard, must be made by measuring these changes and applying the required corrections. The corrections are always subject to errors of observation. If several successive positionings of the standard, relative to the unknown, are required to measure the shape along a diameter, these errors accumulate. Consequently, it is desirable to reduce these errors. The "Method of Averages," [1] which is easy to apply and yields simple working formulas, is used since it provides adequate precision.

In general, large optical surfaces are ground and polished by machines that produce figures of revolution about a known point, usually located at the center of the surface. If the surface is one of revolution,[2] its departure from a straight line that is tangent to it at the center of revolution is a measure of its departure from its tangent plane and consequently determines the shape of the entire surface relative to any other chosen plane. The axis of abscissas (see fig. 2) is chosen as the intersection of a plane through the axis of rotation of the surface with the plane that is tangent to the surface at its center. The unit of abscissas is chosen as the separation of equally spaced reference marks along a chosen diameter of the surface. The chosen diameter is the axis of abscissas. The axis of ordinates is the axis of revolution of the surface. The unit of ordinates is one-half the wavelength of the light used.

The following steps are the chosen procedure for acquiring the data necessary to compute the shape of the surface. The master flat is placed on, and concentric with, the surface (positions A and A', fig. 2). The magnitude of the air wedge between the two surfaces is adjusted to produce a satisfactory number of fringes in the field (see fig. 3). The direction of this wedge is adjusted to make the fringes approximately normal to the line along which the chosen reference points lie. Except for plane surfaces this wedge varies from one reference point to the next. Consequently, if linear interpolation is used in estimating fractions of fringes at the reference points, the error in estimation varies with the fraction observed. For most observers this error is a minimum when the fraction is 0.0 or 0.5, that is, when the reference point falls on the center of a dark or a bright fringe. Best accuracy is obtained by reading abscissas corresponding to the center of all fringes—both dark and bright—and from these data, by nonlinear interpolation, compute the relative

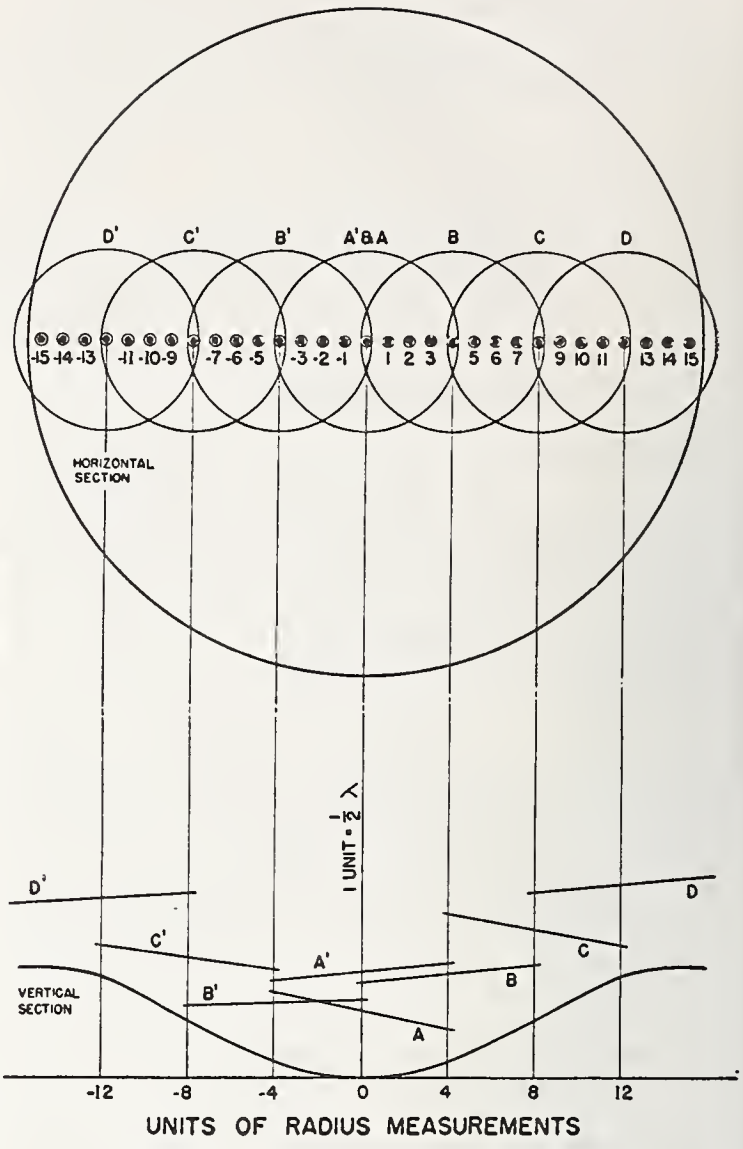


FIGURE 2. *Coordinate system of reference.*

The vertical section (below) cuts the horizontal section (above) along the line of reference points—dots concentric with small circles. Circles A, A' B, B', etc. represent the several positions of the standard test plate. The corresponding straight lines below represent sections through the standard surface of reference.

orders at the chosen reference points. The procedure will be explained in greater detail with the held of the fringe configuration shown in figure 3.

The fringes (relative orders of interference) in position A are evaluated at the several equally spaced reference points by the method indicated above. The standard is then moved to position B, which overlaps an appropriate amount of the area covered in its first position, A. Again the fringe readings at all reference points, covered by the standard in this new position, are evaluated. The standard is then moved to position C and the corresponding fringe readings noted. This procedure is repeated until the surface covered by the master extends to the edge of the unknown surface of revolution. The positions represented by primed letters are a second independent set, useful in checking the precision of the method.

In general, the fringe values at the reference points will not be integral. Consequently, the fractional parts must be obtained by interpolation and, in marginal cases, by extrapolation. Linear interpolation

[1] J. B. Scarborough, Numerical mathematical analysis, p. 446 (Johns Hopkins Press, Baltimore, Md., 1950).

[2] W. A. Granville, Elements of differential and integral calculus, p. 264 (Ginn & Co., New York, N. Y., 1911).

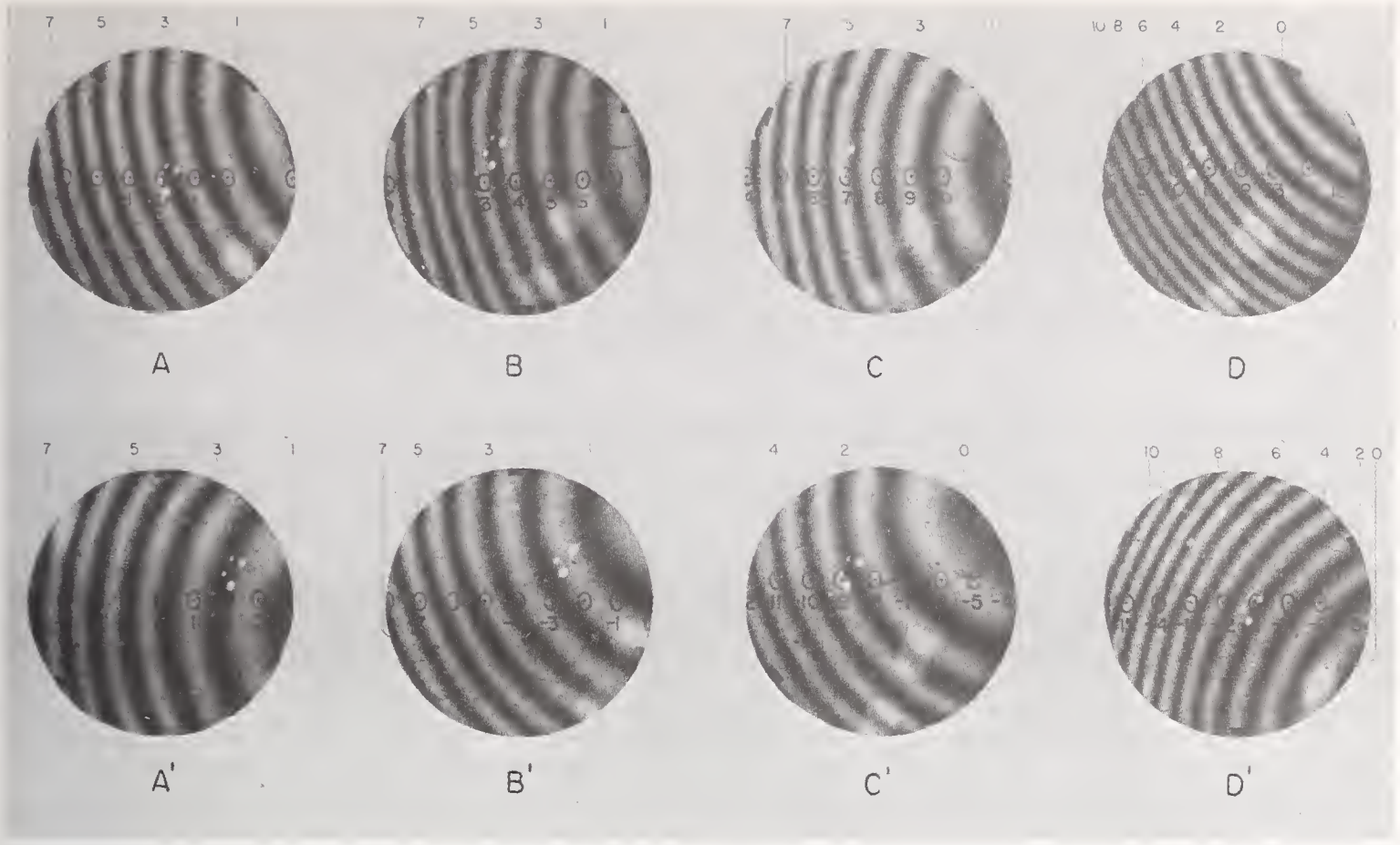


FIGURE 3. *Fringes for the several positions of the standard test plate.*

The numbers above each picture represent relative orders of interference.

or extrapolation is accurate only when the fringes are equally spaced along the straight line on which the reference points are chosen. If they are equally spaced, the surface is plane and no further test is necessary. In general, the fringes will be curved and nonlinear interpolation should be used if highest accuracy is to be obtained. Of the several methods of performing nonlinear interpolation, the graphical method is simplest and, for these purposes, adequate. The details of this are best shown by a description of its application to actual data. This will be given in a later section, where the shape of the surface of a glass disk will be computed from photographs of interference fringes that are obtained when the glass disk is tested interferometrically against a standard flat.

3. Derivation of Formulas

A set of formulas are now derived for use in the above-mentioned computation. In figure 4 the curved line represents the surface to be measured. The chosen reference plane is represented by the axis of abscissas. The chosen reference points are indicated by circles on the curved line. The several positions of the master plane relative to the coordinate system are represented by straight lines at various angles. The coordinates of all chosen reference points, which are marked on the surface of the unknown, are (R, Y_R). The observed fringe (a relative order of interference) at (R, Y_R) is designated A_R for the first position of the standard, B_R for the second position of the standard, C_R for the third position, and so forth. The slopes and intersections of the straight lines with the axis of ordinates are represented by M_A and A_0, respectively, for the A position, M_B and B_0 for the B position, M_C and C_0 for the C position, and so forth.

The slope-intercept form of the equation for the straight line representing the standard in the X position ($X=A$, B, C, etc.) is

$$Y_R+(X_R-X_0)=M_XR. \tag{1}$$

In position A, for example, the ordinates of the line are $(A_R+Y_R)=A_0+M_AR$, where A_R is the separation of the two optical surfaces. The absolute values for X_R are unknown, but differences in X_R, for any given position of the standard, are directly observable. Consequently, the integral part of the smallest X_R will be subtracted from all X_R's for purpose of computation. The quantity (X_R-X_0) is unaffected by this operation.

We will evaluate M_X and X_0 by applying the method of averages (footnote 1) to the data. The grouping of observations will be made in such manner that there is no overlapping of the two groups of points except when an odd number of reference points are to be applied to an evaluation. When the number of reference points in the two groups is odd, two observations are considered to be made at each of the points, and the two observations made at the central

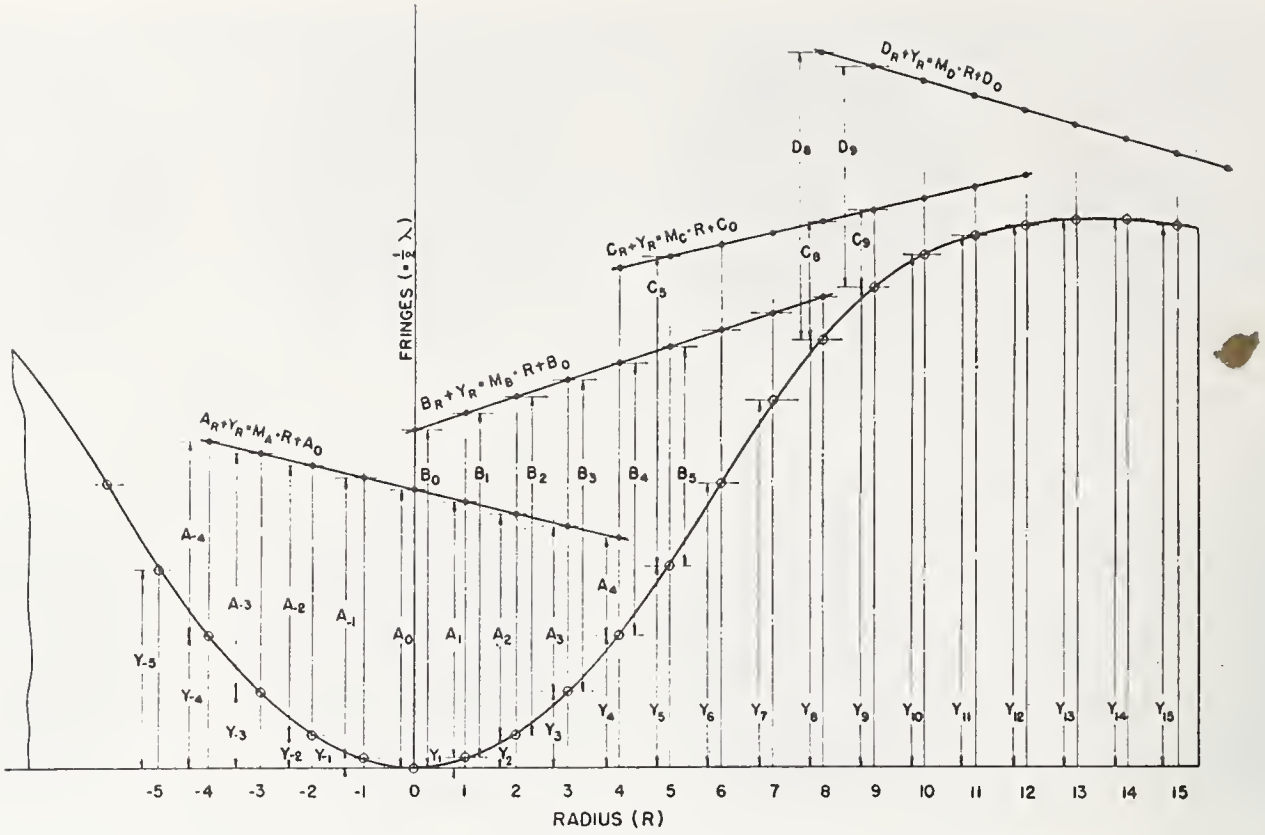


FIGURE 4. *Analytical representation of the surface and standard for mathematical analysis.*
The slanting straight lines represent the several positions of the standard and the adjacent formulas the equations of its corresponding locii.

point are divided between the two groups. The two groups then have an equal number of observations, and all observations are thereby assigned equal weights. When the total number of reference points is even, there is no overlapping of groups for an equal number of points to be allocated to each of the two groups for equal weighting of datum.

In applying the method of averages (footnote 1), we require that the algebraic sum of errors be zero for each of the two groups of observations. This requirement is represented by the following equations:

$$\left.\begin{aligned}\sum_{R_1}^{R_2}(Y_R+X_R)&=M_X\sum_{R_1}^{R_2}R+\frac{1}{2}N_X X_0,\\ \sum_{R_3}^{R_4}(Y_R+X_R)&=M_X\sum_{R_3}^{R_4}R+\frac{1}{2}N_X X_0,\end{aligned}\right\}\qquad(2)$$

where summations in the first of this pair of equations are from the lowest value of $R(=R_1)$ to the largest value of $R(=R_2)$ in this group and summations in the second equation are from the lowest value of $R(=R_3)$ in the second group to the largest value of $R(=R_4)$ in the second group. Equality of R_2 and R_3 will result when the number of reference points is odd, whereas they will differ by unity when the total number of reference points is even. The number of reference points used in each evaluation is N_X.

Solving for M_X and X_0 from the pair of equations (2), we get

$$M_X=\frac{\sum_{R_3}^{R_4}(Y_R+X_R)-\sum_{R_1}^{R_2}(Y_R+X_R)}{\sum_{R_3}^{R_4}R-\sum_{R_1}^{R_2}R}\qquad(3)$$

$$X_0=-\frac{2\left[\sum_{R_3}^{R_4}R\sum_{R_1}^{R_2}(Y_R+X_R)-\sum_{R_1}^{R_2}R\sum_{R_3}^{R_4}(Y_R+X_R)\right]}{N_X\left[\sum_{R_3}^{R_4}R-\sum_{R_1}^{R_2}R\right]}\qquad(4)$$

Equations (1), (3), and (4), together with the assumption that the surface to be measured has revolution symmetry, permit a computation of Y_R for all chosen reference points and consequently the shape of the surface.

The foregoing analysis permits a statistical evaluation of M_X and X_0 for all values of X (i. e., A, B, C, . . .) except the quantity A_0. An error in A_0 represents an error in the position of the reference plane and since this has no effect on the computed shape of the surface, the observed A_0 is assumed to be free from error.

As the surface is assumed to have revolution symmetry,

$$Y_R=Y_{-R}.\qquad(5)$$

From eq (1) $Y_R=M_AR+(A_0-A_R)$, and $Y_{-R}=-M_AR+(A_0-A_{-R})$. On eliminating M_A and Y_R from these three equations, we have

$$Y_R=A_0-\tfrac{1}{2}(A_R+A_{-R}).\qquad(6)$$

As M_A does not enter into the computation, it need not be evaluated.

4. Illustrative Example

Equations (1), (3), (4), and (6) form a set of fundamental equations from which all values for Y_R may be computed.

To clarify any possible lack of understanding in the development and application of these formulas, we will apply them to a set of observations and will compute the resultant shape of a surface. In this case the large circle in figure 2 represents the periphery of a 31.1-cm-diameter disk that is to be measured by means of a standard whose diameter is 8.2 cm. The seven circles, centered on one diameter of the large disk, represent the several successive positions of the standard. The eight circular fringe patterns shown in figure 3 (forming two independent sets, primed and unprimed) were obtained by photographing the fringe patterns formed by light reflected normally from the top surface of the large disk and the standard flat when in the several positions indicated in figure 2. The differences in values of the ordinates ($Y_R - Y_{-R}$), obtained from the two sets of photographs, is a measure of the accuracy obtainable. These differences are due to errors of observation and to an error in the choice of the center point. If the unknown is ground and polished on a spindle that is not concentric with it, the axis of rotation of the unknown surface will not be centered at the chosen origin of coordinates. However, the average value of Y_R and Y_{-R} from the two sets of data, primed and unprimed, respectively, will be almost free from the error of centering.

In order to obtain more precise values for X_R, the following procedure is followed: Reference marks with their associated R-values, or abscissas, are placed on the glass surface along the chosen reference line. These reference points are indicated by black dots centered in the small circles shown in figure 2. The standard is placed in the desired position and the air wedge adjusted so that a desirable number (5 to 10) of fringes cross the reference line. The present author prefers to have the fringes approximately perpendicular to the reference line. Considerable departure from this, however, is usually tolerated. The fringe pattern is photographed.[3] This is repeated for all desired positions of the standard. A fine straight line, using a needle point and a straight edge, is drawn across the photograph through the nine reference points on it. The positions of the centers of the references points are marked by pricking the photograph with a needle. This is to permit accurate readings on their positions. If no distortion is introduced in the photographic reproduction, the separation of all adjacent points will be equal. The intersections of all fringes, both dark (i. e., integral orders of interference) and light (integral plus half orders), with the reference line are marked also by pricking the photograph. The direction of increasing orders of interference is ascertained from tests and notations made when the fringes are photographed. The relative orders of interference, beginning with 1, are indicated by inserting numbers on the dark fringes. In figure 3 these numbers appear above each of the fringe patterns. The lesser visible numbers, adjacent to the reference points, represent the abscissas (or R-values) of the several reference points that were marked on the unknown surface.

In general, the scale of the photographic reproduction will differ from that of the chosen abscissa scale. For interpolation purposes any convenient scale may be used to measure the relative positions of points on the photograph. A centimeter scale with millimeter marks was used for this compilation. We will define the scale used as the P-scale and the readings from it as P-values. The P-values corresponding to all reference points and fringes are read off the centimeter scale and inserted in column 1 of table 1, which represents a typical data sheet. The corresponding R-values (numbers associated with the fiduciary reference points and appearing in the photographs) are inserted in column 2; the A_R-values, corresponding to integral and half-integral orders of interference, are inserted in their respective places in column 3. If an observed P-value, corresponding to an R-value, falls precisely on a fringe (reference point centered on a fringe), two identical P-values are inserted as for $R=-2$ and $A_R=5.5$ in the case given here.

The object of this listing and the following computations is to obtain more precise A_R-values (orders of interference) than can be obtained by direct interpolation or reading from the photographs.

A direct plot of the P-values versus A_R-values first entered in column 3, table 1 (the blank spaces to be filled later), shown as small black dots in figure 5, may be made. A smooth curve is then drawn through these points. The values for A_R in column 3, which correspond to R-values in column 2, are located on the curve. The corresponding ordinates represent the interpolated A_R-values desired. The values with asterisks, shown in table 1, are more precise values obtained by an interpolation process described in the following paragraphs.

The values for A_R, determined above from the direct plot, are not as precise as the directly observed values from which the plot was made. More precise

TABLE 1. *Arrangement of data for computation*

Values for A_R in column 3 that are followed by an asterisk are interpolated values from the curve of figure 5.

P	R	A_R	0.7 P	(0.7 P+A_R)
0.21	−4	8.23*	0.15	8.38*
.44		8	.31	8.31
.93		7.5	.65	8.15
1.39		7	.97	7.97
1.54	−3	6.80*	1.08	7.88*
1.85		6.5	1.29	7.79
2.40		6	1.68	7.68
2.94	−2	5.50*	2.06	7.56*
2.94		5.5	2.06	7.56
3.54		5	2.48	7.48
4.20		4.5	2.94	7.44
4.29	−1	4.42*	3.00	7.42*
4.80		4	3.36	7.36
5.43		3.5	3.80	7.30
5.64	0	3.33*	3.95	7.28*
6.05		3	4.24	7.24
6.85		2.5	4.80	7.30
7.00	1	2.41*	4.90	7.31*
7.70		2	5.39	7.39
8.37	2	1.62*	5.86	7.48*
8.60		1.5	6.02	7.52
9.67		1	6.77	7.77
9.73	3	0.98*	6.81	7.79*
10.80		.5	7.56	8.06
11.08	4	.39*	7.76	8.15*

[3] A micrometer eyepiece may be used to read the data directly and thus dispense with photography.

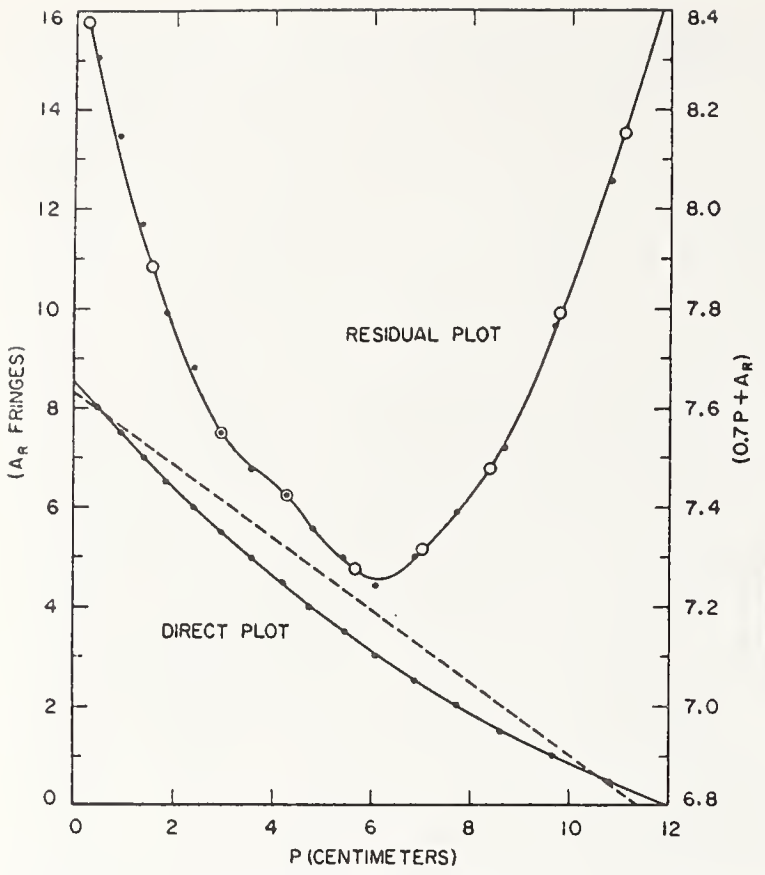


FIGURE 5. *Direct and residual plot of data for nonlinear interpolation of fringe orders.*

values may be obtained from a residual plot.[4] To obtain data for the residual plot, a column of values (table 1, column 4), computed with a slide rule, is added to the table of observed quantities. This column is a product of corresponding P-values and an appropriately chosen constant. This constant represents the slope of a straight line that crudely approximates the plotted points. It is indicated in figure 5. The approximate slope of this line is 0.7. A relatively large error in the choice of the slope contributes little or no error in the final results. Values of $0.7P$ are accordingly computed to two decimal places. The sums of corresponding values from columns 3 and 4, except where vacancies exist in column 3, are entered in column 5. These values are plotted, on an appropriately chosen scale, against corresponding P-values, and a smooth curve that best fits the points is drawn in by inspection (residual plot, fig. 5). One may now locate the points, shown as circles, on this curve, whose abscissas are those observed for the chosen reference points. The corresponding ordinates are read off and inserted in the $(0.7P+A_R)$ column. These values are marked with an asterisk for distinction. The corresponding values of $0.7P$ are subtracted from $(A_R+0.7P)$, and the resultant A_R values, shown also with asterisks, are inserted in the vacant places of the A_R column. These are the desired orders of interference at the several reference points. Values for $X_R(X=B,C,D, \ldots)$ are obtained from all photographs in this manner.

[4] H. M. Goodwin, Elements of the precision of measurements and graphical methods, p. 60 (McGraw-Hill Book Co., New York, N. Y., 1920).

TABLE 2. *Computation sheet for statistical evaluation of data*
The average of corresponding values in columns 3 and 5 represents the most acceptable results

R	A_R	Y_R	B_R	Y_{-R}
−4	8.23	------	------	------
−3	6.80	------	------	------
−2	5.50	------	------	------
−1	4.42	------	------	------
0	3.33	0.00	8.71	0.00
+1	2.41	−.09	7.13	−.09
2	1.62	−.23	5.75	−.27
3	0.98	−.56	4.49	−.56
4	.39	−.98	3.32	−1.04
5	------	−1.57	2.34	−1.58
6	------	−2.30	1.40	−2.28
7	------	−3.16	0.77	−3.13
8	------	−4.26	.29	−4.20
9	------	−5.67	------	−5.39
10	------	−7.05	------	−6.74
11	------	−8.64	------	−8.28
12	------	−10.30	------	−9.96
13	------	−12.37	------	−11.86
14	------	−14.51	------	−13.93
15	------	−16.86	------	−16.21

In order to determine the shape of the surface from the fringe readings and their corresponding abscissas, or R-values, these quantities are now applied to the formulas of eq (3), (4), and (6). Table 2 is a typical computation sheet. From eq (6) one computes $Y_4=3.33-\frac{1}{2}(8.23+0.39)=-0.98$, and likewise for all values of Y_R in the range covered by the standard in position A (column 3 of table 2). In order to proceed to the evaluation of Y_R for other values of R, eq (3) and (4) are used. Values for B_R, found in the same manner as described previously for A_R, are inserted in column 4, table 2. This permits a summation to be made of all terms shown in eq (3) and (4) and consequently an evaluation of M_B and B_0. As the number of reference points covered by the standard in each of the two positions, A and B, is odd (i. e., from $R=0$ through $R=4$, or 5 points), they are each assumed to represent 2 observations, making the total number of observations 10, which may now be divided into 2 equal groups. With this in view, it will be seen that $\sum_{R_1}^{R_2} R=2\times0+2\times1+1\times2=4$, and likewise for all summations shown in eq (3) and (4). The resultant values for M_B and B_0 are, respectively, −1.580 and +8.67. Slide-rule computation is adequate except for final precision. In general, the computed parameter B_0 will differ slightly from the observed value found in table 2. On substituting these values for M_X and X_0 in eq (1), and using values for R and B_R in columns 1 and 4, values for Y_R may be computed from $R=5$ through 8. These operations are now repeated, using consecutively the C and D data, to compute additional values for Y_R to the edge of the surface. These values are inserted in column 3, table 2. A similar treatment of the primed data yields an independent set of value for Y_{-R} shown in column 5, table 2, which should agree approximately with the above set. An average of Y_R and Y_{-R} values for each value of R is the accepted value. This averaging operation tends to decrease the effects of error in choice of the axis of rotation of the unknown surface as well as observational errors.

WASHINGTON, December 3, 1953.

Journal of Research of the National Bureau of Standards Vol. 61, No. 6, December 1958 Research Paper 2917

Parallel Testing Interferometer

James B. Saunders

The conventional methods of testing the parallelism of opaque bodies, such as gage blocks, by interferometry require wringing of the body to an optical flat. This operation disturbs the temperature equilibrium, necessitating long periods between tests, especially for long blocks. It often injures the surfaces of both the optical flat and the test body. Also, if the body is a standard gage block, repeated wringings during use ultimately change the dimension. This paper describes an interferometer for measuring the parallelism of gage blocks and other bodies of any reasonable length without the necessity of the wringing operation. Two forms of this instrument are used—one for testing long blocks and another for testing short blocks. Either form can be constructed for testing blocks of any length, but two forms are found to be more practical.

1. Introduction

The conventional procedures for measuring the parallelism of gage blocks [1, 2] require the wringing of the blocks onto an optical flat. The wringing operation often injures the contacted surfaces and repeated wringings necessitate frequent refinishing of the optical flat that is used as a base. A method for measuring parallelism, without the wringing operation, has significant advantages because the danger of injury to the contacted surfaces is eliminated. Accordingly, two instruments that utilize this method are described: one is for testing very long blocks and similar bodies, whereas the other is designed for short blocks. Both instruments use low orders of interference and neither requires the use of a standard.

2. Optics of the Interferometer

A description of the optics for either form of this instrument covers a large portion of that for the other. For distinction we will designate them as "the long-block interferometer" and "the short-block interferometer". The double-image prism used in these instruments is adjusted during construction [3] so that a ray of light, shown in the plane of figure 1A, after division into two component rays, 1 and 2, at P_0 will, on reflection at P_1 and P_2, lie in planes that are parallel to the semireflecting plane of the prism but deviate equally toward or from opposite sides of the plane of figure 1A. The projection of the light rays on the dividing plane is shown in figure 1B, which is perpendicular to the plane of figure 1A. This deviation is effected by rotating one component of the prism relative to the other about an axis normal to the dividing plane of the prism. If this deviation is held constant, the width of the interference fringes in the direction normal to the plane of figure 1A is fixed. This component of fringe width is, therefore, frozen into the system [4] when the cement between the component prisms becomes hard by cooling after adjustments are complete. Furthermore, the tilting of any plane surface outside the prism that affects the two component beams between division and recombination, will produce equal effects in this direction and, consequently, will not affect the fringe width. However, the rotation of plane surfaces about an axis normal to the plane of figure 1A will produce equal effects on the two component beams but in opposite directions, thus producing a proportionate effect on the fringe width in the direction parallel to this plane.

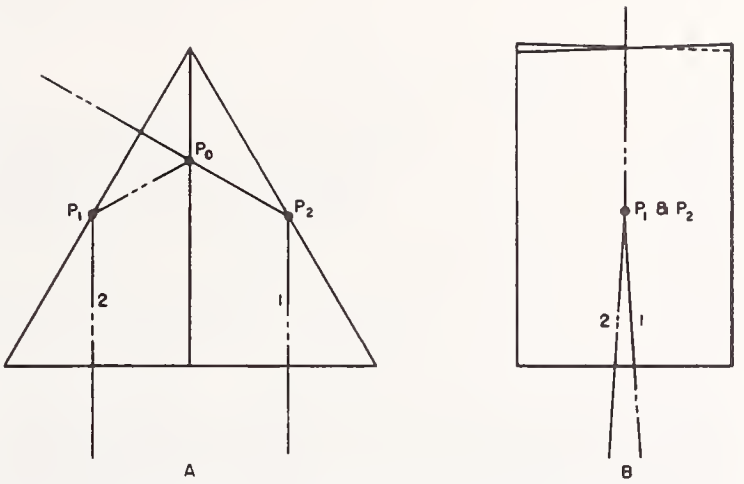


FIGURE 1. *Kösters double-image prism.*

A small angle is formed by the two 30° edges shown exaggerated in B.

Because of the above-described properties of this prism, adjustments of the instrument in which it is used affect the fringe width in one direction only. Consequently, when measuring the parallelism of gage blocks the test can be applied to parallelism in only one direction at a time. To test for parallelism in other directions the block must be rotated.

2.1. Long-Block Interferometer

Figure 2A is a horizontal section through the optical elements of this instrument. The light from a source at $S_1(S_2)$ is collimated by lens $L_1(L_2)$ and divided into two equal components by the beam-dividing plane $B_1(B_2)$. Each component suffers total internal reflection in the prism and emerges in planes

[1] Gauges and fine measurements, by F. H. Rolt, I, p. 204 (Macmillan and Co., Ltd., London 1929).
[2] The science of precision measurement, by Tbe DoAll Co., p. 143 (1953).
[3] Construction of a Kösters double-image prism, by J. B. Saunders, J. Research NBS 58, 21 (1957) RP2729.
[4] NBS Tech. News Bul. 42, 30 (1958).

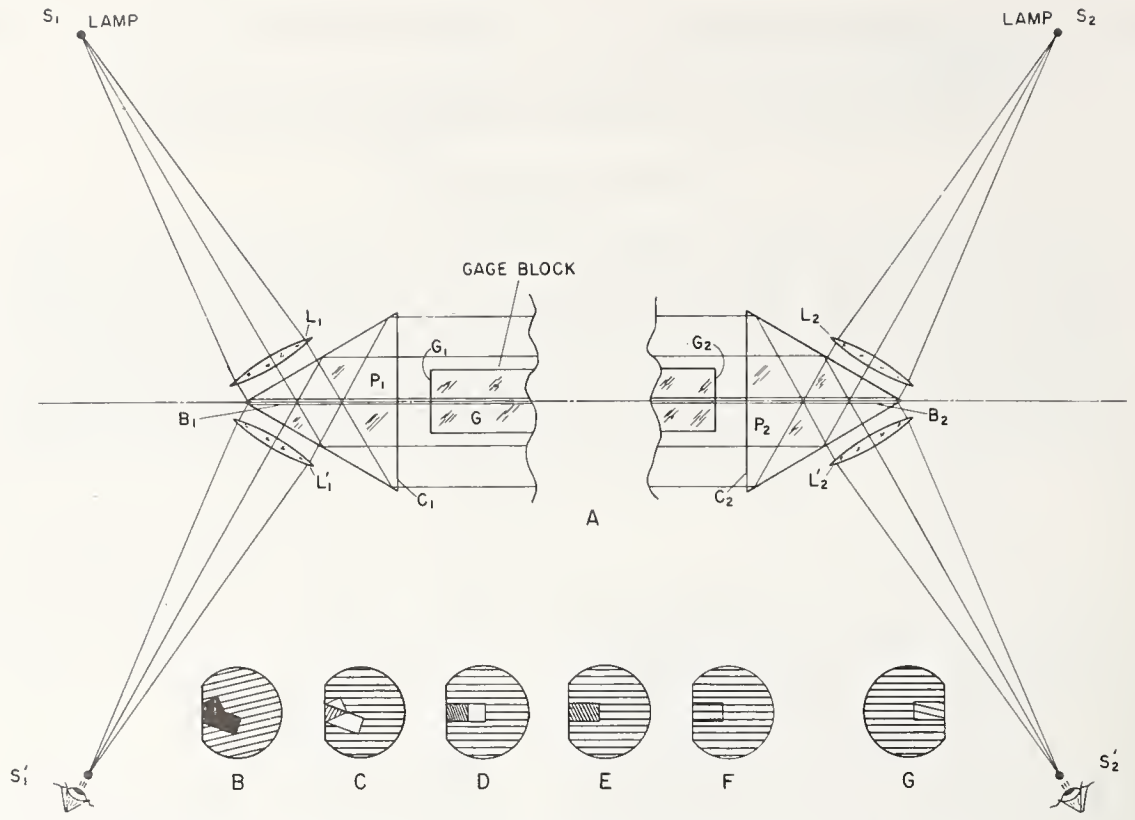


FIGURE 2. *Optics of interferometer for testing parallelism of long blocks.*

Figures B through G represent observed interference fringes.

parallel to $B_1(B_2)$, but at a small angle to the base surface $C_1(C_2)$. The latter condition is obtained by a slight rotation of $P_1(P_2)$ about an axis normal to the dividing plane $B_1(B_2)$. This serves to permit the elimination of light reflected from $C_1(C_2)$.

The two prisms, P_1 and P_2, are separated by a distance exceeding the length of the longest block to be tested. If desired, this distance may be made adjustable. The two dividing planes, B_1 and B_2, are adjusted to be coplanar and the two base faces, C_1 and C_2, are adjusted parallel to each other. A line joining the centers of P_1 and P_2 is adjusted to form a small angle with the normal to faces C_1 and C_2.

The light emerges from prism $P_1(P_2)$ as two separated components, one on each side of the dividing plane. It enters prism $P_2(P_1)$, again suffers total internal reflection, and each pair of component rays recombines in the plane of $B_2(B_1)$. One-half of each beam proceeds to the neighborhood of source $S_2(S_1)$ and the other half to $S_2'(S_1')$. An observer at $S_2'(S_1')$ sees a set of interference fringes that cover the entire aperture.

If a gage block, G, is inserted in the position shown with its end face, G_1, adjusted normal to the light beams, it will reflect equal and corresponding parts of the two component light beams from S_1 back through P_1 to S_1'. Accordingly, the observer at S_1' sees a background set of fringes, produced by light from S_2 and another set on the face of G_1, that is produced by light from S_1. Since G_1 is normal to the light beams these two sets of fringes will be parallel to each other and to the plane of figure 2A. If the other end, G_2, of the gage block is parallel to G_1, the fringes seen at S_2' will likewise be parallel to each other and to the plane of figure 2A. If, however, G_2 is not parallel to G_1, in the plane of figure 2A, it will not be normal to the light beams, and the sets of fringes seen at S_2' will not be parallel to each other. The angle between these two sets of fringes is a measure of the angle between G_1 and G_2 in the plane of figure 2A. The component of the angle between G_1 and G_2 that is perpendicular to the plane of figure 2A (or horizontally in figs. 2B to 2G) does not affect the fringes because it affects all pairs of component beams equally. If the component of the angle between the gage-block surfaces that is normal to the plane of figure 2A (or vertically in figs. 2B to 2G) is desired, the block must be rotated 90° and the operation repeated.

Since each prism is adjusted for complete compensation in the plane of figure 2A, white light can be used. A measure of the vertical width of the fringes (perpendicular to fig. 2A) for a known monochromatic light, with a micrometer eyepiece at S_2', gives a calibration of the micrometer scale in units (microns, millionths of an inch, etc.) of length for measuring the displacement of white-light fringes from a chosen reference point on the gage-block surface.

The procedure for adjusting a gage block is explained with the aid of inserts in figure 2. In general, when the block is placed on its supports, the light reflected from its end will not reach the observer because of excessive angular deviation from the

eyepiece. The block will appear in silhouette, as indicated in figure 2B. When the surface G_1 is adjusted approximately normal to the light, fine fringes will usually be visible in the area covered by both images of it, as shown in figure 2C. The images of the two parts of G_1 are made to coincide by rotating G about the center line of figure 2A, causing the image to change from that of figure 2C to figure 2D. A lateral motion, without rotation, will then change the image from that shown in figure 2D to figure 2E. A further small rotation of G about an axis normal to figure 2A brings G_1 normal to the light and the fringes on G_1 will appear horizontal and parallel to the background fringes as shown in figure 2F. The observer then moves to position S_2' and observes the set of fringes shown in figure 2G. The angle between these two sets of fringes corresponds to the angle between surfaces G_1 and G_2.

A photograph of the long-block interferometer is shown in figure 3. The base of the instrument is designed for rigidity so as to avoid flexure. Rigidity is quite important. The two prisms are mounted on rigid tables at each end of the base. Three screws (D_1 and two others not shown) permit raising and lowering of prism P_2. They also permit rotation or tilting of this prism about any chosen horizontal axis. The prism housings are fastened to the table tops by means of large-headed screws, in oversize holes, that permit lateral adjustments of the prisms relative to each other. Two screws, D_2 and another concealed by the housing of prism P_2, permit small rotations of this prism about a vertical axis by applying lateral torques to the legs of the table. Similarly, screws D_3 and D_4 permit rotary adjustments of the other prism housing.

The adjustments described above permit the alinement of the two prisms. This adjustment is critical, rather difficult to attain, but when once obtained is very stable. The final adjustments are executed while observing interference fringes produced by lens L'_2, in one of the many images of the source S_2. This author uses a pinhole source of approximately 1-mm diam and a short-focus lens to observe its image at S_2'. There are two sets of fringes, superimposed upon each other, in the proper image to be used. When these two sets of fringes are horizontal and very broad, fringes can be seen with the two eyepieces when focused on the pinhole. These fringes, in white light, are horizontal with the zero order in the center of the field.

The supports on which the long blocks rest are located at the Airy points so as to reduce changes in the angle between G_1 and G_2 due to gravitational distortion.[5] These supports rest on an adjustable plate which, in turn, is supported at one end by two points and at the other by one point. This plate is adjustable at one end, laterally with screw D_5 and vertically with another screw D_6 which is concealed in figure 3.

2.2. Short-Block Interferometer

The optics of the short-block interferometer are shown in figure 4. The double-image prism, lenses light source, and viewing position in figure 4B are identical to that of either end of the long-block instrument described above. A reflecting prism, P_R in figure 4, replaces one of the prism assemblies of the long-block instrument. Also, the optical axis of the instrument is vertical instead of horizontal. Figure 5 is a photograph of the short-block interferometer. Figure 4A may be considered a section through the center of 4B, coincident with the dividing plane of prism P. The indicated rays, 1 and 2 in figure 4A, however, do not lie in this plane. Their positions relative to it are indicated in 4E, which is a vertical view through 4B. The two surfaces, G_1 and G_2, of the gage block (figs. 4A and 4D) appear as G and G′ in figure 4E. G'_2 in figure 4A is an image of G_2 as seen by light reflected from the right-angle prism, P_R.

[5] F. H. Rolt, Gauges and fine measurements **II**, 340 (Macmillan and Co., 1929).

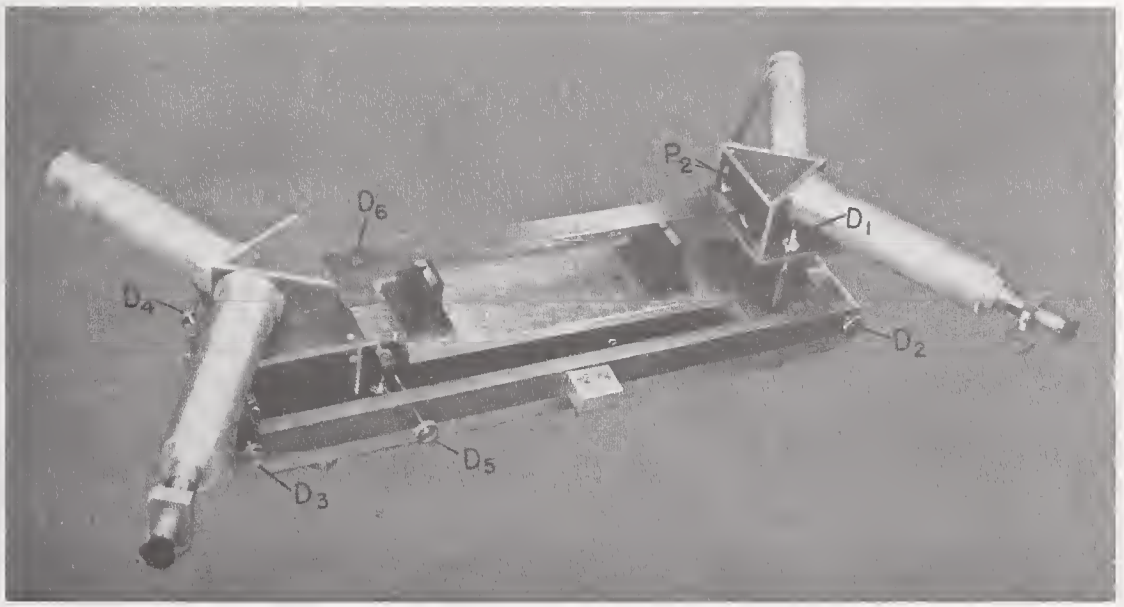


FIGURE 3. *Photograph of the long-block interferometer.*

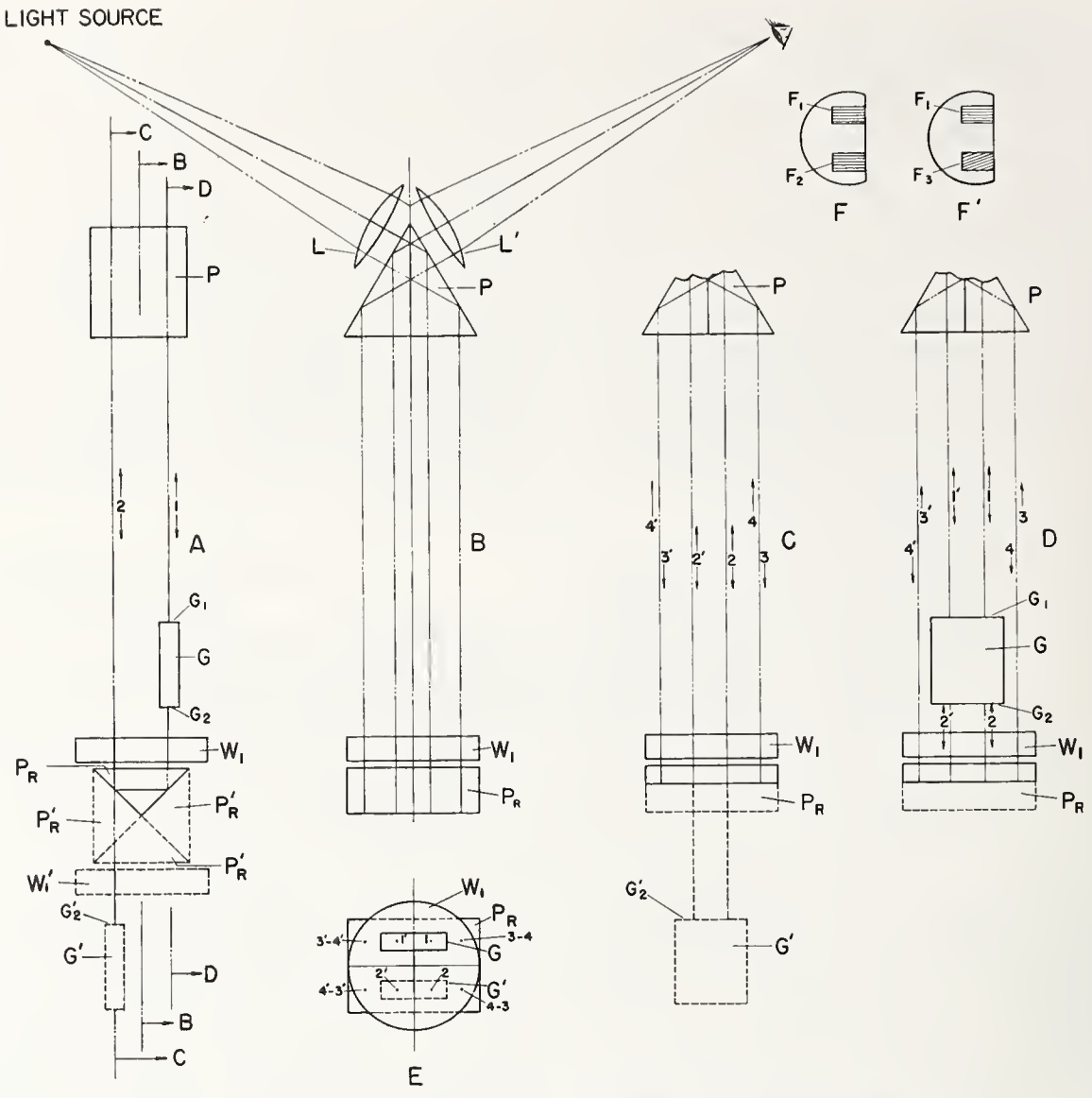


FIGURE 4. *Optics of the interferometer for testing parallelism of short blocks.*

Figures B, C, and D are sections through A. Figure E is a vertical view. Figures F and F′ represent interference fringes for a parallel and a nonparallel block, respectively.

A plane optical wedge, W_1, and its image, W'_1, (in figure 4A) are shown between G and its image G′. Figures 4B, 4C, and 4D are sections through 4A, along the lines indicated. The gage block is not located in the center as was the case in the instrument described above.

The two component rays of light, 1 and 1′ (figs. 4A and 4D), are caused to reflect normally from G_1 by adjusting G with a leveling screw. They return into P where they recombine to produce the interference fringes, F_1, shown in figure 4F. The two component rays, 2 and 2′ (figs. 4A and 4C) are transmitted downward through the optical wedge, W_1, suffer two internal reflections in P_R, and if P_R is properly adjusted, return upward and parallel to their directions of incidence through W_1 to G_2.

The wedge, W_1, is adjusted initially by rotation so that its thickness is constant at all points in either of the planes B, C, and D. When in this neutral position it does not affect the interference fringes because of compensation in each pair of component beams that pass through it. The function of W_1 will be explained later.

In order to measure the angle between G_1 and G_2, the deviation of the light by P_R toward or from the dividing plane of P must either be reduced to zero or its effect eliminated by measuring the observed angle for two orientations of G, which are 180° apart. The light that is not intercepted by G forms an interference pattern of uniform tint (or color) that fills the background about and between the two images of the gage block, shown in figure 4F or 4F′. Figures 4F and 4F′ represent the conditions observed when the ends of the block are parallel and nonparallel, respectively. A typical pair of component rays, which form this interference pattern, is indicated by 3 and 3′. They travel downward in figure 4C and upward in figure 4D. If the right-angle edge of P_R is normal to the dividing plane of P, the pair of rays 3 and 3′ in figure 4C can be made to return in planes that are parallel to the dividing plane, by rotating P_R about an axis parallel to the plane of 4A and normal to the incident light. This condition is attained when the background fringe is infinitely broad. The direction of the background fringes, when not infinitely broad,

FIGURE 5. *Photograph of the short-block interferometer.*

remain parallel to the dividing plane, because for each pair of component rays, such as 3 and 3′, there is a corresponding pair, 4 and 4′, that travel identical paths but in opposite directions. The optical path differences are, therefore, equal to each other and also equal to that for any other pair of component rays in the plane of 4C and 4D. The order of interference along the dividing plane corresponds to the optical path difference that was introduced into the double-image prism by the built-in wedges at the point that corresponds to the point of intersection of the right-angle edge of P_R and the dividing plane of P. This point is located in the center of figure 4E.

In general, due to the inherent error of judging when the background fringes are infinitely broad and to imperfections in the optical elements, a more precise method of evaluating the wedge between G_1 and G_2 is to measure the wedge for two positions that differ by 180°. If the background fringes are unaltered, the instrumental errors will be equal for the two positions and the value of the wedge unchanged except in sign. Consequently, the albebraic difference yields twice the value of the wedge.

There are three ways that one might evaluate the wedge between G_1 and G_2. The first is to rotate P_R until the order of interference at points C and E (fig. 6A) are equal; then rotate the gage block until the orders at A and B are equal; and finally, observe the difference in order of interference at points F and H. The second method is: After performing the above operations, instead of reading the order difference between F and H, reduce this order difference to zero by rotating W_1 and read the resultant change on a scale attached to W_1, figure 4E. This scale may be calibrated with monochromatic light and the units may be radians, degrees, or the corresponding variation in height of the block. A third method is to leave the wedge in its neutral position, adjust G so that the orders of interference of A and B are equal, and change the order at H to equal that at F by rotating P_R about an axis normal to the incident light and parallel to the plane of figure 4A. The order of interference between two points such as C and D (fig. 6A) after rotating P_R will be equal to one-half of that between F and H before this rotation was performed. By choosing a point, such as E in figure 6, such that CE equals K times CD, the order difference between C and E will be K times that between C and D.

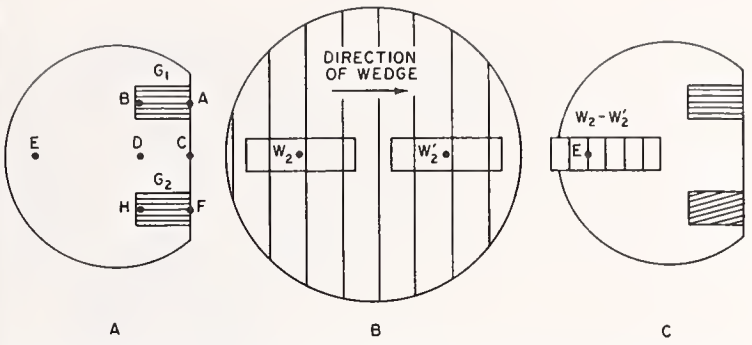


FIGURE 6. *Figures A and C represent the positions of reference points relative to the different sets of interference fringes.*

Figure B shows the wedge from which the two smaller wedges, W^2 and $W^{2\prime}$, were cut and the difference in optical thickness between them.

When using this last method for testing gages that are almost parallel, the angle between G_1 and G_2 will be small and the background fringes will be too broad for reading fractions of fringes. To eliminate this difficulty, an optical wedge, illustrated in figure 6B, is constructed and from it two sections W_2 and W_2' are cut and placed on P_R as shown in figure 7. The wedges W_2 and W_2' are equal, but when placed in the position shown, the effect is to narrow the background fringes seen through them. The results are illustrated in figure 6C. The difference in thickness of W_2 and W_2' at a selected reference point E (fig. 6C) is determined by the choice of the corresponding positions on the plate from which they were cut. This difference in thickness is chosen so as to cause the zero order of interference to pass through the chosen point when the background fringes about W_2–W_2' are infinitely broad.

If the angles of wedges W_2 and W_2' are properly chosen, the width of the fringes seen through them will be most favorable for measuring the fractional parts of fringes. Also, the position of the zero-order fringe, relative to point E, may be calibrated to read directly the angle between the ends of the gage blocks.

The recommended procedure for measuring a block is: (1) Adjust the two sets of fringes seen on the ends of the block so that they are perpendicular to the dividing plane, as in figure 4F; (2) note the position

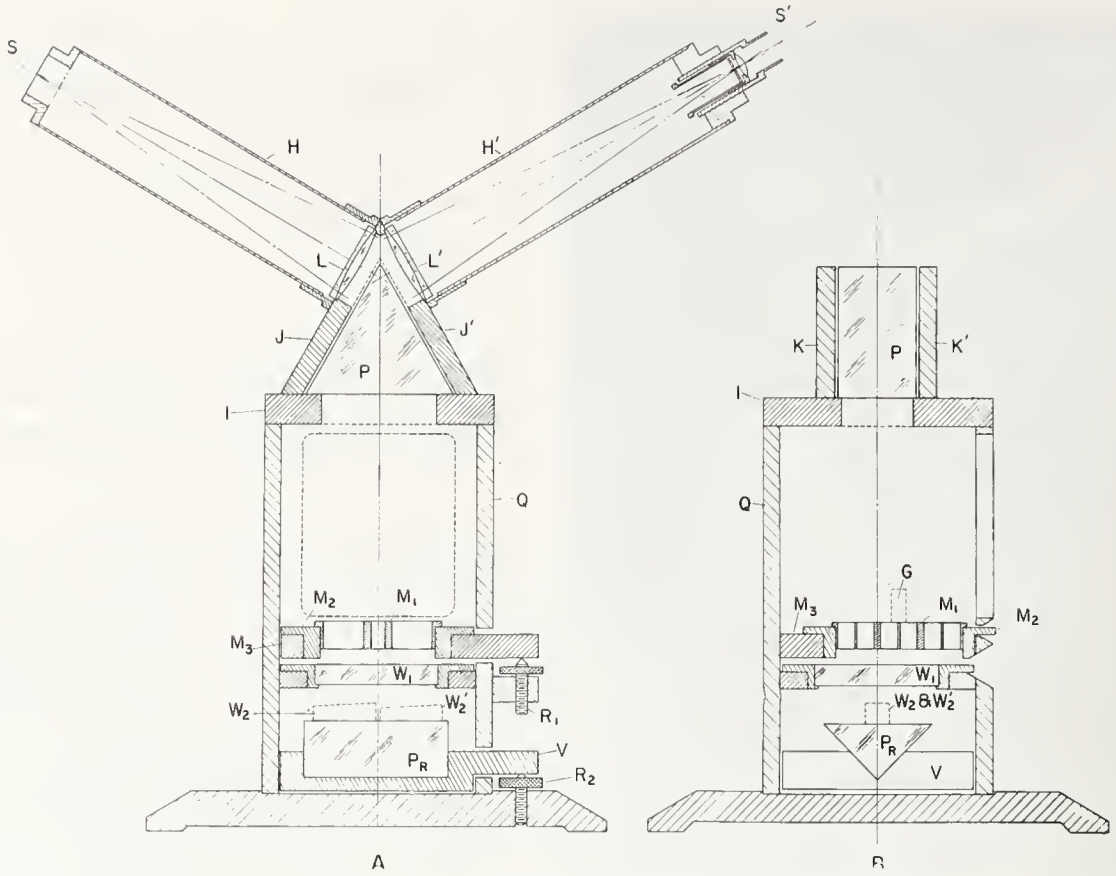


FIGURE 7. *Two vertical sections, at right angles to each other, through the short-block interferometer.*

of the zero-order fringe (or absolute order at E, fig. 6C); (3) rotate the block 180° about a vertical axis through its center; (4) readjust the fringes to restore the condition of (1) above; (5) again note the position of the zero-order fringe (or absolute order at E); (6) the difference in the two observed orders at E, or positions of the zero-order fringe, multiplied by the constant K, described above, is a measure of the angle between the ends of the block.

Figure 7 shows two vertical sections through the center of the short-block interferometer that are mutually perpendicular to and through the centers of each other. A pinhole, S, illuminated either with monochromatic or polychromatic light from outside the tube H, serves as source. The position of the pinhole is adjustable in the focal plane of the collimator lens, L, and the collimator tube is adjustable in length. The prism, P, rests on a thick plate, I, to which is fastened the lens-holding plates, J and J′. The plates, K and K′, which cover the ends of P, are not fastened to J or J′. Consequently, small stresses applied to tube H′ while adjusting the eye-piece or manipulating a micrometer in it, are not transmitted to P.

The material of the instrument, except for the collimator, the eye-piece tubes and optical elements, is made of steel. Steel was chosen because its expansivity approximates that of the glass elements more nearly than other usable materials.

The gage block, G, rests on a rotatable plate, M_1, that is perforated so as to transmit the required parts of the light beams used for making measurements. See also figure 8 for a vertical view of M_1 and its supporting parts. The plate, M_1, rotates in an annulus ring, M_2, which in turn is rotatable, from outside the instrument, in another annulus, M_3. The aperture in M_3 has its center displaced from the center of the instrument in a direction parallel to the dividing plane of P and by an amount equal to one-half the horizontal separation between the centers of the gage block, G_1, and its image, G′. The gage block rests on the center of M_1 (see fig. 6B). Accurate placement of the block is facilitated by stops.

The eccentric annulus, M_3, has an arm that projects through the wall, Q, of the instrument and is supported by this arm at one of its three supporting points by an adjusting screw, R_1. The other two supports for M_3 are steel balls, T_1 and T_2, (fig. 8B), which are held in conical holes by means of two screws, U_1 and U_2, respectively. The ends of these screws have eccentric conical depressions that permit a limited amount of rotation of M_3 about an axis normal to the dividing plane of P. The screw R_1 permits fine adjustment of M_2, and consequently the gage block, which it supports, about a horizontal axis parallel to the dividing plane of P. A similar pair of balls, screws, and the adjustable screw R_2, permits rotation of P_R about two axes parallel to those used for adjusting M_3.

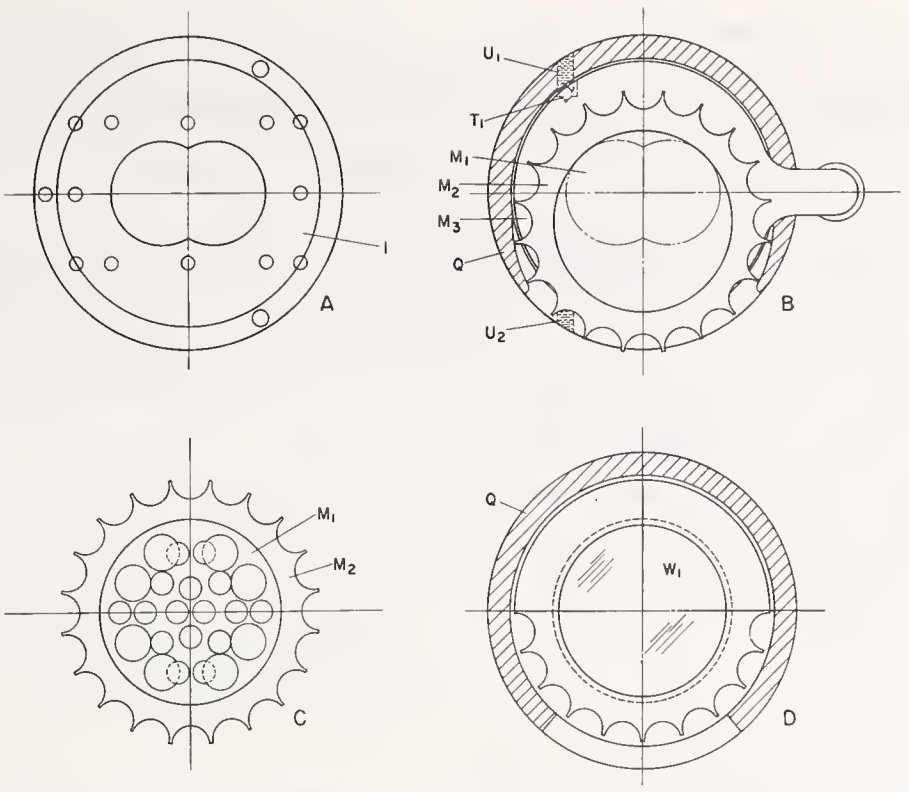


FIGURE 8. *Several views showing individual parts of the short-block interferometer.*

If the refracting edge of the optical wedge, W_1, is made perpendicular to the dividing plane of P (i. e., parallel to the plane of fig. 7B), each pair of component rays will traverse this plate at points of equal thickness. Consequently, W_1 in this orientation, does not affect the fringes of interference. It does, however, serve as a window, protecting the prism P_R from the accumulation of dust. Other functions of W_1 will be discussed later.

The prism, P, is centered over the aperture in plate I (figs. 7A and 8A) with its ends parallel to plates K and K′ (fig. 7B). Using the base surface of P as a plane mirror, the pinhole aperture S is located in the focal plane of lens L by varying the length of tube H and at a point in this plane where the light will form an image of the pinhole upon itself. This places the light beams, after division at the dividing plane, normal to the prism base and, consequently, parallel to each other.

A gage block, whose end faces are parallel to each other, is placed on the center of plate M_1 with its lower surface parallel to the top surface of M_1. When the surface of the block is adjusted parallel to the base of P by means of screws R_1, U_1, and U_2, the light from S is reflected normally from the top surface of the block and observed at the eyepiece as interference fringes.

The parts of the light beams that are not intercepted by the block and its support M_1, traverse the optical wedge W_1, and enter the right-angle prism P_R. After two internal reflections in P_R, the light returns along a path that is symmetrical to its incident path with respect to the 90° edge of P_R. The 90° edge of P_R is made normal to the dividing plane of P by means of screws in its support similar to U_1 and U_2 and the fine-adjusting screw R_2 (fig. 7A). This light forms the background fringes used in the test, but no interference is observed until this prism edge is nearly normal to the dividing plane. When it is not normal the two images of this edge, formed by the two component light beams, intersect in the extension of the beam dividing plane. The prism P_R is rotated until the two images coincide—perfection being attained when the background fringes have maximum contrast.

The above-mentioned contrast in the background fringes is not affected by screw, R_2, since it has no vertical rotational effect on the prism. Its effect is to change the width of the background fringes only. When the background fringes are made infinitely broad, the collimated beam of light returns toward P parallel to the incident beam—all rays having suffered a horizontal shift in P_R, as illustrated in figure 4A.

The holes in the gage-block support, M_1 (fig. 8C), are so spaced that when M_1 is rotated to one of the four positions for which the rectangular sides of the block are either parallel or perpendicular to the dividing plane, all light that goes through M_1 returns again through it. That is, the apertures in M_1 are symmetrical both with respect to the dividing plane and to the 90° edge of P_R. Parts of the beam (rays 3, 3′, 4, and 4′ in fig. 4) will pass downward through M_1 to P_R, shift horizontally in P_R, and pass upward through other apertures in M_1 to P. Other parts of the beam (rays 2 and 2′ in fig. 4) will pass downward through M_1, shift horizontally in P_R, pass

upward through M_1 to the lower surface of the gage block, return through M_1 to P_R and again upward through M_1 to P. If the two end faces, G_1 and G_2, are parallel, the light will be incident on G_1 and G_2 at equal angles. When G_1 is adjusted normal to the light, G_2 will also be normal to it.

We have three sets of fringes, shown in figure 4F′, to consider: (1) The set F_1 is formed by light reflected from the top surface, G_1, of the gage block; (2) the background fringe between and about F_1 and F_3, formed by light reflected from P_R but not incident on the gage block; and (3) the set F_3 formed by light reflected from P_R to G_2 and back through P_R. The direction or orientation of F_1 determines the angle between G_1 and the incident wavefront; the width of the background fringes determine the direction between the incident and reflected beams to and from P_R; and the orientation of F_3 determines the angle between G_2 and the wavefront that is reflected from it.

The set of fringes, F_1, is adjusted by means of R_1 normal to the dividing plane of P, for which condition G_1 is normal to the incident light. The background fringe is made infinitely broad, for which condition the light beams returning from P_R are parallel to the incident beams. If the two surfaces of G are parallel, the set of fringes appearing on G_2 will be parallel to those on G_1 and indicated as the set F_2 in figure 4F. If G_1 and G_2 are not parallel, the fringes seen on them and indicated as F_1 and F_3 in figure 4F′ will not be parallel to each other. The angle between these two sets of fringes is a measure of the angle between the two ends of the gage block.

Washington, May 5, 1958.

Journal of Research of the National Bureau of Standards Vol. 62, No. 4, April 1959 Research Paper 2943

Interferometer for Large Surfaces

James B. Saunders and Franz L. Gross[1]

An interferometer is described that permits the testing of large areas, such as layout plates. The extension to large areas is obtained by causing a collimated beam of light to reflect from the specimen at a large angle of incidence. The resultant fringe pattern is a contour map of the surface relative to an arbitrarily chosen plane and the contour interval is a function of wavelength and angle of incidence.

1. Introduction

An interferometer for testing large surfaces has been described by Linnik.[2] However, in spite of certain advantages, the Linnik interferometer has not been used extensively in this country. This lack of application may be due to the complexity of the instrument.

The instrument described here is relatively simple and easy to operate, is twice as sensitive as the Linnik interferometer, and is relatively free from vibrations.

2. Optics of the Interferometer

The optics of this interferometer are shown schematically in figure 1. The light from a source at S is collimated by the lens L_1, and separated into two coherent beams by the semireflecting dividing plane of a Kösters double image prism.[3,4] The entrance angle of the collimated light may be adjusted to give any desired deviation θ between the component beams 1 and 2. Beam 1 is reflected normally from mirror M_1 and returns upon itself. Beam 2 is reflected from the surface to be tested at an angle of incidence of $90-\beta$ degrees, then normally from mirror M_2 and returns along its previous path to the dividing plane of the prism, where it recombines with beam 1. The observer at E sees interference fringes on the superimposed images of the two mirrors.

The gross aspects of the fringe pattern (fringe direction and spacing) are controlled by the wedge angle between the wave fronts of beams 1 and 2, while the small irregularities of the fringe pattern are a function of the irregularities of the test surface. If the test surface were perfectly flat, the fringes would be straight and parallel. Any curvature of the test surface introduces a corresponding curvature in the wave front of beam 2, and this wave front, when compared with the plane wave front of beam 1, introduces curvature into the otherwise straight fringes. By adjusting the tilt of the test surface (i.e., setting $\beta=\theta$ and introducing a slight tilt across the width of the test area), it is possible to adjust the fringes so that they run parallel to the long dimension of the test area (as seen by the observer at E). In this case, the curvature of the fringes is a direct and precise measure of the curvature of the test area (in its long dimension). The fringe pattern is a contour map of the area of the specimen surface that is being tested.

3. Sensitivity

The sensitivity of the instrument depends upon the value of β. Since the light is reflected twice from the specimen surface, the sensitivity is double that obtained with the Linnik interferometer for the same angle of incidence. When the instrument is adjusted as described in the preceding section, one fringe departure from straightness corresponds to a departure from flatness of $\lambda/(4 \sin \beta)$, where λ is the wavelength of light.

[1] Present address: University College, London, England.
[2] V. P. Linnik, Compt. rend. acad. sci. U.R.S.S. **35**, 16 (1942).
[3] J. B. Saunders, Construction of a Kösters double-image prism, J. Research NBS, **58**, 21 (1957) RP2729.
[4] The Kösters double image prism is made from two 30°–60°–90° prisms, one of which is partially silvered on the face opposite the 60° angle. The two prisms are then cemented together to form the equilateral prism shown in figure 1. The partially silvered surface becomes a semireflecting plane.

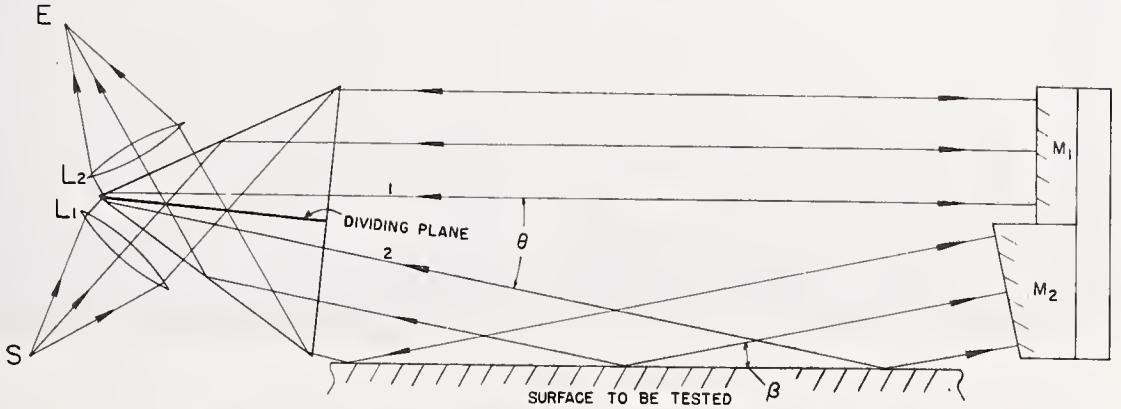


FIGURE 1. *Optical arrangement of the surface plate interferometer.*

In a finished instrument, the angles θ and β could be made adjustable so that any length surface could be made to fill the aperture of the system. This would give maximum sensitivity for any size surface measured. However, since the value of β must be known for evaluating the fringe pattern, it may be more practical to use fixed values for θ and β (with M_1 and M_2 bound into a rigid unit) and bind the mirror unit to the prism housing. This would eliminate the necessity of making frequent measurements of β and also add stability to the instrument. The resulting instrument would have a fixed adjustment and could be used by an unskilled operator.

The maximum length of surface that may be covered with one setting is $A \csc \beta$, where A is the aperture of the prism. Thus by decreasing β, any length surface could be covered with a prism of a given aperture, but of course the sensitivity of the instrument would also decrease.

4. Experimental Model

Figure 2 is a photograph of the pilot model that was used to test the interferometer. The surface plate used in this assembly was 91 cm long. The wedge angle between the wave fronts was controlled by adjusting the end mirrors, which produced the same effect as tilting the surface plate. The Kösters double image prism and lens assembly shown in figure 2 were taken from another instrument that was designed for use with a clear circular aperture of

FIGURE 2. *Pilot madel of surface plate interferameter for testing feasibility of the optical principals with a typical granite surface plate.*

The reflectivity of the end mirrors is designed to give approximately equal intensity for the two lightbeams at the receiving point E.

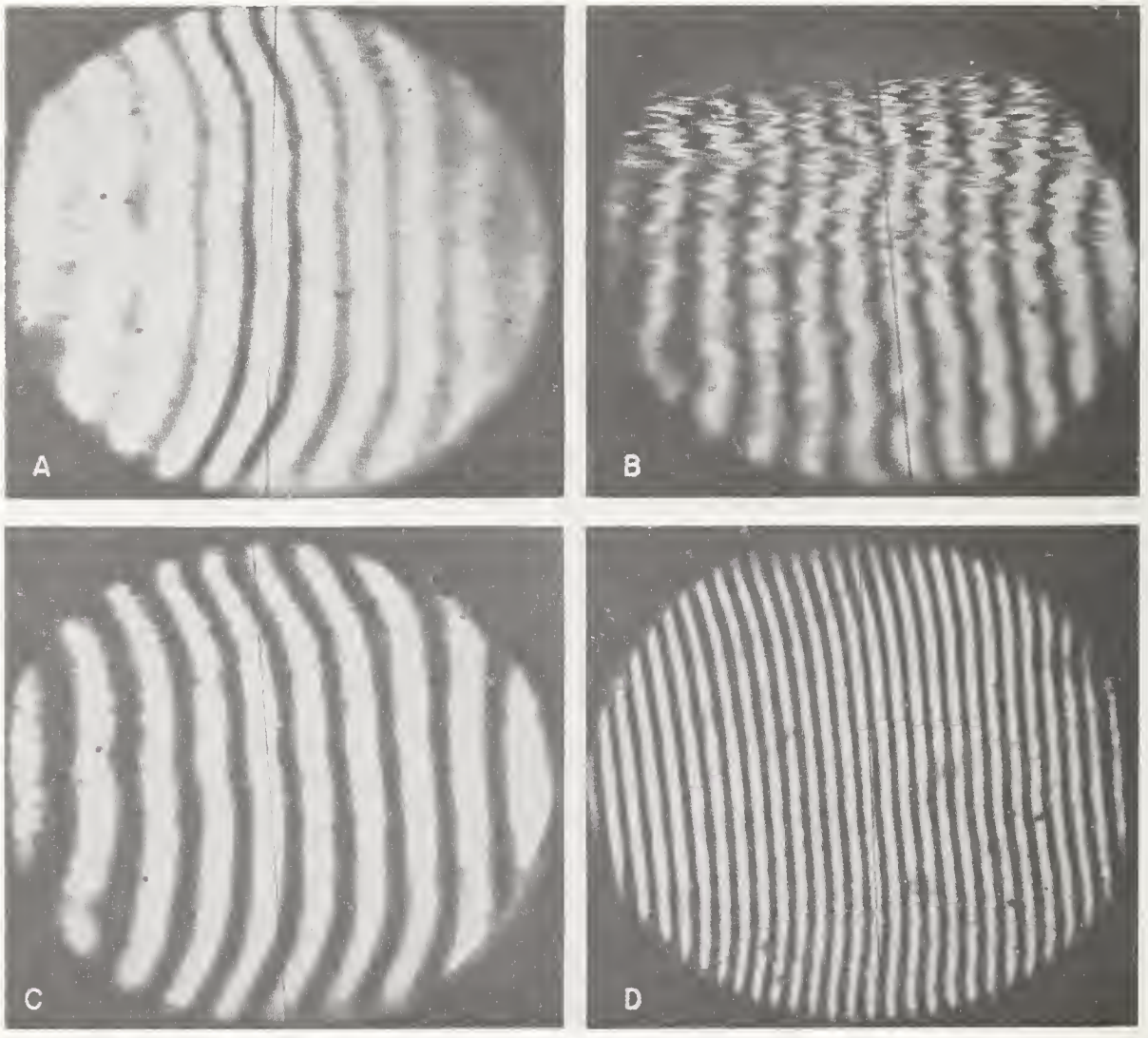


FIGURE 3. *Fringe patterns.*

A, White light from granite surface plate; B, monochromatic light from a badly scratched cast iron surface; C and D, monochromatic light from a granite surface. Angle of incident on surface plate is approximately 2½°.

51-mm diam (2 in.).[5] Since the prism had an aperture 51-mm square, better coverage of the specimen surface could have been obtained by increasing the circular aperture to $51\sqrt{2}$ mm. This would have left the square aperture of the prism unobstructed, so that the area of the specimen being tested would have been a rectangle 51 mm by 91 cm, and the unfavorable narrowing of the field at the ends of the specimen due to the circular aperture would have been eliminated.

5. Results

The interference fringes shown in figure 3 were taken with the instrument described above and shown in figure 2. Photograph A was made with white light and the others with monochromatic (λ=5876 A) light. The difference between photographs C and D was obtained by changing the angle between the two interfering wavefronts, thus changing the width of the fringes. Photographs A, C, and D of figure 3 were made with the plate of black granite, shown in figure 2, whereas photograph B was made with an old cast iron plate (dated 1918) that was badly scratched and marred. In this model the angle β was 2½° and thus a departure from straightness of one fringe corresponds to a deviation from flatness of 5.75 λ (approximately 0.00013 in.).

The optical performance of this instrument was found to be highly satisfactory. Some mechanical development is necessary in order to produce a more practical instrument. Accordingly, a few suggestions given here may be of help to the designer. The prism should be designed to use its entire rectangular aperture. The apertures of the lenses should be sufficiently large to prevent constriction of this aperture. The prism and end mirrors should be rigidly bound together if vibrations of the fringes are to be avoided. White light may be used but the resulting difficulty of finding and adjusting the fringes exceeds the inconveniences associated with monochromatic sources.

[5] This prism assembly was designed for use with θ=0°. Using it at θ=2½°, the circular aperture became an oval aperture with a vertical diameter of 45 mm.

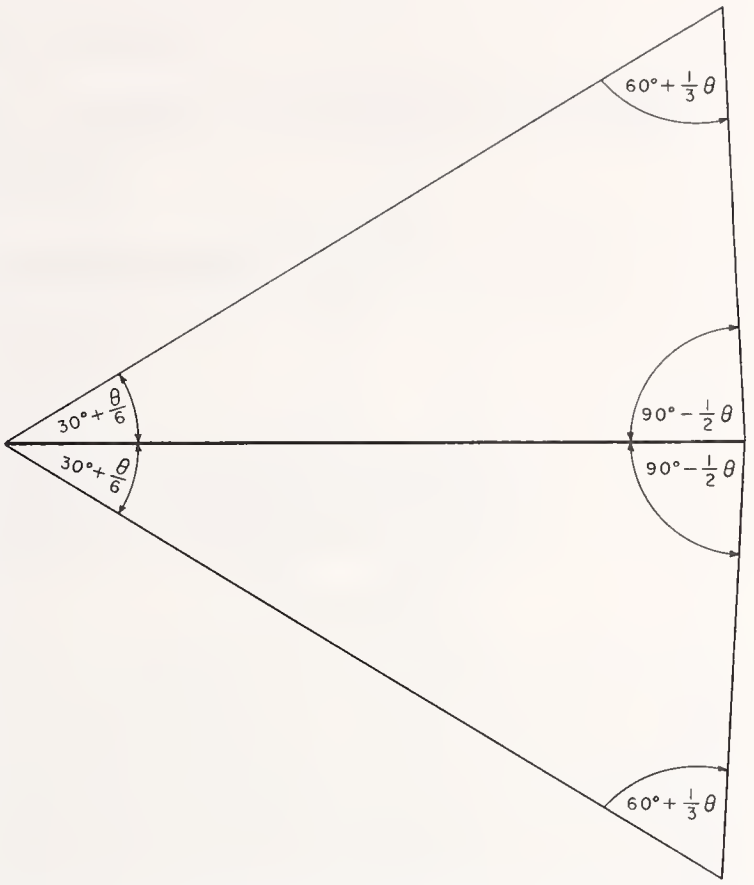


FIGURE 4. *Modified Kösters double-image prism for use with monochromatic light.*

A cover that encloses the light beams, between the prism and end mirrors, greatly enhances the stability of the fringes, and should be used unless the surrounding air is in a steady state. If white light is to be used, differential refraction may be reduced to a minimum by using the form of Kösters prism shown in figure 4. The procedure for making such a prism differs from that described in footnote 4 in that all surfaces are finished before cutting the prism into two parts.

WASHINGTON, December 8, 1959.

518-743 O - 73 - 11

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 65B, No. 4, October–December 1961

Measurement of Wave Fronts Without a Reference Standard:

Part 1. The Wave-Front-Shearing Interferometer

James B. Saunders

(June 26, 1961)

The wave-front-shearing interferometer may be used to test any converging wave front regardless of whether or not it is symmetrical. A mathematical operation is described that permits complete analysis of the data. This operation yields values of the deviations of wave fronts under test from a close fitting sphere. The reference surface may be chosen statistically so that the results are the deviations from a best fitting sphere.

Introduction

The wave-front-shearing interferometer [1, 2, 3] [1] provides a comparison of a converging wave front with a sheared image of itself. Similarly, the wave-front-reversing interferometer [4] provides a comparison of either a converging or plane wave front with a reversed image of itself. A detailed description of a method of analyzing the data from these interferometers, except for the special case of revolution symmetry, has not been published. Consequently, the absolute shape of a completely unknown wave front could not be obtained.

The method of analysis involves a mathematical operation. This operation is similar for the two aforementioned interferometers. Its description is more clearly portrayed in its application to the wave-front-shearing interferometer (≡WSI) than to the wave-front-reversing interferometer (≡WRI), although the method evolved from a study of the latter. There are, however, sufficient differences in the two operations to warrant separate treatments. Accordingly, part 1, which follows, will describe the analysis of data from a WSI and part 2, to be published later, will describe the analysis of data from a WRI. Part 3 of this series will describe the absolute testing of optical flats. This method is not new but has not been described adequately in the literature.

Part 1

1. Analysis of WSI Fringes

The WSI will yield unique solutions of a wave front. The assumption of symmetry (i.e., that it have revolution symmetry) is unnecessary. Any convergent wave front may be tested along any chosen diameter or along the line of its intersection with any plane that passes through the point of convergence. Several equally spaced reference points are chosen along this line. The line of reference points must be parallel to the direction of shear and the shear must be an integral multiple of the separation of the points. This arrangement and spacing of the reference points causes their images, in the region of overlapping wave fronts, to coincide in pairs as is shown in figure 1A.

The separation of the reference points will be defined as unity for this discussion. If the magnitude of the shear is unity also, one adjustment of the system (one set of fringes) is sufficient to evaluate the deviations of the wave front at the chosen reference points from a statistically chosen sphere. For a shear that exceeds unity, more than one fringe pattern (i.e., two or more shear values) must be used to obtain a solution.

In figure 1, $P_\nu(\nu=0, 1, 2 \ldots N)$ represents the $(N+1)$ reference points in one of the images, W, of the sheared wave front and P'_ν the corresponding points in the other image, W'. Let δ_ν (see fig. 1B) equal to the deviation of the wave front at P_ν, from the corresponding point, T_ν, on a reference circle, C, (to be chosen later), ϵ equals the angle between the two images, C and C', of the reference circle at their point of intersection and μ equals the distance from the intersection point to P_0. The distance, μ, is positive if P_0 is below (in fig. 1B) the intersection of the two circles and negative if above. The deviation, δ_ν, is positive if P_ν is on the concave side of the circle, and negative if it lies on the convex side of the reference circle. The distance from P_0 to P_ν (or from T_0 to T_ν), measured along the circle, is always positive and is represented by ν.

The separation, S_ν (distance from T_ν to $T'_{\nu-1}$), of the two images of the reference circle at any pair of points, P_ν and $P'_{\nu-1}$, (assuming unit shear) may be obtained from figure 2. The centers of the two circles are located at C_1 and C'_1. The distance $C_1C'_1$ equals $2h$, R equals the radii of the two circles, ρ equals the distance from E to T_ν, E bisects the line $C_1C'_1$ and ϕ is the angle subtended by $(\nu+\mu)$ at E. Applying the law of cosines (from trigonometry) to triangles $EC'_1T'_{\nu-1}$ and EC_1T_ν we obtain

$$R^2=h^2+(\rho+S_\nu)^2-2h(\rho+S_\nu)\cos(90-\phi) \text{ and}$$

$$R^2=h^2+\rho^2-2h\rho\cos(90+\phi).$$

[1] Figures in brackets indicate the literature references at the end of this paper.

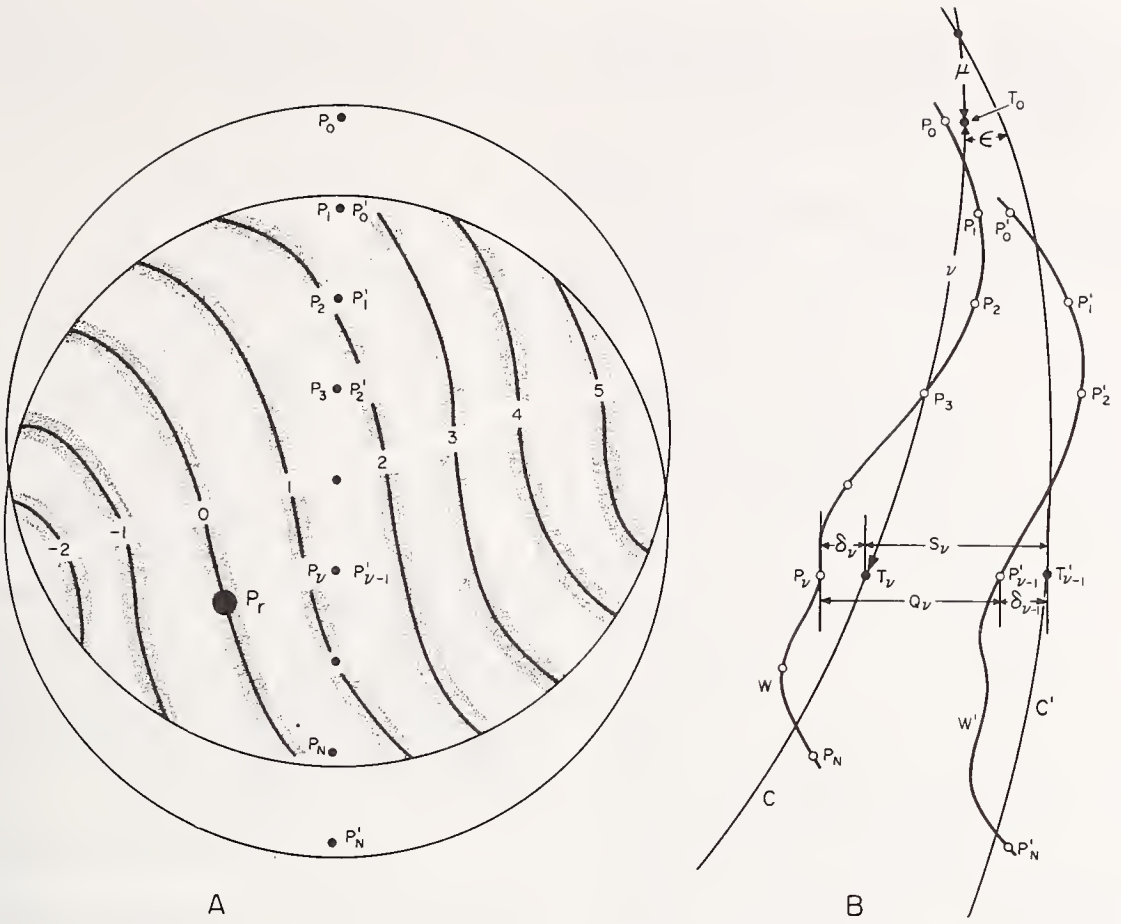


FIGURE 1A *Two images of a wave front sheared laterally relative to itself.*
B *Illustration of the two wave fronts relative to the images of a reference circle.*

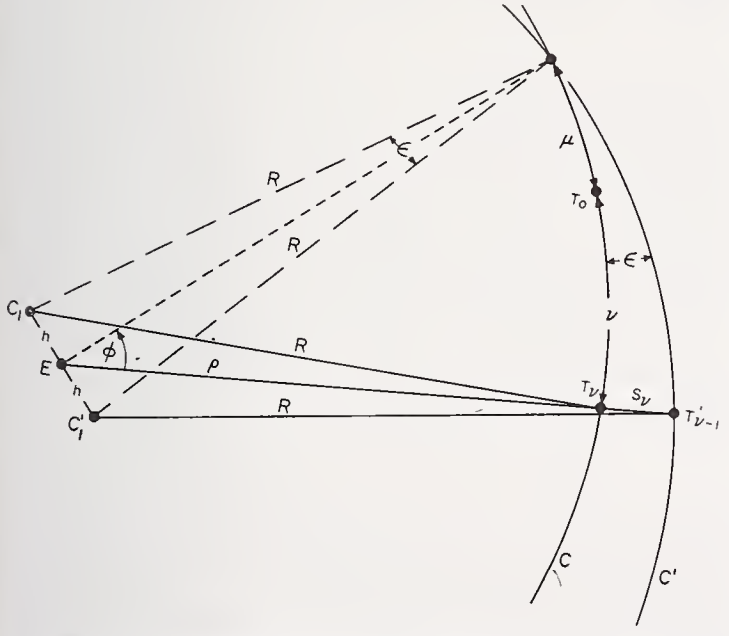


FIGURE 2 *Relationship between parameters and other variables.*

On taking differences and solving for S_ν we obtain,

$$S_\nu = 2h \sin \phi. \tag{1}$$

This equation may be applied rigorously, but if the focal ratio (radius of curvature of the circle divided by the diameter of the wave front) exceeds six, no significant error is introduced by replacing sin ϕ by its approximate equivalent, $(\nu+\mu)/R$. Also, since the angle ϵ will always be quite small, the value of $2h$ is approximately equal to $R \cdot \epsilon$ and the value of S_ν in eq (1) becomes

$$S_\nu = (\nu+\mu)\epsilon. \tag{2}$$

If we let Q_ν represent the separation of the two wave fronts, W and W', at P_ν, it is seen from figure 1B that

$$\left.\begin{aligned} &\delta_\nu = \delta_{\nu-1} + Q_\nu - S_\nu = \delta_{\nu-1} + Q_\nu - \epsilon(\nu+\mu) \\ &\nu = 1,\ 2,\ 3\ \ldots\ N. \end{aligned}\right\} \tag{3}$$

All quantities that represent optical distances are given in units of the wavelength that is used. Accordingly, these quantities should be multiplied by the wavelength to convert to standard units of length. The quantity Q_ν may be observed directly if white light is used to adjust the zero order to a known point, P_r, and then, using monochromatic light of known wavelength, observe the order difference between points P_r and P_ν. It is inconvenient and unnecessary to use white light. Let Q_r represent the unknown order of interference at an arbitrarily chosen point, P_r, in the fringe pattern (fig. 1A) and q_ν the difference in order (number of fringes) between this point and P_ν. The order of interference, Q_ν, at

P_ν is, therefore, equal to $(Q_\tau+q_\nu)$. This introduces another unknown into the equations of observation; but this may be absorbed by replacing the product $\epsilon\mu$ with a new parameter, r, such that

$$\epsilon\mu=(Q_\tau-r). \tag{4}$$

With these values for Q_ν and $\epsilon\mu$ replaced by their respective equals, eq (3) becomes

$$\left.\begin{aligned}&\delta_\nu=\delta_{\nu-1}+q_\nu+r-\nu\epsilon\\&\nu=1, 2\ \ldots\ N\end{aligned}\right\} \tag{5}$$

Equation (5) represents N equations that contain $(N+3)$ unknowns: namely, $(N+1)$ δ_ν's plus the two parameters ϵ and r. We need three additional equations relating these $(N+3)$ unknowns if a solution is to be possible. It appeared that one could use a second set of fringes formed with a shear of two units, as shown in figure 3A. This would yield $(N-1)$ additional equations of observations with only two new unknowns, i.e., two more parameters, ϵ_2 and r_2, to be added to the above-mentioned $(N+3)$ unknowns. If N is 6 or larger, we would have as many equations as unknowns and a solution would seem to be possible, but it can be shown that these equations are not entirely independent.

The δ_ν's do not have significance until the reference circle, C, shown in figure 1B, is defined. Any circle may be defined by three conditions. In analytical geometry, these might be the coordinates of the center and the radius of the circle. A circle might also be defined as the one that passes through three given points. The circle of reference may be fully defined by three equations of condition. When these are combined with the set of observation equations (eq (5)) data from a single set of fringes is sufficient for a complete solution.

The values of the deviations may be obtained without knowing the value of R. However, when testing a parabolic mirror, for instance, the deviations represent the difference between the parabola and a sphere. If comparison is to be made between the measured deviations of the wave front and the computed deviations of a parabola from a circle, the radius of curvature of the circle must be known. The radius, R, can be measured independently of the interferometer. The ideal image point, E, (fig. 2) or the mean point of convergence of the wave front, can be located and its distance from any chosen point can be measured directly. The center of the reference circle will be located at the ideal image point if the reference circle is chosen as the circle that best fits the observed wave front. This may be done most precisely by the method of least squares. The value of R is the measured distance from the ideal image point (point of convergence) to any chosen reference point plus the computed deviation of this reference point from the reference circle.

In the test of a concave mirror, R is approximately equal to the radius of the mirror. In the test of a lens (simple or compound system), R is the distance from the point of convergence (image point) to the back surface of the lens.

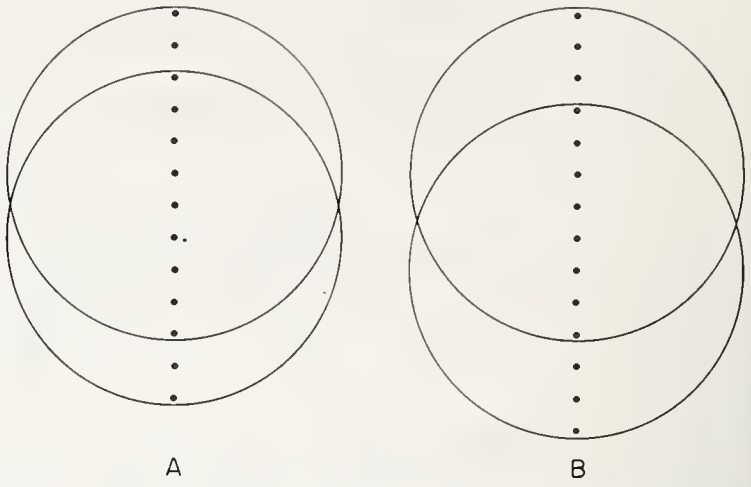


FIGURE 3A *Wave fronts sheared 2 units.*
B *Wave fronts sheared 3 units.*

2. Specifications for the Reference Circle

There are three relatively simple methods of defining the circle of reference. Each method requires three equations of condition. These conditions involve three well-known principles: namely, the method of coincidence, the method of averages, and the method of least squares.

2.1. The Method of Coincidence

We may define the reference circle, shown in figures 1B and 2, by requiring it to pass through any three of the chosen reference points. Thus, we might require that it pass through the two end points, P_0 and P_N, plus some other point near the axis of the mirror (or lens). If the points, P_0 and P_N, are too close to a dubbed edge, a better fitting circle is obtained by requiring it to pass through the two points adjacent to P_0 and P_N, respectively. These two choices for reference circle are made by equating the corresponding δ_ν's to zero. Any three of the points can be chosen to define the circle and the results would be the deviations of the wave front at all chosen reference points from the chosen circle.

For illustration we will choose the fringe pattern shown in figure 1A. There are 8 reference points. We will arbitrarily require the reference circle to pass through points P_0, P_3, and P_7. This is done by requiring that

$$\delta_0=\delta_3=\delta_7=0. \tag{6}$$

For this illustration, eq (5) represents 7 equations of observation. These 7 equations, when combined with the 3 condition equations, (6), form a set of 10 linear equations with 10 unknowns; and all δ_ν's that are not evaluated in eq (6) can be computed in terms of known quantities (the q_ν's and ν's). Table 1 and figure 4 show the results obtained by this method.

2.2. The Method of Averages

The method of averages will usually yield a better fitting circle than the method of coincidence, because of the averaging of the δ_ν's. This method requires an assignment of weights. A reasonable correlation

TABLE 1. *Deviations of a wave front from reference circles that are chosen as follows: Column 3, by the method of coincidence; column 4, by the method of averages; and column 5, by the method of least squares*

ν	q_ν	δ_ν		
		MC	MA	LS
0		0.00	−0.20	−0.10
1	2.74	.23	.13	.14
2	2.10	.15	.11	.07
3	1.70	.00	.00	−.07
4	1.45	−.06	−.05	−.13
5	1.30	.06	.04	−.01
6	1.00	.21	.14	.15
7	0.30	.00	−.16	−.06

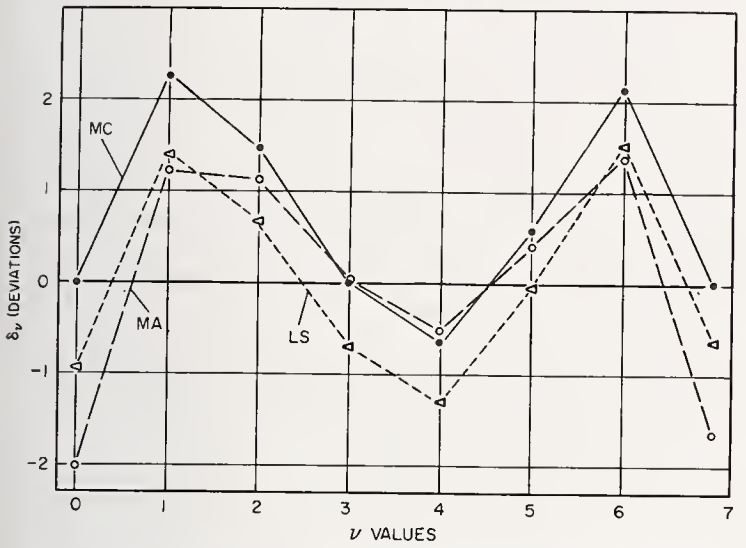


FIGURE 4 *Graphical representation of the deviations of a wave front from reference circles that are chosen by the method of coincidence (MC), the method of averages (MA), and the method of least squares (LS).*

of image quality, formed by the measured wave front from an optical, image-forming system, with wave-front shape would normally require that the weights should be proportional to some function of the distance from the reference points to the center of the circular aperture of the optical system. This would allow the assignment of equal weights to equal areas of the wave front. Equal weights will, however, be used in this discussion.

In previous applications of the method of averages [5], the deviations of the observations from a curve, which is to be fitted to the observations by adjustment of parameters, are equated to functions of the parameters. These are then grouped into as many groups as there are parameters to be adjusted. The algebraic sum of the deviations in each group is equated to zero. When the functions of the parameters (eq (5)) are substituted for the deviations, there are as many equations as parameters and the parameters can be evaluated. Any manner of grouping the deviations will yield a solution. A very poor and sometimes useless solution will be obtained if the points that are allocated to each group are scattered among those of the other two groups. This author prefers to avoid any intermingling of the points of one group with those of other groups.

The parameters of the reference circle shown in figure 1A are ϵ, r and any one of the deviations, δ_ν.

The parameter ϵ (see figs. 1B and 2), together with R, determines the distance, h, from the center, C_1, of the reference circle to the point of convergence, E, of the wave front. The center of this circle is displaced from the ideal image point, E (mean point of convergence) for two reasons. One is that the axis of the mirror or lens cannot be adjusted to pass absolutely through the image of the source, and, secondly, the interference fringes would be too broad for reading when the wave front is nearly spherical. Consequently, the angle ϵ is adjusted to be large enough to provide fringes of a most convenient width for accurate readings.

The parameter, μ, together with R, determines the direction of the point C_1 from E, relative to the axis of the mirror. The parameter r is related to and replaces μ through eq (4). Consequently, the parameters r and ϵ define the center of the circle of reference relative to the wave front.

Any one of the δ_ν's may be chosen as a parameter. By adding δ_ν to the directly measurable distance from E to P_ν we have the radius R of the circle of reference. We will choose δ_0 for the third parameter because of simplification in the formulas that follow.

This relationship of the three parameters to the coordinates of the center of the circle and its radius is somewhat vague, involved, and perhaps not clearly explained, but it is sufficient to understand that they correspond to the three parameters that define the circle of reference.

If equal weights are to be assigned, there must be an equal number of points allocated to each of the three equations of condition. In the example chosen above (fig. 1A), the number of points are not integrally divisible by three. However, we may consider that each of the observation equations represents three identical observations. The total number of observation equations is then $3N$ and division by 3 is now possible. The three equations of condition are:

$$3\delta_0+3\delta_1+2\delta_2=\delta_2+3\delta_3+3\delta_4+\delta_5=2\delta_5+3\delta_6+3\delta_7=0. \quad (7)$$

The set of equations represented by (5) may be replaced by the equivalent set

$$\delta_\nu=\delta_0+\sum_{\sigma=1}^{\nu} q_\sigma+\nu\cdot r-\epsilon\sum_{\sigma=0}^{\nu-1}\sigma=f_\nu, \quad (8)$$

where f_ν is a function of the three parameters, δ_0, ϵ, and r. Any equation in set (8) is obtained by adding the first ν equations in set (5). The functions, f_ν, may now be substituted for the corresponding δ_ν's in (7). It will be noted that there is no f_0 corresponding to δ_0. This, however, does not present any difficulty because δ_0 is itself a parameter and may be left in eq (7) as is.

The three equations of (7), after replacing δ_ν by its equal, f_ν, contain the three parameters as unknowns which may be evaluated. The computed values for the parameters are then substituted in eq (8) for evaluating the remaining δ_ν's. The results

for the chosen example, shown in figure 1A, are given in table 1, column MA, and are also represented graphically in figure 4.

The number of reference points, in the unit-shear method, is determined by the shear angle and diameter of the element being tested. For testing elements (mirrors or lenses) during the polishing and figuring operations, the optician is only interested in departures from a desired figure. The easiest method of computing this is, therefore, the most practical. The shear may be adjusted to such a value that the resultant number of points is some multiple of three. The computation is a minimum when the number of points is the same in each group.

2.3. The Method of Least Squares

The method of evaluating a wave front by least squares is more elaborate than the others but would usually be used when highest precision is desired. In applying this method, the sum of the squared deviations $\left(\sum_{\sigma=0}^{N}\delta_\sigma^2\right)$ is to be minimized with respect to the three parameters required to define the reference sphere.

This application of least squares differs from the usual method of least squares in that the observations are assumed to be free from error and the number of observations needs not exceed the number of unknowns whose values are sought (namely the δ's). In fact, a unique solution may be obtained even if the number of observations is one less than the number of chosen reference points.

To minimize the sum of the squared deviations of the wave front from the reference sphere we require that

$$\sum_{\nu=0}^{N}\delta_\nu^2=\delta_0^2+\sum_{\nu=1}^{N}f_\nu^2=\text{minimum.} \tag{9}$$

This is effected by equating to zero the differentials of eq (9) with respect to r, ϵ, and δ_0, respectively. On performing the differentiation and dividing through by 2, we obtain

$$\delta_0+\sum_{\nu=1}^{N}f_\nu\frac{\partial f_\nu}{\partial\delta_0}=\sum_{\nu=1}^{N}f_\nu\frac{\partial f_\nu}{\partial r}=\sum_{\nu=1}^{N}f_\nu\frac{\partial f_\nu}{\partial\epsilon}=0. \tag{10}$$

The partial differentials, as obtained from eq (8), are

$$\left.\begin{aligned}&\frac{\partial f_\nu}{\partial\delta_0}=1,\ (\nu=0,1,2\ldots N)\\&\frac{\partial f_\nu}{\partial r}=\nu,\ \text{and}\ \frac{\partial f_\nu}{\partial\epsilon}=-\sum_{\sigma=0}^{\nu-1}\sigma,\ (\nu=1,2,3\ldots N)\end{aligned}\right\} \tag{11}$$

On substituting these differentials into (10), we have for the example of 8 reference points (fig. 1A),

$$\left.\begin{aligned}&\delta_0+\delta_1+\delta_2+\delta_3+\delta_4+\delta_5+\delta_6+\delta_7=0\\&\delta_1+2\delta_2+3\delta_3+4\delta_4+5\delta_5+6\delta_6+7\delta_7=0\\&\delta_2+3\delta_3+6\delta_4+10\delta_5+15\delta_6+21\delta_7=0\end{aligned}\right\} \tag{12}$$

On combining sets of eqs (8) and (12), we again have 10 equations with 10 unknowns. Using the same set of fringes (fig. 1A) as was used above for illustration, we obtain for the δ_ν's the values given in column LS of table 1. These values are also plotted, along with those obtained by the other two methods, in figure 4.

3. Sensitivity

The sensitivity of the WSI varies with the absolute magnitude of the shear. If the shear is maintained at unity the sensitivity of the observations increases with a decrease in the number of points because the unit, as defined above, increases. The reduction of the number of points is, of course, limited.

In testing surfaces that depart very greatly from a sphere, as for a large aperture and large aperture-to-focus-ratio parabolic mirror, it is desirable to reduce the sensitivity. Otherwise, such a mirror cannot be tested interferometrically with the WSI. As the number of points is increased, the absolute spacing of the points decreases. The smaller the shear relative to aperture, the smaller can the optical path difference be made. By reducing the absolute magnitude of the shear the maximum optical path difference may be reduced at will and, consequently, the fringes may be adjusted to any desired width.

The sensitivity may also be increased, without reducing the number of reference points, by using two or more shear values that are greater than unity (fig. 3). This may yield more equations than there are unknowns, but averaging of computed points is always acceptable. Each additional set of fringes, obtained with a different shear value, introduces two additional unknowns—new values for ϵ and r.

4. Symmetrical Surfaces

Optical surfaces that are generated mechanically generally have revolution symmetry about the optical axis. If this is known and the axis is known to be at the center of the mirror (assuming circular apertures) then the deviations are known to be symmetrical because the center of the reference circle can be assumed to lie on the axis. Two other conditions are then sufficient to define the reference circle. The condition for revolution symmetry is defined by the specification

$$\delta_\nu=\delta_{N-\nu} \tag{13}$$

and the condition equations, (6), (7) and (12), become, respectively,

$$\delta_0=\delta_3=0, \tag{14}$$

$$3\delta_0+3\delta_1+2\delta_2=6\delta_3+2\delta_2=0, \tag{15}$$

and

$$\left.\begin{aligned}&\delta_0+\delta_1+\delta_2+\delta_3=0\\&21\delta_0+15\delta_1+11\delta_2+9\delta_3=0\end{aligned}\right\}. \tag{16}$$

ERRATA

1. Change the upper limit of the 2nd summation in equation (8) from $\nu - 1$ to ν.

2. Change upper limit of summation in euqation (11) from $\nu - 1$ to ν.

3. Change 3rd line of equation (12) from

$$\delta_2 + 3\delta_3 + 6\delta_4 + 10\delta_5 + 15\delta_6 + 21\delta_7 = 0, \text{ to read}$$

$$\delta_1 + 3\delta_2 + 6\delta_3 + 10\delta_4 + 15\delta_5 + 21\delta_6 + 28\delta_7 = 0$$

4 Change 2nd line of equation (16) from:

$$21\delta_0 + 15\delta_1 + 11\delta_2 + 9\delta_3 = 0, \text{ to read}$$

$$14\delta_0 + 11\delta_1 + 9\delta_2 + 8\delta_3 = 0$$

The third equation in each of the sets (6), (7), and (11) becomes identical, respectively, to the first equation of the set.

The assumption of symmetry (eq (13)) reduces the number of δ_p's by approximately $\frac{1}{2}N$ (exactly $\frac{1}{2}N$ if N is an even number) and reduces the number of condition equations by one. If N is larger than 2, the number of equations exceeds the number of unknowns, for a unit shear arrangement. Consequently, larger shears (2 or more units) may be used with increased sensitivity, as explained above.

5. References

[1] W. J. Bates, A Wavefront Shearing Interferometer, Proc. Phys. Soc. **59,** 940 (1947).

[2] R. L. Drew, A Simplified Shearing Interferometer, Proc. Phys. Soc. (London) sec. B, **64,** 1005 (1951).

[3] Andre Danjon, Mach Interferometer for the Measurement of Fringe Visibility; Arthur Beers, Vistas in Astronomy, p. 383 (Pergamon Press, 1955).

[4] Pol Mollet, Optics in Metrology, The Wavefront Reversing Interferometer, p. 227 (Pergamon Press, 1960).

[5] J. B. Scarborough, Numerical Mathematical Analysis, p. 357 (The Johns Hopkins Press, Baltimore, Md. (1930).

(Paper 65B4–62)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 66B, No. 1, January–March 1962

Measurement of Wave Fronts Without a Reference Standard: Part 2. The Wave Front-Reversing Interferometer

J. B. Saunders

(November 28, 1961)

Interferometers permitting the measurement of shape and altered distribution of fringes are usually used to compare either an unknown wave front with a known one or an unknown surface with a standard reference surface. Any error in the reference surface introduces a corresponding error in the results sought. This paper describes an absolute measuring interferometer and the associated mathematical operations necessary for analysis of the data. The reference surface is purely mathematical in nature and is therefore free from error. Deviations of converging wave fronts from perfect spheres, paraboloids or ellipsoids are readily measured without the use of tangible reference surfaces.

The equation of the reference surface may contain one or more parameters whose value is sought. Thus, the eccentricity of the conicoid that best fits a mirror and deviations of the surface from the true mathematical curve are obtained simultaneously. The sensitivity of the test can be varied so that when testing large aspherical elements which depart very far from spheres the number and width of the fringes can be adjusted to any desired values.

1. Introduction

Many tests made by interferometry consist in measuring the variation in separation of two wave fronts. Usually, the shape of one of the wave fronts is known so that the variation in the observed separation is a measure of the absolute shape of the unknown. The shape of the wave front is a function of the optical elements (lenses, mirrors, etc.) that produce it. Consequently, the aberrations of lenses, the shapes of mirror surfaces, the variations in densities of fluids, etc., are obtainable from the measurement of wave-front shapes.

Usually the known wave front is produced by a mirror of know shape called the reference surface and the unknown wave front which is produced by the specimen is compared with it. If the reference surface is imperfect, any error in it is reflected into, or transferred to, the measured shape of the unknown.

This paper describes a method of comparing wave fronts against a reference surface which is purely mathematical in nature. Since the Koesters double-image prism, which is the basic element of this interferometer, can be made with a high degree of perfection [1] [1] the results are practically free from instrumental error and the necessity for making reference mirrors and lenses of various foci is eliminnated. The interferometer is quite stable, compact, and is easily operated by nontechnical personnel.

This method of analyzing interferometer data, with the necessary modifications required to adapt it to the wave-front-shearing interferometer (≡WSI), has been described in part 1 of this series of papers [2]. Without this development, neither the WSI nor the wave front reversing interferometer (≡WRI), will yield unique results except when applied to wave fronts which are known (from some other type of test or source of information) to have revolution symmetry.

2. Description of the Method

To describe the principles used in this method of analysis its application to the testing of a simple lens will be used for illustration. The operations described here apply, without change, to the testing of lenses by autocollimation (one conjugate at infinity), the testing of mirrors, and to all other tests where the results sought can be obtained from a knowledge of wave-front shapes. This treatment is applicable to three dimensional space but for simplification the mathematics will be limited to a plane.

2.1. Symbols and Abbreviations

The discussions and treatment of this problem require a considerable number of symbols. To provide a ready reference, all quantities used are defined below.

a=diameter of the circular aperture of the lens or mirror under test.
B=dividing plane of the double-image prism.
C and C'=images of conic section.
C_1, C_2, and C_2'=arcs of reference curves representing known wave fronts; (i.e., arcs of circles if a corrected lens or spherical mirror is considered and arcs of a conic section if a comcoid is being considered).
D=position of small, monochromatic light source.
E=center of curvature of prism base.
e=eccentricity of conic section.

[1] Figures in brackets indicate the literature references at the end of this paper.

F=ideal image point (conjugate of D in fig. 1).

F_1, F_2, F'=images of F (F_1 and F_2 are imaged by the outer plane surfaces of the prism and F' is imaged by the dividing plane B.)

F_c and F_c'=centers of curvature of conic section corresponding to points of extreme curvature.

h and k=coordinates of F_c'.

h and k=components of the line from F to E, perpendicular and parallel to the dividing plane, respectively.

$N_{r\sigma}$=observed order of interference at P_ν.

P_ν and P_σ=any chosen pair of superimposed reference points. (These points are equally distant [optically] from the light source and, therefore, fall on the same wave front, Their positions on the wave front are indicated to a sufficient approximation by spots on the lens surface, by beads on a thin wire or string, or by lines on a scale.)

R=radius of curvature of a conic section; minimum for e^2 positive and maximum for e^2 negative.

S=separation of reference curves C_1 and C_2' (or C and C') along the line of sight from E through P_ν.

T, T', T_1, T_1', T_2, T_2'=reference points on a conic section.

u and v=coordinates of E.

V and V'=perihelion of conic section.

W, W_1, W_2, and W_2'=respectively. the wave front before it is incident on the lens or mirror, the part of the transmitted wave front that lies above the dividing plane, the transmitted part that lies below the dividing plane and the image of W_2 with respect to the dividing plane.

WSI=wave-front-shearing interferometer.

WRI=wave-front-reversing interferometer.

x and y=coordinates of T.

α=angle between the dividing plane and the optical axis.

β=angle between the dividing plane and the radius vector to ρ_ν.

δ_ν=separations between C_1 and W_1 along the radius vector to ρ_ν.

δ_σ=separation between C_2 and W_2 along the radius vector to P_σ, (or separation between C_2' and W_2' along the radius vector to P_ν).

ϵ=angle between two images of conic section.

ϵ_μ=angle between C_1 and C_2' at their point of intersection with the dividing plane.

λ=wavelength of light.

μ=distance from the extension of the dividing plane to the axial center of the lens (fig. 1). It is positive when the dividing plane is above the axis of the lens and negative when below.

ν and σ=distances from P_ν and P_σ, respectively, to the optical axis. These distances are measured along the arc of the reference curve, C_1C_2. They are positive when the corresponding points are above the axis of the lens and negative when below.

ρ, ρ', ρ_{c1}, and ρ_{c2}=radius vectors from E to T, T', T_1, and T_2, respectively.

ρ_ν and ρ_σ=radius vectors from E to P_ν and P_σ. respectively.

ϕ=the sum of the squares of the deviations (defined by eq (13)).

2.2. Apparatus

A simple interferometer for testing lenses, without a reference surface or element for comparison, is shown in figure 1. It consists of a small monochromatic light source at D, the lens to be tested and a double-image prism [1]. The lens received the diverging wave front, W (spherical when three dimensional space is considered and circular for two dimensional space) and changes it into a converging wave front. If the lens were perfectly corrected for the indicated position of the source, the converging wave front would be an arc of a circle (curve C_1C_2) and the light would form a perfect image (airy disk) of the source at the ideal image point F. The wave front produced by an imperfect lens (greatly exaggerated) is indicated by curve W_1W_2 and its deviation from the ideal wave front, C_1C_2, is a measure of the abberations of the lens. The circle that represents the ideal wave front exists in mathematical form only and its parameters are adjusted statistically or otherwise chosen so that it fits the real wave front in accordance with some arbitrarily selected specification. Methods of choosing the reference circle will be described later.

In practice the prism is mounted so as to pivot about E, and E is placed near F. The semi-reflecting dividing plane, B, is adjusted so that its extension cuts the circular lens along a chord at a chosen distance, μ, from the axis of the lens. Each new adjustment of the prism (change in value of μ) introduces and additional arbitrary value for ϵ_μ. The quantity μ is positive when the dividing plane is above the axis of the lens and negative when below.

The effect produced by this prism on the wave-front is as follows: the wave front W, after division by wave-front division into two parts, W_1 above the dividing plane and W_2 below, is further subdivided by amplitude division at B into two beams each, making four beams altogether, two arriving at each of the points F_1 and F_2. One of the beams that arrives at each of these points suffers two reflections and the other a single reflection. Consequently, a mirror image of everything seen below the dividing plane appears above it and vice versa. The arc C_2 (fig. 2A) appears as C_2'. The center of curvature of C_2, [3] located at F, is imaged at F' which is the center of curvature of C_2'. The part W_2 of the wave front below the dividing plane appears as W_2', so that W_1 and W_2 appear inverted with respect to each other and superimposed in the overlapping region W_2', where fringes of interference (fig. 2B) are observed.

The two halves of the prism are slightly rotated about an axis normal to the dividing plane (as described in references 1 and 6). A wave front having symmetry with respect to the dividing plane would therefore give rise to straight fringes normal to the line representing the dividing plane in the field of view. If the two halves of the prism are not rotated relative to each other the fringes are straight and parallel to the dividing plane.

2.3. Equations of Observation

The separation of wave fronts, W_1 and W_2' (fig. 2A), is the optical path difference $\rho_\nu - \rho_\sigma$ which is

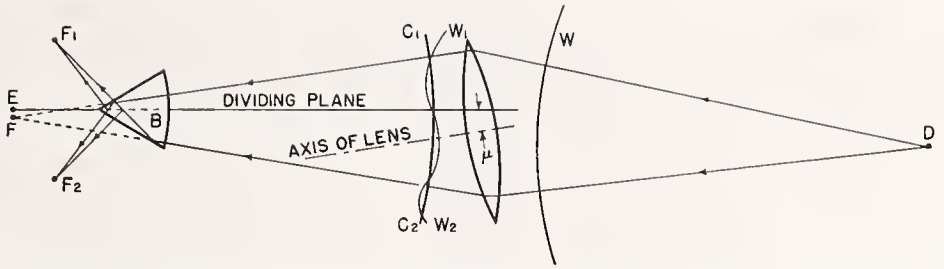


FIGURE 1. *Optical arrangement for testing a lens for a finite object distance.*
This arrangement requires the use of polarized light because of polarized effects in the prism.

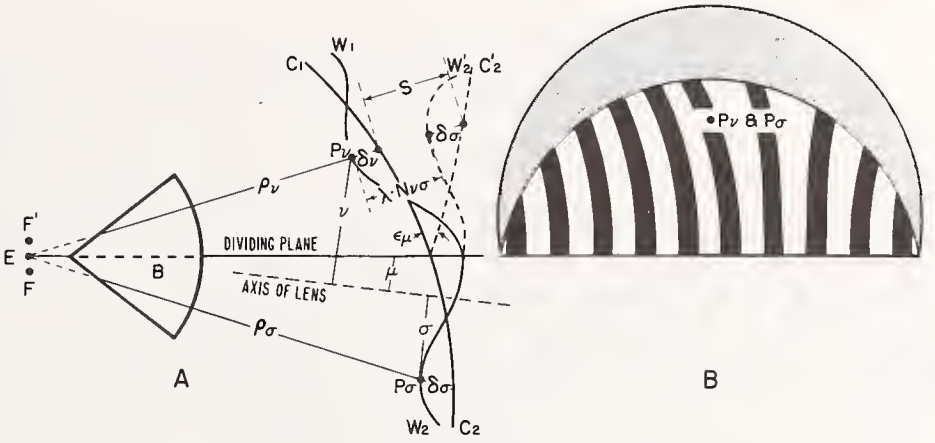


FIGURE 2. A. *Illustration of image formation by the double-image prism.*
B. *Interference fringes obtained with the assembly of figure 1.*

equal to the produet $\lambda N_{\nu\sigma}$. The corresponding separation of the referenee cirele C_1 and C_2' at P_ν is, approximately [2].

$$S=(\nu-\mu)\epsilon_\mu=\epsilon_\mu(\mu-\sigma). \tag{1}$$

The two quantities in parentheses are the distanees, along the cirele, from P_ν and P_σ, respeetively, to the dividing plane. Sinee P_σ, in figure 2, is below the axis of the lens, σ is a negative quantity.

It is apparent from figure 2A that the relation between the four separations, δ_ν, δ_σ S, and $\lambda N_{\nu\sigma}$ is given by the equation

$$\delta_\nu+S=\delta_\sigma+\lambda N_{\nu\sigma}. \tag{2}$$

See seetion 2.1 for definitions of these quantities. On substituting for S its value from eq (1) and transferring terms we have

$$\delta_\nu-\delta_\sigma=\lambda N_{\nu\sigma}-\epsilon_\mu(\nu-\mu) \tag{3}$$

which are the basie equations of observation, providing relationships between the deviations of the wave front from the abstraet eurve C_1C_2.

2.4. Distribution and Number of Reference Points

If the surfaees of the lens are figures of revolution, the optie axis is well defined, and ean be aeeurately loeated by observing the fringe pattern while ehanging the value of μ. When the dividing plane eoineides with the optie axis the fringes will be straight regardless of aberrations in the lens. If the lens surfaces are not figures of revolution the changes in curvature and spaeing of the fringes will indicate the most probable position of the axis.

The referenee points are ehosen in the aperture of the lens, along a straight line normal to the dividing plane of the prism and through the optieal axis, regardless of whether the axis is or is not eentered in the aperture of the lens. The positions of these points are measured from the optieal axis. They are positive when above and negative when below the axis. The points are equally spaeed and their separation is defined as unity. An odd or an even number of reference points, symmetrically plaeed above and below the axis, may be used. An odd set, therefore, would be loeated at integral units of length from the axis, as in figure 3–A, while an even set would be plaeed at odd integral half units of length from the axis, as shown in figure 3–B. If a point at the center is desired one must ehoosc an odd number of points. Figure 3–C illustrates the arrangement of an odd set when the optie axis is not eentered in the aperture. The interferometer is sufficiently sensitive to deteet deviations of the position of the axis from the center of the aperture in most lenses.

Any number of points may be seleeted along the prineipal diameter (the diameter that is normal to the dividing plane) by adjusting the separation of adjaeent points. If the mirror or lens to be tested is unsymmetrical and/or irregular the test should be made with more points than when a smooth wave front is formed.

Even when the wave front is smooth and symmetrical about the axis, many referenee points will sometimes be required. The number of fringes in the field of view depends upon three quantities. These are: (a) The magnitude of the wedge that is built into the prism, (b) the angle between the wave fronts at the dividing plane, and (e) the relative shapes of the two images, W_1 and W_2', of the wave front. The effects of these three quantities are more fully described as follows:

(a) The built-in wedge of the prism produees an invariant wedge between the wave-front images. Its

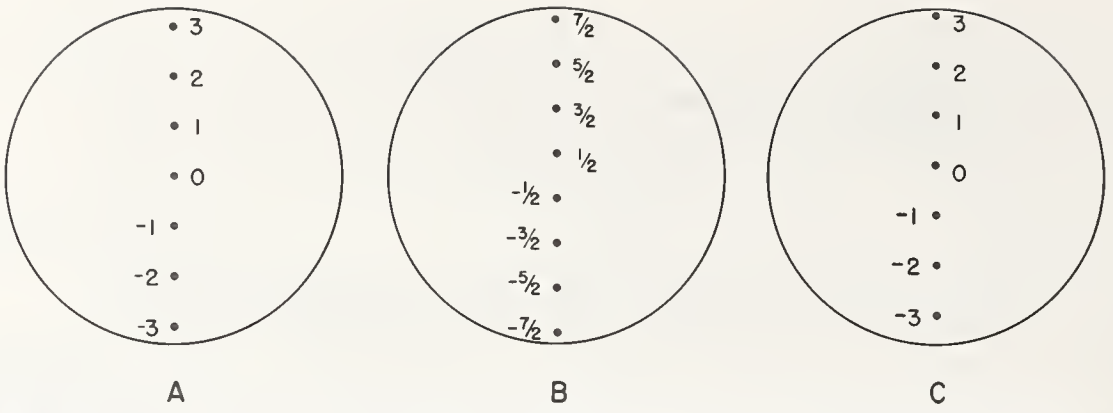


FIGURE 3. *A, B, and C are, respectively, the distribution of the reference points when the number is odd, when the number is even, and when the optic axis is not centered in the lens.*

vertex is perpendicular to the dividing plane. Since the component of this wedge, along the line of reference points, is zero, its effect on the interference at the reference points is to add a constant to the observed orders of interference at all reference points. If the vertex of the built-in wedge is adjusted to coincide with the line along which the reference points are to be selected, this wedge does not affect the order of interference at the reference points. It does, however, ensure that the fringes never get too broad for precision readings.

(b) The angle, ϵ_μ, between the wave fronts at the dividing plane is equal to the angle between lines drawn from points F and F' to the vertex of this angle (see fig. 13, A). This angle is adjustable by varying the position of F relative to E, by displacing the prism laterally. Thus the component of fringe widths perpendicular to the dividing plane and the order of interference at any chosen reference point can be adjusted to any desired value. If the two wave fronts, W_1 and W_2', are identical in shape (i.e., if wave front W_1W_2 is spherical) the difference in order of interference between all pairs of adjacent points is approximately constant and adjustable at will.

(c) If the wave front W_1W_2 is not spherical the two images of it, W_1 and W_2', will not be identical unless W_1W_2 has revolution symmetry and the value of μ is zero. When testing a wave front that departs very far from a sphere, as, for instance, when testing an $f/1$ parabola of large aperture, the minimum number of fringes that can be obtained by adjusting ϵ_μ depends upon μ. Since, in operation of the instrument, pairs of points are brought into coincidence for conditions other than $\mu=0$, and since the absolute magnitude of μ depends upon the unit separation of points, it follows that the choice of the number of reference points determines the minimum attainable number of fringes. If the number of reference points is too small the minimum attainable number of fringes may be too large for accurate readings. Consequently, a large number of points may be required to obtain the optimum reading condition. Since the sensivity of the interferometer depends upon the ratio μ/a, the sensivity also depends upon the number of reference points. The variation of sensivity with the ratio μ/a and its effects on the required number of reference points are fully explained under "Sensitivity" in part 1 of this series [2].

2.5. Solution of Wave-Front Deviations

We shall select (fig. 4) quite arbitrarily for this discussion, seven reference points along the principal diameter of the lens and with their spacings subtending equal angles at E. These points will be considered as lying on a common wave front at the time it emerges from the lens, and their separations will be defined as one unit of length. Since the points are observed from position E they may be indicated with small beads on a fine wire that is stretched across the face of the lens (fig. 4A).

If the dividing plane is adjusted to include one of the inner points (any point except an end one) as in figure 4C, some of the points will appear to coincide in pairs. If it is adjusted to bisect the angle between any two of the inner points, as in figure 4B, some of the points again appear to coincide in pairs. The points will coincide in pairs when μ is an integral number of half units of length but will not coincide for other values of μ.

We may substitute, in eq (3), the observed values for $N_{\nu\sigma}$ and $(\nu-\mu)$ at each pair of points, successively. This yields as many equations of observation as there are pairs of coinciding points. We may then change the value of μ, thereby changing the combination of paired points, and obtain additional equations. Each new value for μ (i.e., $\mu=0, \pm 0.5$ ± 1.0, . . .) introduces several additional equations but only one additional unknown—i.e., a new value for ϵ_μ. In this manner, we can obtain more equations than we have unknowns; and it would seem that a statistical analysis of these equations would permit an evaluation of the several deviations (values for the δ's). However, such is not the case. The relation between these equations are such that the resultant number of independent equations is always less by three than the number of unknowns.

When the wave front is completely unknown, the minimum number of observation equations that permit a unique solution of the δ's require two sets of fringes (two values for μ). Each adjustment of the interferometer introduces a new value for ϵ_μ because the value of ϵ_μ can not be retained from

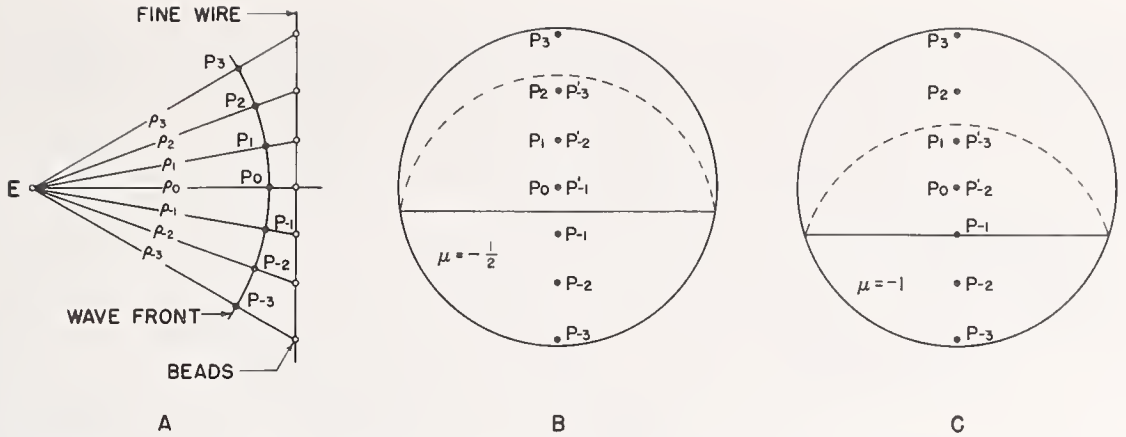


FIGURE 4. A. *Illustration to show the manner of locating the reference points,* Pν, *by means of beads on a fine wire.*
B. and C. *Distribution of points and images of points for* $\mu=-0.5$ *and* $\mu=-1$, *respectively.*

one adjustment to another. For simplicity we will use the two adjustments (2 sets of fringes) that are obtained with $\mu=0$ and $\mu=0.5$. These are (for the 7 chosen reference points) for $\mu=0$

$$\delta_1=\delta_{-1}+\lambda N_{1,-1}-\epsilon_0, \tag{4}$$
$$\delta_2=\delta_{-2}+\lambda N_{2,-2}-2\epsilon_0, \tag{5}$$
$$\delta_3=\delta_{-3}+\lambda N_{3,-3}-3\epsilon_0, \tag{6}$$

and for $\mu=0.5$

$$\delta_1=\delta_0+\lambda N_{1,0}-0.5\epsilon_{0.5}, \tag{7}$$
$$\delta_2=\delta_{-1}+\lambda N_{2,-1}-1.5\epsilon_{0.5}, \tag{8}$$
$$\delta_3=\delta_{-2}+\lambda N_{3,-2}-2.5\epsilon_{0.5}. \tag{9}$$

Equations 4 through 9, form a set of 6 simultaneous equations having 9 unknowns (7 δ's and 2 ϵ_μ's); consequently, we either have an excess of 3 unknowns or a shortage of 3 equations for obtaining a unique solution. It is quite obvious that we cannot evaluate deviations from an unidentified curve. The persistent shortage of three equations is associated with the three conditions necessary to define an arbitrarily chosen 3-parameter reference circle.

The method used in combining the three condition equations with the set of observation equations is fully described under the heading "Specifications for the Reference Circle" that appears in part 1[2]. Since the details of the application to the WRI are so similar to those given for the WSI, this treatment will be limited to details that are necessary for a clear understanding of the WRI. Since three points fully define a circle we may define the reference circle, C_1C_2, as that particular circle which passes through any three of the chosen points, by equating the corresponding δ's to zero. Thus, if we impose the three conditions,

$$\delta_3=\delta_0=\delta_{-3}=0 \tag{10}$$

we are requiring that the reference circle shall pass through points P_3, P_0, and P_{-3}, as indicated in figure 5B. This may be considered as either a reduction in the number of unknowns by three or an increase of three in the number of independent simultaneous equations. In either case we get as many equations as unknowns so that a unique solution of all unknowns is possible. Any three of the chosen reference points may have been selected to fall on the circle. The simplest manner of specifying and computing the reference circle (called the "Method of Coincidence") may not produce a good fitting circle; but this does not render any less obvious the location of the high and low places on the wave front. Consequently, an optical worker would have no trouble locating the high and low places for further polishing to reduce the deviations.

A better fitting reference circle is obtained by applying the method of averages [4]. Figure 5C shows the same wave front with a circle of reference that is determined by this method. The method of averages allows a variety of groupings of points and assignments of weights. For instance, since the areas of the wave fronts that are associated with the several points increases with ν we might multiply all δ's in eq (10) by the quantity $(\nu+1)$ so as to assign unit weight to the axial point, P_0, and corresponding higher weights to other points, thus distributing the weight more nearly proportional to the areas of the several zones.

A good fit, based on the method of averages, is obtained by imposing the following simple conditions:

$$\left.\begin{aligned}\delta_3+\delta_2&=0\\ \delta_1+\delta_0+\delta_{-1}&=0\\ \delta_{-2}+\delta_{-3}&=0.\end{aligned}\right\} \tag{11}$$

The circle that fits the wave front shown in figure 5, in accordance with the conditions of eq (11) is shown in figure 5C. Equations (11), when added to the six equations of observation (eq (4) through (9)), again supply the necessary additional independent equations for a complete solution of all δ's.

The best fitting circle, shown in figure 5D, is defined by a third set of three conditions,

$$\frac{\partial\phi}{\partial\delta_0}=\frac{\partial\phi}{\partial\epsilon_0}=\frac{\partial\phi}{\partial\epsilon_{0.5}}=0 \tag{12}$$

where

$$\phi\equiv\sum_{\nu=-3}^{3}\delta_\nu^2=\delta_{-3}^2+\delta_{-2}^2+\delta_{-1}^2+\delta_0^2+\delta_1^2+\delta_2^2+\delta_3^2 \tag{13}$$

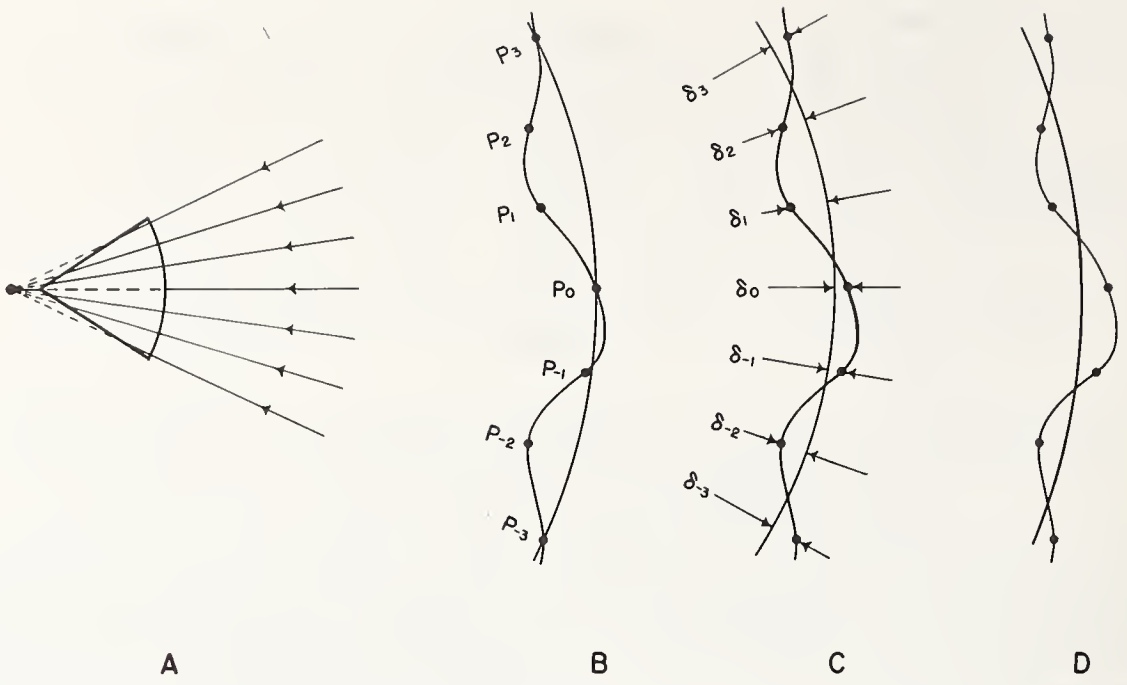


FIGURE 5. A. *Illustration to show the distribution of the reference points relative to the prism.* B, C, and D. *The positions of the reference circles relative to the wave front when the reference circle is chosen in accordance with the method of coincidence, the method of averages, and the method of least squares, respectively.*

which requires that the sum of the squares of the deviations be a minimum. Differentiation with respect to any other three of the nine unknowns (i.e., 7 δ's and 2 ϵ_μ's) will yield the same set of condition equations.

In order to determine the equations of condition corresponding to eq (12), it will be convenient to solve for the δ_ν's. These are:

$$\delta_0=\delta_0 \quad \text{(obvious)} \tag{14}$$

$$\delta_1=\delta_0+\lambda N_{1,0}-0.5\epsilon_{0.5} \text{ (from 7)} \tag{15}$$

$$\delta_{-1}=\delta_0+\lambda(N_{1,0}-N_{1,-1})-0.5\epsilon_{0.5}+\epsilon_0 \text{ (from 4 and 15)} \tag{16}$$

$$\delta_2=\delta_0+\lambda(N_{1,0}-N_{1,-1}+N_{2,-1}) -2.0\epsilon_{0.5}+\epsilon_0 \text{ (from 8 and 16)} \tag{17}$$

$$\delta_{-2}=\delta_0+\lambda(N_{1,0}-N_{1,-1}+N_{2,-1}-N_{2,-2}) -2.0\epsilon_{0.5}+3\epsilon_0 \text{ (from 5 and 17)} \tag{18}$$

$$\delta_3=\delta_0+\lambda(N_{1,0}-N_{1,-1}+N_{2,-1}-N_{2,-2}+N_{3,-2}) -4.5\epsilon_{0.5}+3\epsilon_0 \text{ (from 9 and 18)} \tag{19}$$

$$\delta_{-3}=\delta_0+\lambda(N_{1,0}-N_{1,-1}+N_{2,-1}-N_{2,-2}+N_{3,-2}-N_{3,-3}) -4.5\epsilon_{0.5}+6\epsilon_0 \text{ (from 6 and 19)} \tag{20}$$

On performing the differentiations indicated in eq (12), and substituting the corresponding values for these differentials from eqs (14) through (20), we obtain the set of condition equations for the method of least squares. These are:

$$\left.\begin{aligned}&\delta_3+\delta_2+\delta_1+\delta_0+\delta_{-1}+\delta_{-2}+\delta_{-3}=0\\&\delta_{-1}+\delta_2+3(\delta_{-2}+\delta_3)+6\delta_{-3}=0.\\&(\delta_1+\delta_{-1})+4(\delta_2+\delta_{-2})+9(\delta_3+\delta_{-3})=0\end{aligned}\right\} \tag{21}$$

Any one of the sets of conditional equations (given by eqs (10), (11), or (21) combined with the observation equations (4) through (9)) permits a complete solution of all unknowns. The operations that lead to a solution of these equations are performed as follows: (1) By adjusting the interferometer so that $\mu=0$ (i.e., with the dividing plane cutting the wave front along a line through the axis of the lens; (2) substituting the observed values for $(\nu-\mu)$ and the observed orders of interference at each of the three pairs of coinciding points into eq (3) thus obtaining eqs (4), (5), and (6): (3) readjusting μ to equal 0.5 (fig. 4B): (4) again substituting the observed quantities $N_{\nu\sigma}$, ν and μ into eq (3), thus obtaining eqs (7), (8), and (9): (5) solving for the 9 unknowns in the 9 linear, simultaneous equations (eqs (4) through (9) and any one of (10), (11), or (21)).

2.6. Symmetrical Systems

Lenses and mirrors that have been polished mechanically produce symmetrical wave fronts when tested on axis. For such symmetrical elements all δ's on one side of the axis of symmetry are equal to corresponding δ's on the other side and for $\mu=0$

the fringes of interference will be straight. All δ's with negative subscripts can, therefore, be replaced with the corresponding positive subscripts. In the case of seven reference points, chosen here for illustration, the left hand side of eqs (4), (5), and (6) becomes zero and the right hand side of each of these equations represents a solution of ϵ_0. Since ϵ_0 is superfluous, the adjustment of μ to zero for a symmetrical wave front serves only as a check on symmetry. Consequently, if symmetry is observed, eqs (4), (5), and (6) are exempted from the group of simultaneous equations. The elimination of negative subscripts reduces the number of unknowns so that a single adjustment (μ=0.5, eqs (7), (8) and (9) of the interferometer is sufficient for a complete solution. Equations (4), (5), and (6) are eliminated and eqs (10), (11), and (12) for the three methods, are reduced, respectively, to $\delta_3=\delta_0=0$, $\delta_3+\delta_2=2\delta_1+\delta_0=0$, and $\frac{\partial\phi}{\partial\epsilon_{0.5}}=\frac{\partial\phi}{\partial\delta_0}=0$.

The number of equations in the complete set is reduced from 9 to 5 with a corresponding reduction in the number of unknowns.

2.7. Illustration

The process described above is illustrated in the following application to the testing of a 12-in. aperture, 15-ft focal length astronomical refractor. Figure 6 shows the optical arrangement for testing with one conjugate at infinity. Figure 7A is a photograph of the fringes taken with the dividing plane coinciding with the optic axis of the lens (μ=0) and 7B shows the fringes when $\mu=\frac{1}{2}$ unit (=2 cm). Eight reference points were used for this test. These were located at distances of ±2, ±6, ±10, and ±14 cm from the optical axis. The unit separation is 4 cm. Therefore the absolute values for ν and σ are 0.5, 1.5, 2 5, and 3.5 units. Table 1 shows the observed orders of interference for the two sets of fringes at the pairs of reference points indicated by the ν and σ columns. The straightness of the fringes in figure 7A shows the wave to be symmetrical about the dividing plane. When straight fringes are obtained for several azimuthal positions of the lens (by rotating it relative to the prism about its optical axis) the test indicates the lens to be a figure of revolution. This means that $\delta_\nu=\delta_\sigma$ and eqs (4), (5), and (6) are not required. The solution for the δ's can, therefore, be obtained by using only three observation equations and two equations of condition. The equations of condition may be based on either one of the three criteria: i.e., method of averages, method of coincidence, or method of least squares.

We shall arbitrarily choose to use the method of averages and select the reference circle by requiring that

$$\left.\begin{aligned}\delta_{0.5}+\delta_{1.5}&=0\\ \delta_{2.5}+\delta_{3.5}&=0.\end{aligned}\right\} \tag{22}$$

The equations of observation are obtained by substituting corresponding values for ν, σ, μ, and $N_{\nu\sigma}$, shown in table 1, in the columns under (μ=0.5), into eq (3). These are:

$$\left.\begin{aligned}\delta_{1.5}&=\delta_{0.5}+0.55 \quad -1 \quad \epsilon_{0.5}\\ \delta_{2.5}&=\delta_{1.5}+0.20 \quad -2 \quad \epsilon_{0.5}\\ \delta_{3.5}&=\delta_{2.5}-0.35 \quad -3 \quad \epsilon_{0.5}.\end{aligned}\right\} \tag{23}$$

Equations (22) and (23) contain 5 unknowns, 4 δ's and $\epsilon_{0.5}$. The simultaneous solution of these equations for the δ's yields the deviations of the wave front from a sphere. These values are -0.23, $+0.23$, $+0.28$, and -0.28, respectively, for $\delta_{0.5}$, $\delta_{1.5}$, $\delta_{2.5}$, and $\delta_{3.5}$. Since the light passes through the lens twice in this test these deviations are twice the values that would have been obtained for one

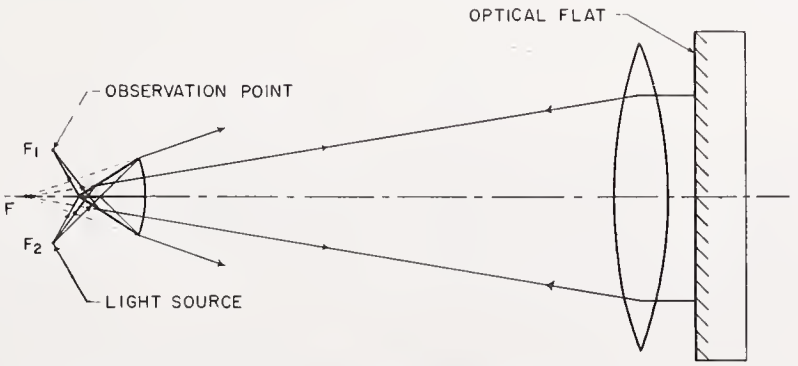


FIGURE 6. *Optical arrangement for testing a lens with one conjugate at infinity.*

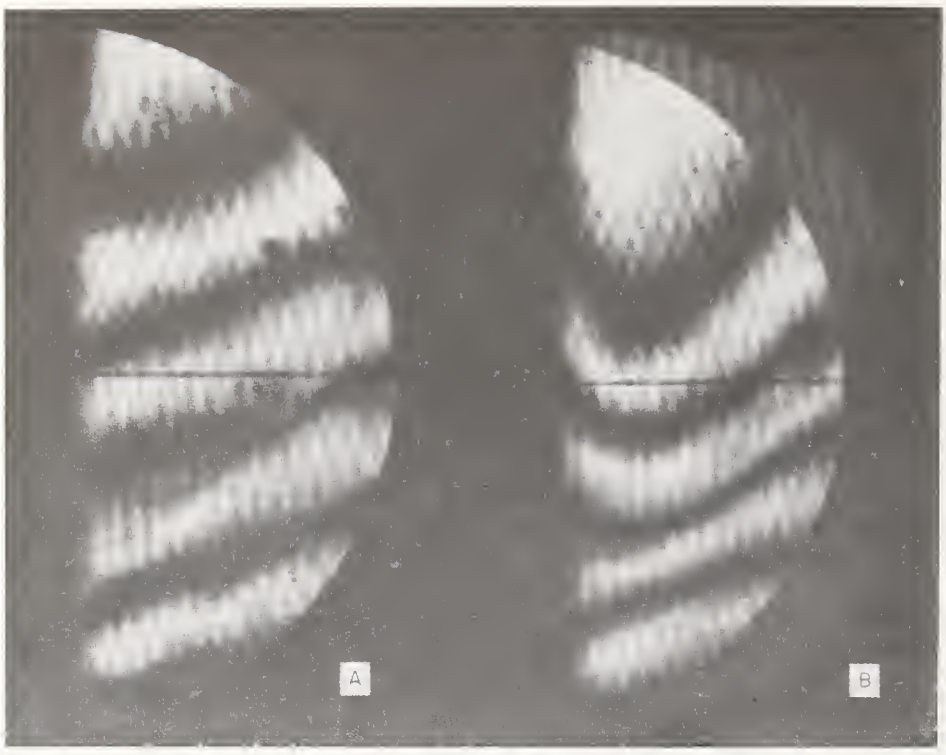


FIGURE 7. *Photographs of fringes with 12-inch aperture, f/15 astronomical objective when* (A) $\mu=0$, *and* (B) $\mu=0.5$.

TABLE 1. *Data from figure 7 for analysis of a 12-inch aperture astronomical objective*

$\mu=0$			$\mu=0.5$		
ν	σ	$N_{\nu\sigma}$	ν	σ	$N_{\nu\sigma}$
0.5	−0.5	−0.10	0.5	0.5	0.00
1.5	−1.5	−0.45	1.5	−0.5	0.55
2.5	−2.5	−0.70	2.5	−1.5	0.20
3.5	−3.5	−1.00	3.5	−2.5	−0.35

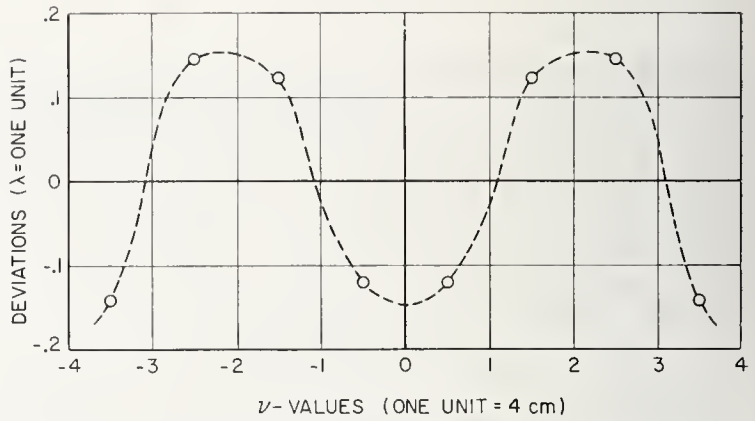


FIGURE 8. *Graphical representation of wave-front-shape computed from the photographs of figure 7.*

transmission through the lens—as for light from a celestial star. The deviations for one transmission are represented graphically in figure 8. The curve indicates that when this objective is used for astronomical work the wave front departs from a perfect sphere by approximately one-third of a wavelength. The light used for this test was the green line of the mercury spectrum ($\lambda=0.546$ microns).

If eq (3) is used in the testing of a parabolic mirror the computed deviations will be the deviations of the mirror from a perfect sphere. Usually, when testing a parabolic mirror one wishes to know how much the mirror differs from a perfect paraboloid. If the mirror is ellipsoidal one wishes to know both the deviations and the eccentricity. The eccentricity, e, enters as an additional parameter (eq (30) of appendix A).

The simultaneous evaluation of the deviations and the eccentricity is obtained with the method of coincidence by requiring the curve to pass through 4 points instead of 3 as was the case (eq (10)) for a 3-parameter reference curve. If the method of averages is to be used the reference points are grouped into 4 groups yielding 4 equations of condition instead of the 3 shown in eq (11). Similarly, with the method of least squares an additional equation ($\partial\phi/\partial e=0$) is to be added to the three equations under eq (12). If the surface is known, or can be assumed, to be a figure of revolution about its axis, the number of condition equations is reduced by 1, as for the circular reference curve.

Figure 9 is a photograph of fringes taken with a 12-inch aperture, 84-inch focal length, mirror that was claimed to be a nearly perfect paraboloid. Figure 10 shows the optical arrangement used for making this photograph. The computed shape of this mirror indicated that it conformed more closely to an ellipsoid than to a paraboloid. The general equation for an ellipse (described in appendix A, eq (31)) was

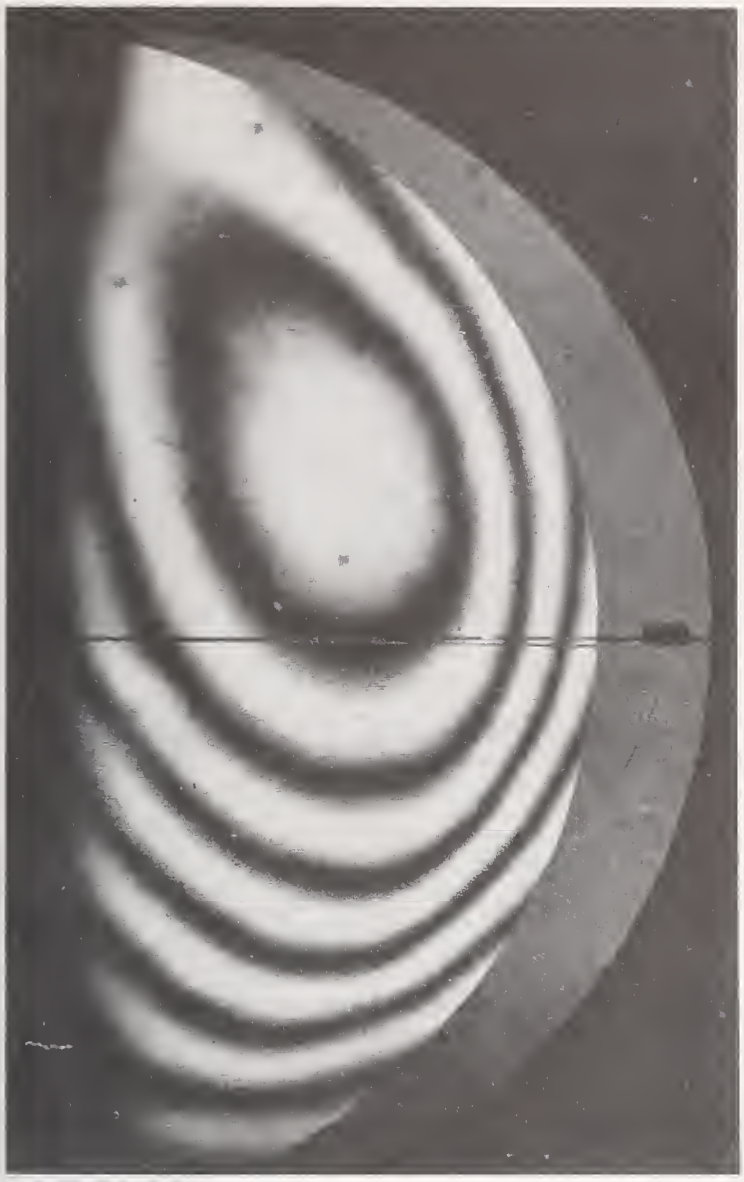


FIGURE 9. *Photograph of fringes taken with a 12-inch aperture f/7 parabolic mirror.*

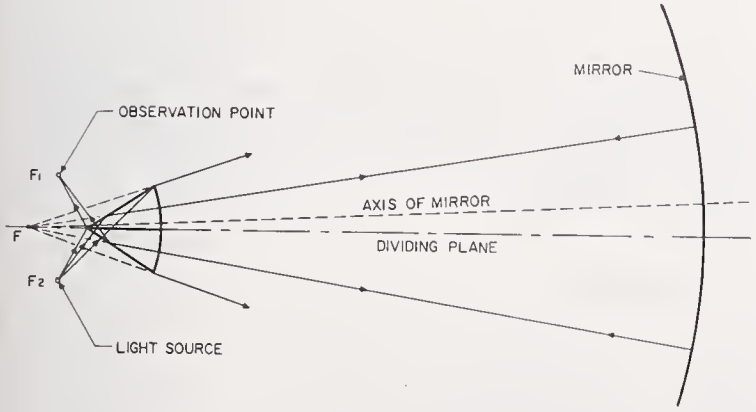


FIGURE 10. *Optical arrangement for testing concave mirrors at their centers of curvature.*

FIGURE 11. *Photograph of two Koesters double-image prisms, one mounted in a cell and the other untrimmed.*

then tried and the computed value for the eccentricity corresponded to an ellipse whose foci are separated by 89 meters. It was subsequently learned that the final test on this "Nearly Perfect Parabolic Mirror" was made with a Foucoult knife edge test, with the source located 90 meters (assumed to be practically at infinity) from the mirror and the knife edge at the point of convergence or image of the source. This test, therefore, confirms the description of the mirror as being a "nearly perfect parabola" but even more nearly a perfect ellipsoid.

Figure 11 shows photographs of the prism, before and after it is fitted into a cell. This prism has been made in sizes from approximately 6 inches down to a fraction of an inch. The one shown here has a one inch aperture and is considered a most favorable size for many applications.

2.8. Summary

The interferometer described here for testing wave fronts is believed to be easier to operate than any previous interferometer yet developed for this purpose. Its operation is simple. Unskilled personnel can be trained to operate it after a short period of training..

This interferometer yields results that are absolute. It does not require a standard of reference. The absolute evaluation of wave-front shapes, however, depends upon a unique method of analyzing the data. The analysis involves the development of a set of linear simultaneous equations. This requires some technical background knowledge. After the equations for a given assembly are developed, unskilled personnel can compute wave-front shapes by substituting observed orders of interference into the equations and solving for the deviations from the chosen reference curve. The chosen reference curve will usually be a circle but more complicated curves may be used.

518-743 O - 73 - 12

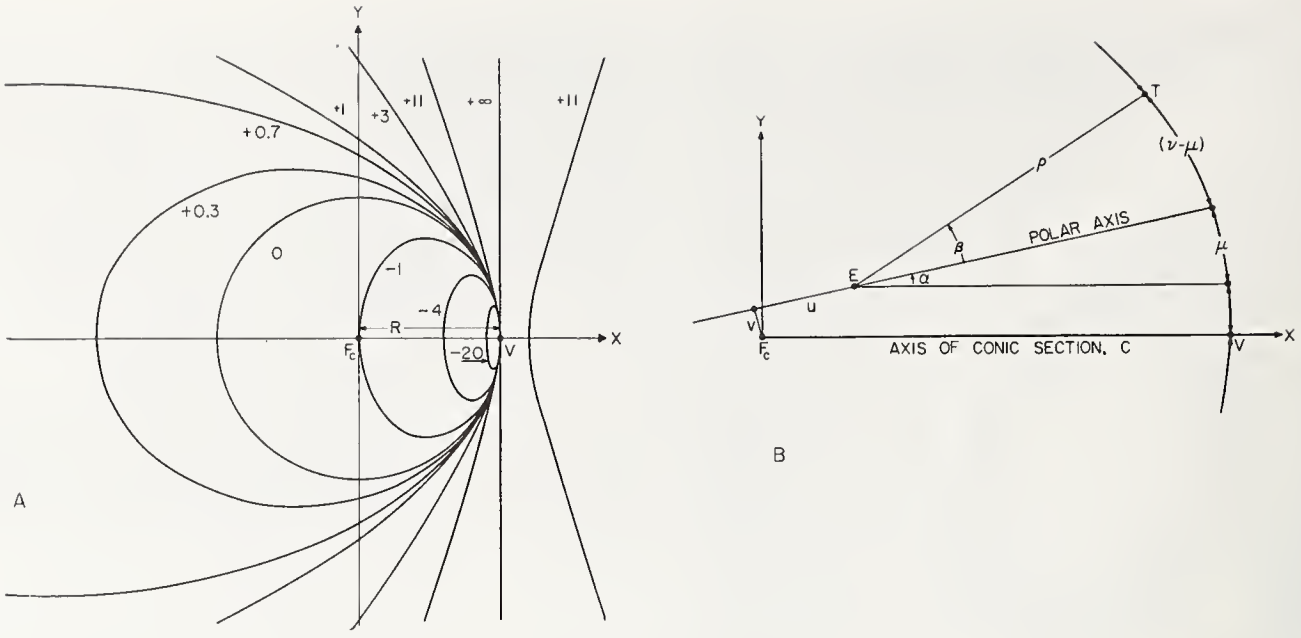


FIGURE 12. *A. The family of curves representing the conic section for various indicated values of e^2 (e=eccentricity). B. Illustration for describing the transformation from rectangular to polar coordinates.*

3. Appendix A. Optical Path Difference Between Any Two Points on a Conic Section to Any Point in the Neighborhood of the Center of Curvature of the Section at Its Vertex

The equation for the conic section in rectangular coordinates, with the origin at F_c, (a distance R to the left of the curve, see fig. 12A), and the x-axis coinciding with the axis of the conic, is given by the 2 parameter equation

$$y^2+(1-e^2)(x-R)^2+2R(x-R)=0. \tag{24}$$

The family of curves, shown in figure 12A represents this equation for the various indicated values of e^2. The parameter R is the radius of curvature of the curves at the vertex V (except for $e^2=\pm\infty$), and e is the eccentricity.

If we shift to polar coordinates (ρ, β) with the polar axis making angle α with the axis of the conic (see fig. 12B) and the pole located at point E, which is displaced from F_c by a small amount u parallel to the polar axis and a small distance v perpendicular to it, we may obtain the polar equation for the conic by means of the transformation equations,

$$\left.\begin{aligned} y&=u\sin\alpha+v\cos\alpha+\rho\sin(\alpha+\beta)\\ x&=u\cos\alpha-v\sin\alpha+\rho\cos(\alpha+\beta).\end{aligned}\right\} \tag{25}$$

The quantities u and v will never exceed the separation of points E and F_c, which is always quite small relative to ρ, because these two points can always be superimposed to a high degree of accuracy.

On eliminating x and y from eqs (24) and (25); solving for ρ; applying binomial expansions to eliminate radicals and negative power terms; expanding $\sin(\alpha+\beta)$ and $\cos(\alpha+\beta)$ in powers of $(\alpha+\beta)$ [5]; and dropping all power terms in u and v higher than unity (because these terms are always insignificantly small), we obtain

$$\left.\begin{aligned} \rho&=f_1(\alpha,\beta)-uf_2(\alpha,\beta)-vf_3(\alpha,\beta),\\ &\text{where}\\ f_1(\alpha,\beta)&=R+1/8\,Re^2(\alpha+\beta)^4-1/48\,Re^2(1+3e^2)(\alpha+\beta)^6+\ldots\\ f_2(\alpha,\beta)&=\cos\beta+1/2\,e^2\beta(\alpha+\beta)^4+\ldots\\ f_3(\alpha,\beta)&=\sin\beta-1/2\,e^2(\alpha+\beta)^4+\ldots\end{aligned}\right\} \tag{26}$$

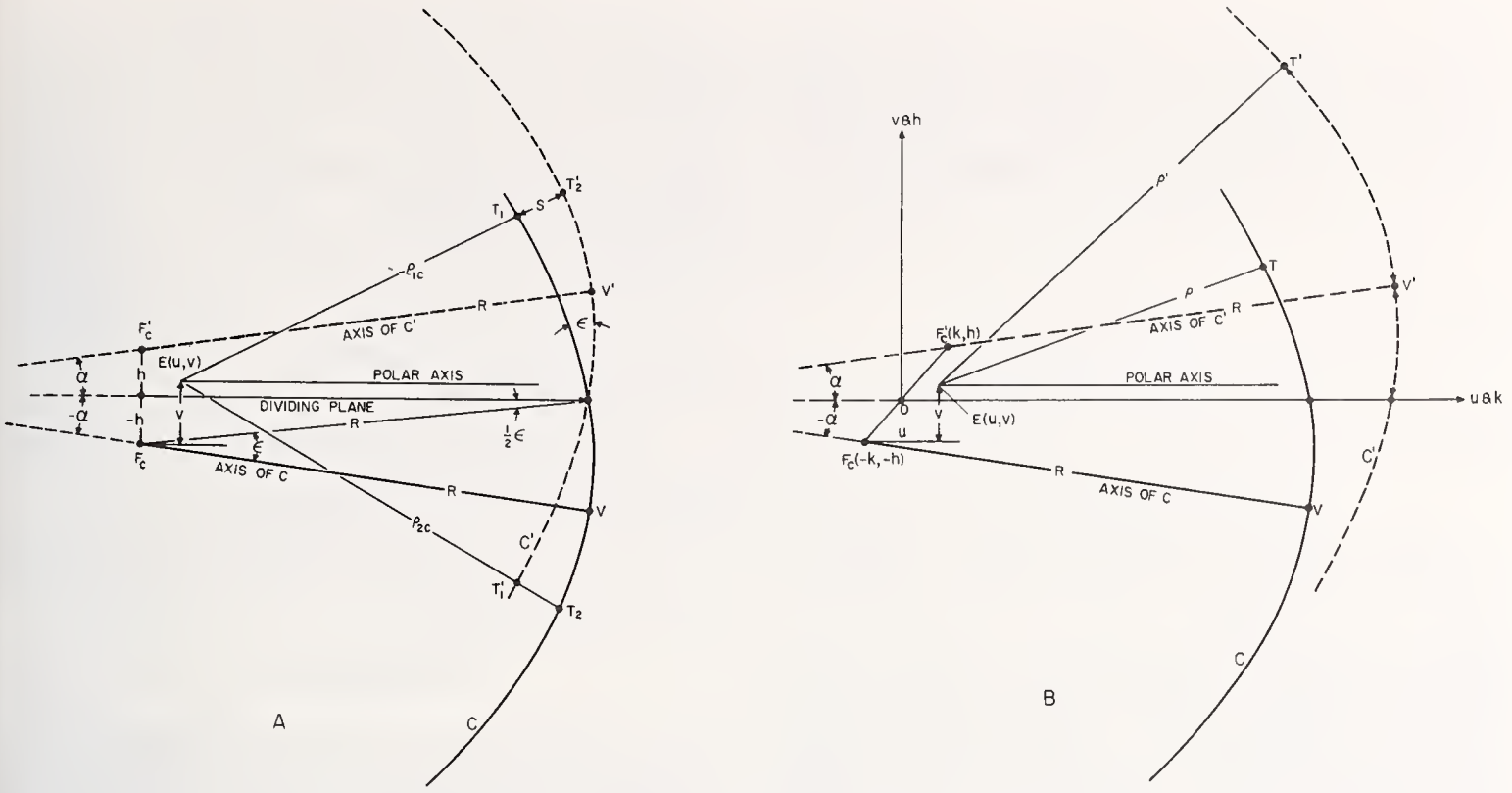


FIGURE 13. A and B. *Illustration of the relative positions of two images of a wave front with the WRI and WSI, respectively.*

The term $f_1(\alpha, \beta)$ is the value of ρ when the pole is located at F_c. When testing optics of large aperture-to-focal-length ratios the quantity $f_1(\alpha, \beta)$ must include higher order terms of the series than for smaller f-values. It will be found that, in some cases, the $f_1(\alpha, \beta)$-series converges too slowly for practical application and that the rigorous, unexpanded value

$$f_1(\alpha, \beta) \equiv \frac{R[\sqrt{1+e^2\sin^2(\alpha+\beta)}-e^2\cos(\alpha+\beta)]}{[1-e^2\cos^2(\alpha+\beta)]} \tag{27}$$

is more practical. For this treatment we shall limit the angles to values comparable to those obtained with an $f/8$ system. This limitation makes the last terms, indicated in each of the functions $f_1(\alpha, \beta)$ $f_2(\alpha, \beta)$ and $f_3(\alpha, \beta)$ of eq (26) negligible so that eq (26) becomes.

$$\rho = R + 1/8Re^2(\alpha+\beta)^4 - u\cos\beta - v\sin\beta. \tag{28}$$

3.1. Equations for the WRI

When the conic is used as the reference surface in a wave-front-reversing interferometer, the optical path difference, S (fig. 13A), is the difference between two radius vectors, ρ_{1c} and ρ_{2c}, forming angles with the dividing plane, which are equal in magnitude but opposite in sign.

The two curves, C and C', shown in figure 13A are images of the conic section. These two images are inverted by the interferometer, with respect to each other, about the dividing plane which is parallel to the polar axis. The two points T_1 and T_1' are images of a common point on the original curve. Similarly T_2 and T_2' are images of a common point; but T_1 and T_2 are images of different points that are at equal angular distances from the dividing plane. The radii vectors to T_1 and T_2 are ρ_{1c} and ρ_{2c}, respectively.

If we define the perpendicular displacement of F_c from the dividing plane as $-h$, the corresponding displacement for F_c' is $+h$. The component displacements of E from F_c, as defined for figure 12, is u and v. The corresponding displacements of E from F_c' is u and $(v-2h)$. The value of ρ_{1c} is, therefore, obtained by replacing ρ in eq (28) by ρ_{1c}. Similarly, since the sign α is reversed in the image, the value for ρ_{2c} is obtained by replacing ρ, α, and v by ρ_{2c}, $-\alpha$, and $(v-2h)$, respectively. On making these substitutions, we obtain

$$\rho_{2c} - \rho_{1c} \equiv S = 2h\sin\beta - Re^2\alpha\beta(\alpha^2+\beta^2). \tag{29}$$

It is convenient to use the measured distances $(\nu-\mu)$, and μ instead of the angles β and α. Also, from figure 13A, it is obvious that $2h=\epsilon R$, since h is quite small relative to R. On substituting for $2h$, α, β (=sin β, approximately), their respective equals, ϵR, μR^{-1}, and $(\nu-\mu)R^{-1}$, the value of S becomes

$$S=\epsilon(\nu-\mu)-e^2R^{-3}\mu(\nu-\mu)[\mu^2+(\nu-\mu)^2] \quad (30)$$

which becomes for a paraboloid $(e=1)$,

$$S=\epsilon(\nu-\mu)-R^{-3}\mu(\nu-\mu)[\mu^2+(\nu-\mu)^2] \quad (31)$$

and for a sphere $(e=0)$,

$$S=\epsilon(\nu-\mu). \quad (32)$$

Equation (32) is the same as eq (1). The use of a conic section instead of a circle for the reference curve, in the derivation of eq (3), leads to the introduction of the additional unknown, e, into eq (3) so that an additional condition equation becomes necessary for a unique solution, thus permitting the incidental evaluation of e. Since four conditions are required to define the chosen conicoid, four condition equations are needed. These may be: The assignment of arbitrary values (usually zeros) to any four of the δ's (if the method of coincidence is used); the grouping of the reference points into four groups and requiring the algebraic sum of the δ's in each group to be zero (if the method of averages is used); or, the addition of the condition equation, $\frac{\partial\varphi}{\partial e}=0$, to eq (12) (if the method of least squares is used).

3.2. Application to the WSI

When the conic is used as a reference surface in a wave-front-shearing interferometer, the optical path difference, S, is $(\rho'-\rho)$, where ρ and ρ' are the radius vectors from the point E to points T and T' (see fig. 13B) on curves C and C', respectively. Points T and T' are images of a common point on the conic section. Curves C and C' are sheared images of the conic section. The angle of shear is $-\alpha$ for C and α for C'. Therefore, the polar axis makes angles α and $-\alpha$ with the axis of C and C', respectively.

The component displacements of E from F_c is the same in figure 13B as in figure 12B. The value for ρ is, therefore, given by eq (28). The component displacements of E from F' is $(u-2k)$ and $(v-2h)$. Consequently, the value for ρ' is obtained from eq (28) by replacing ρ, α, u, and v by ρ', $-\alpha$, $(u-2k)$ and $(v-2h)$, respectively. On making these substitutions and taking differences we obtain

$$\rho'-\rho\equiv S=2h\sin\beta+2k\cos\beta-Re^2\alpha\beta(\alpha^2+\beta^2). \quad (33)$$

It will be noted that k does not appear in eq (29) (relating to WRI) but does appear in (33) (relating to WSI). The term $2k\cos\beta$ is approximately constant when β is limited to small values (as when testing an $f/8$ system) but becomes significant for large values of β (as when testing an $f/1$ system), unless k is sufficiently small. If the WSI is adjusted to place the zero order of interference at the center, the magnitude of k is insignificant and the values of S for the two interferometers are identical. On substituting for $2h$, α, and β their respective equals, as given in section 3.1, above ($\cos\beta=1$, approximately), eq (33) becomes

$$S=\epsilon(\nu-\mu)+2k-e^2R^{-3}\mu(\nu-\mu)[\mu^2+(\nu-\mu)^2] \quad (34)$$

which becomes for a paraboloid $(e=1)$,

$$S=\epsilon(\nu-\mu)+2k-R^{-3}\mu(\nu-\mu)[\mu^2+(\nu-\mu)^2] \quad (35)$$

and for a sphere $(e=0)$,

$$S=\epsilon(\nu-\mu)+2k. \quad (36)$$

It is inconvenient and unnecessary to reduce k to an insignificant magnitude. It can be combined with other constants, as described in reference [2]. In fact, the constant k is a part of the constant r of reference [2]. Unfortunately, the sign convention for measuring μ and ν is different in this paper (part 2) from that of part 1 [2]. The constant, r, computed from eq (3) of reference [2] in combination with eq (36) above, becomes $r=Q_r+\epsilon\mu-2k$. Since r is a constant and serves as a parameter only, its form does not affect the problem.

4. References

[1] J. B. Saunders, Construction of a Koesters double-image prism, J. Research NBS **58,** 21 (1957) RP2729.

[2] J. B. Saunders, Measurement of wave fronts without a reference standard: Part I. The wave-front-shearing interferometer; J. Research NBS **65B,** 239 (October–December 1961).

[3] Pol Mollet, Optics in metrology, the wavefront reversing interferometer, p. 227 (The Pergamon Press, 1960).

[4] J. B. Scarborough, Numerical mathematical analysis, p. 357 (The Johns Hopkins Press, Baltimore, Md., 1930).

[5] W. A. Granville, Elements of the differential and integral calculus, p. 228; Ginn and Co. (or any other textbook on differential and integral calculus).

[6] J. B. Saunders, Parallel testing interferometer, J. Research NBS **61,** 491 (1957) RP2917.

(Paper 66Bl–70)

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 68C, No. 3, July–September 1964

Wave Front Shearing Prism Interferometer

J. B. Saunders

(April 9, 1964)

A small prism interferometer is described that is a modification of the Bates' wave front shearing interferometer. The prism is small, rugged, and versatile. The shape of any converging wave front from any optical element or compound instrument may be measured with this instrument.

Several variations of the prism are described that have characteristic properties. Special features are introduced for eliminating spherical aberration effects, for producing fringes of different characteristics, and for testing all wave fronts up to and exceeding that of an F.1 cone of rays.

Details for making and adjusting the prisms are given. Examples are given for interpreting the interferograms.

1. Introduction

The principles of the wave front shearing interferometer (WSI) have been described by Professor J. W. Bates [1]. This interferometer is particularly well suited for measuring the shapes of converging wave fronts. The difficulty of making the Bates instrument and others that have subsequently been described has limited the applications. This paper describes a simple form of WSI that is easily constructed, easily adjusted, and is practical. It is very small and can be mounted in an eyepiece cell for testing astronomical telescopes or as a convenient mount for other applications.

The WSI described by Bates consists of two beam dividers, two front-face mirrors, and two compensator plates. The angle of shear could be varied at will in the Bates model. Later models of the WSI by Drew [2] and Brown [3] are more compact. The Brown version is a relatively compact instrument of fixed shear. The instrument described here is more compact than any of the previous models, is quite easy to adjust, and is relatively free from vibration effects.

The WSI is quite suitable for testing lenses, mirrors, and combinations of these. It is also quite practical for studying atmospheric effects on light. Professor Bates described a method for analyzing the results from a WSI when the converging wave front is axially symmetrical. D. S. Brown [4] described another method of interpretation of sheared interferograms. The author of this paper has given a rigorous mathematical operation [5] for interpretation of sheared interferograms that does not depend upon symmetry of wave form. The solution is unique for any chosen "family" of reference points.

The elements forming this wave front shearing interferometer have been combined into a small compound prism that may be made into a variety of forms. Each form may be constructed with or without a built-in wedge that either permits or prevents rotation of the fringes. The two components of the prism are cemented together for stabilization of those adjustments that should remain fixed, without limiting the adjustments that are essential to its operation. The unit is small, rugged, compact, easily constructed, easily adjusted, and simple to use.

2. Basic Optical Principles

The wave front shearing interferometer (WSI) is a modification of that developed by Mach and Zehnder [6]. It differs from the MZ-interferometer in that convergent light is used instead of collimated light, the reflecting and beam-dividing surfaces are not parallel, the component beams emerge nonparallel, and the fringes are nonlocalized. The angle between the emerging components of any ray represents the angle of shear. To get observable interference between these two beams a source that is small [1] in the direction of shear is necessary.

The differential deviation of the component beams is accomplished in the Bates WSI by tilting one of the beam dividers, whereas in the Drew WSI it is effected by rotating one of the front-face mirrors. The vertex, or point of convergence, of one of the component wave fronts (images of the source) is always located near the optical element that is used to effect the differential deviation (angle of shear). The lateral separation of the two images of the source determines the width of the fringes (as in the Fresnel's mirrors [7] arrangement.)

In figure 1A, the lens, L, forms two images of the light source, S, at S_1 and S_2. The two component beams (represented by the two components of the principal ray and designated as 1 and 2) suffer total internal reflection at points P_1 and P_2. Each of these beams again divides at plane B into two components—the transmitted beam, 1, dividing into beams 1′ and 1″ and the reflected beams, 2, dividing into beams 2′ and 2″. An observer's eye located close to the prism so as to receive either pair of rays 1′ and 2″ or 1″ and 2′ will be able to see two images of L (L_1 and L_2, figure 1B).

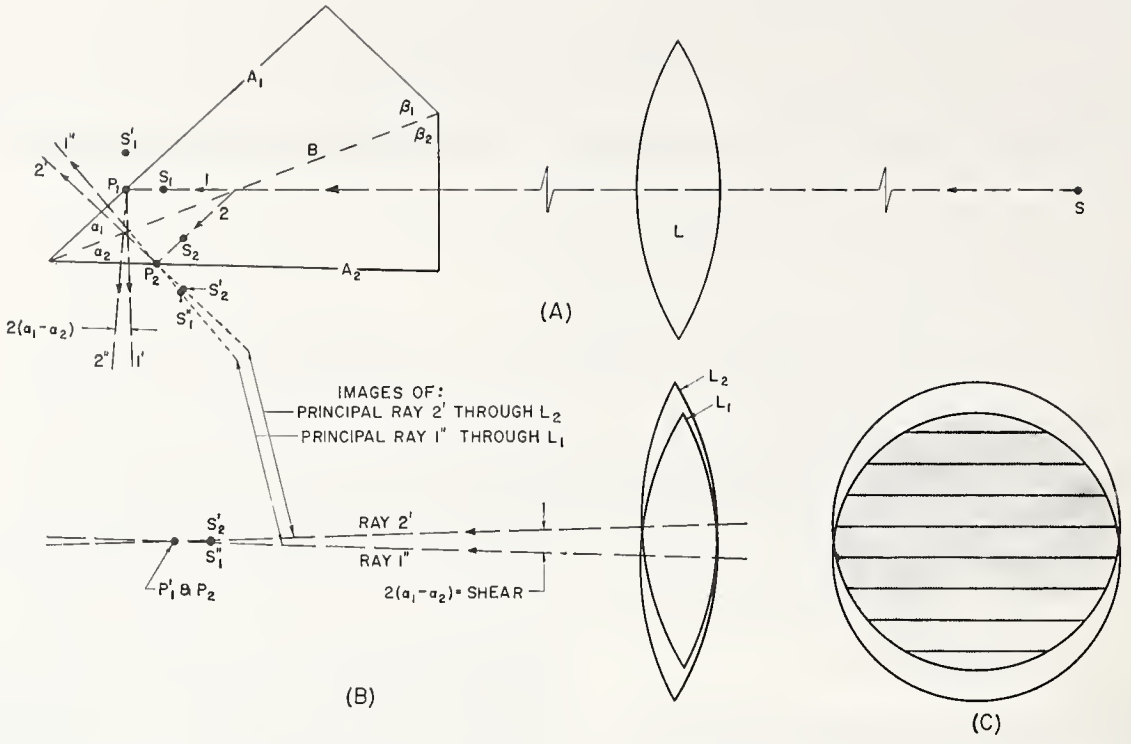


FIGURE 1A. *Optical arrangement for testing a lens, L. The two elements of the prism, A_1B and A_2B, are cemented together, with a semireflecting coat of aluminum on one of the cemented surfaces. Refraction at the two exit faces is assumed to be negligible.*
B. *L_1 and L_2 are images of L, as seen through prisms A_1B and A_2B, respectively, by an observer's eye located behind the prism and receiving rays 1'' and 2'. The image P_1' of P_1 appears to coincide with P_2.*
C. *Fringes of interference that appear to be localized in the overlapping areas of the two images, L_1 and L_2.*

The two component prisms, shown in figure 1A, have angles α_1 and α_2 that differ by a chosen amount. Prism A_1B is assumed to have been moved parallel to the plane of figure 1 and relative to prism A_2B so that the image of $P_1[P_2]$, with respect to the semireflecting plane B, coincides with $P_2[P_1]$. This places P_1 and P_2 equal distances from both the source and the observer. The two emerging beams, $1'$ and $2''$, appear to intersect at P_1 and the other pair of beams, $1''$ and $2'$, appears to intersect at P_2. Images of S_1 and S_2 appear to be located at S_1'' and S_2', respectively. This adjustment of the prisms makes the two paths of light through the center of the lens equal so that when white light is used the zero order of interference is observed to pass midway between the centers of the two images of the lens. This fringe is almost colorless with a relatively pure colored fringe on each side of it. All other fringes are colored with varying tints due to overlapping of colors.

If the points P_1 and P_2 are further from L than S_1 and S_2, images of S are formed between the lens and these points. If, however, the points P_1 and P_2 are nearer the lens than S_1 and S_2, the images of S are beyond the points P_1 and P_2. By moving either the prism, the lens, or the source along the axis of the lens, the emerging components of the principal rays remain fixed relative to the prism and the two component images of the source, S_1'' and S_2' (see figure 1B), appear to move along the two components of the principal ray. If the two images are made to coincide with P_1' and P_2 they will also coincide with each other and the two components, $1''$ and $2'$ (or $1'$ and $2''$), of the principal ray intersect at P_2 (or P_1'). Consequently the lateral separation of the two coherent images, S_1''' and S_2', (figure 2A) may be varied at will by a simple linear translation of the prism along the axis of the lens.

Since the two images, S_1 and S_2, are coherent, of equal intensity, and equally distant from both the observer and the source, S, the two emerging cones of light will combine to produce low-order interference. The two coherent images act exactly as the two coherent sources in the Fresnel's mirrors [7] arrangement. The fringes of interference are nonlocalized. If either of the two pairs of beams, $1'$ and $2''$ or $1''$ and $2'$, is projected onto a photographic emulsion (on film or paper) a photograph of the fringes is obtained. Also using a camera to receive either pair of beams, the fringes may be photographed in good contrast. The focusing of the camera does not vary the contrast, but if the lens, L, is imperfect the nature of the interferogram (fringe pattern) will vary with changes in the focus of the camera because the shape of the wave front changes as it is propagated from one position to another.

Interference fringes appear only in the overlapping areas (fig. 1C) of the two component beams. These fringes will be straight if the lens, L, is free from axial aberrations. Figure 2A shows several sets of fringes that illustrate the effect of moving the prism along the axis of the lens. As in the Fresnel's mirrors arrangement, the fringes are always perpendic-

ular to the direction of the straight line joining S_2' and S_1'' and their width (or separation) is inversely proportional to the lateral distance between the two images S_2' and S_1''.

Since the two rays, $1''$ and $2'$, intersect (lying in a common plane) and since the two images, S_2' and S_1'', are equally distant from the point of intersection, the relative direction of the two images does not change. Consequently, the fringes do not rotate with movement of the prism along the principal ray.

If the compound prism (fig. 1A) is moved perpendicular to the axis of the lens and in the plane of figure 1A, the distance from the two images, S_2' and S_1'' (fig. 2B), to the observer becomes unequal and the zero order of interference moves away from the center of the field. The movements of the fringes, due to prism movements along the axis, will be similar to those in figure 2A and the zero order of interference will be displaced from the center of the field. The zero-order fringe is indicated in figure 2 by representing it with lines that are heavier than for other orders.

The analysis of wave fronts by the method described in reference [5] requires that a "family of reference points" (appendix A) be chosen along a line parallel to the direction of shear.

If prisms A_1B and A_2B (fig. 3) are each rotated a few degrees about an axis normal to B and through points P_1 and P_2, rays 1 and 2 will be deflected out of the plane of figure 3A. If the rotation is equal for the two prism components but in opposite directions the two component rays will be deflected by equal amounts. This results in a shearing of the wave front. The direction of shear is perpendicular to the plane of figure 3A, whereas the shear indicated by figure 1B was in the plane of figure 1A. The two emerging rays $1'$ and $2''$ (or $1''$ and $2'$) lie in a plane that is normal to that of figure 3A. Figure 3B shows the observed images of the lens as seen by beams $1''$ and $2'$, in a plane normal to that of figure 3A. The plane of figure 3B is also normal to that of figure 1B.

The two prisms shown in figures 1 and 3 shear the wave front in directions that are perpendicular and parallel, respectively, to the vertices of the prism angles. The spacing of the fringes in both prisms varies when the prism is moved along the principal ray. The fringes in both prisms move laterally, without change of spacing, when the prism is moved in the direction of shear and do not change for small movements of the prism perpendicular both to the principal ray and to the direction of shear.

If the prisms in figures 1 and 3 have been adjusted to place the images of P_1 and P_2 coincident, the two sources will always be equally distant from the intersection of the components of the principal rays. Consequently, the direction of the line joining the two image sources will not change and the interference fringes will not rotate. If the fringes are to be rotatable the components of the principal ray must not intersect each other.

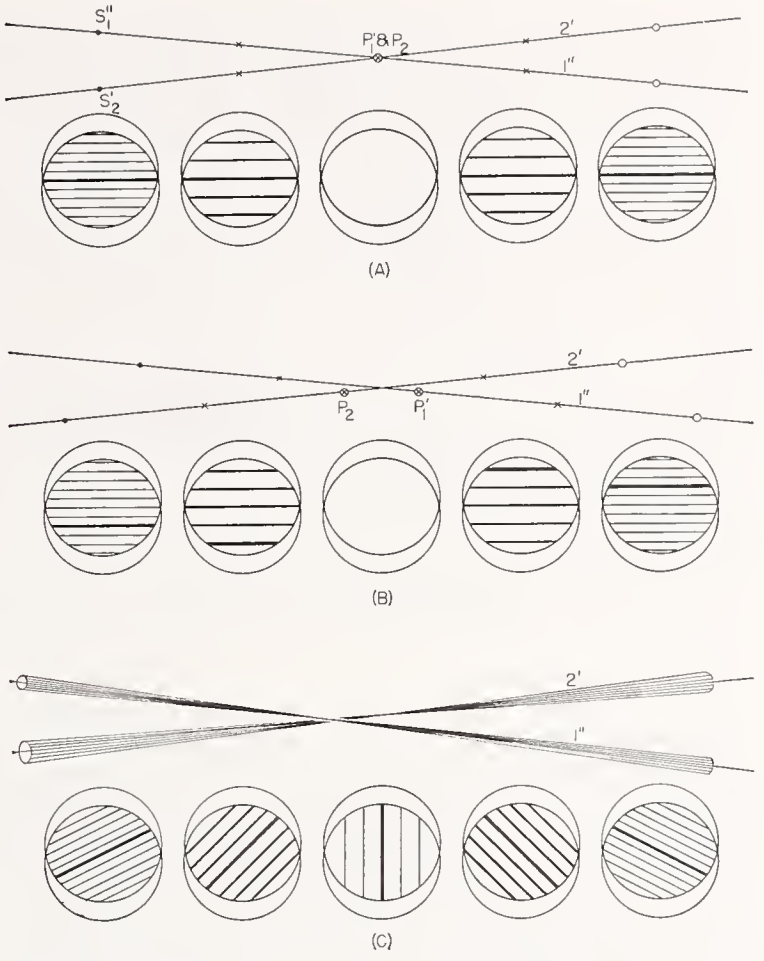


FIGURE 2A. *The components of the principal ray ($1''$ and $2'$ of fig. 1 B) intersect at coincident points P_1' and P_2. The fringe spacing varies with separation of images S_2' and S_1'' but do not rotate.*

B. *Components of principal ray intersect, but not at P_1' or P_2. P_1' and P_2 are equally distant from the point of intersection. The zero-order fringe is displaced from the center of the pattern.*

C. *The two component rays, $1''$ and $2'$ do not intersect. The horizontal component of the fringe width is invariant and the vertical component varies as in the above figures, A and B. This results in a rotation of the fringes. The fringe width is a maximum when the separation of the two images of the source are closest together.*

It is desirable, in many cases, to use white light sources for testing lenses—particularly for testing astronomical telescopes with light from celestial stars. When white light is used for the source the number of usable fringes is limited. In this case it is desirable that the fringes be as nearly parallel to the line of reference points as possible and with the zero-order fringe (the fringe of best visibility) located on or among the reference points. The prism described above and illustrated in figure 1 does not permit this adjustment without making the fringes too broad for accurate readings, unless the wave front has considerable distortion.

The WSI prism may be designed to permit rotation of the fringes. In the preferred design for rotatable fringes, the shearing of the component beams is accomplished by using two identical prism compo-

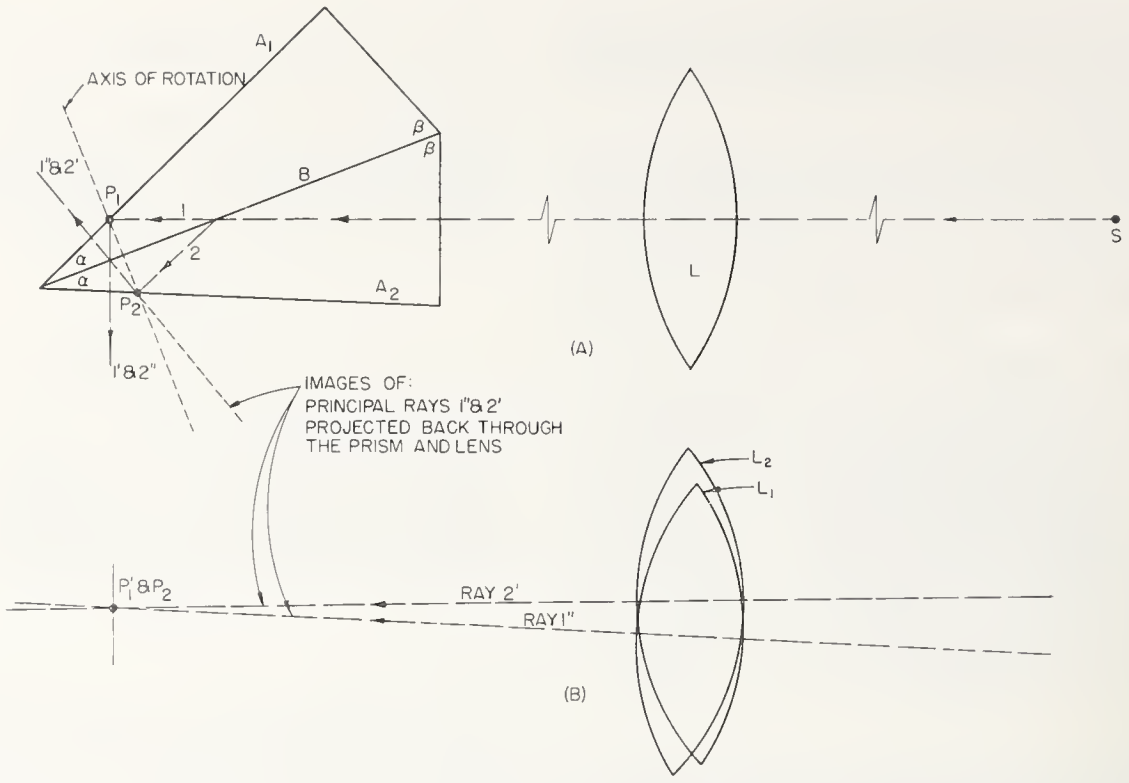


FIGURE 3A. *The two components of the prism are identical, with the aluminum coat on one prism only. Shear is obtained by rotating the two components about an axis normal to B, but in opposite directions.*

B. *The planes of figures A and B are normal to each other.*

nents with a relative rotation of these components about an axis that is normal to the dividing plane, B. If the angles of the two component prisms are equal and no shear is introduced, the two component rays 1′ and 2″ or 1″ and 2′ (figure 3) will coincide. The two images of L (shown in fig. 3B as L_1 and L_2) will coincide. The optical path difference between rays 1 and 2, and all other rays in the beam, will be zero. No fringes will be observed because the phase difference is constant over the entire beam.

To obtain rotation of the fringes the two component rays 1″ and 2′ (and also 1′ and 2″) are shifted, relative to each other, in the direction normal to the plane of figure 3B. Figure 4 shows the ray trace through a prism that has been adjusted to perform this shift. Dividing planes B_1 and B_2 are parallel to each other and separated by a distance, t, that is proportional to the resultant lateral separation, d, of the two component emerging beams.

The prism of figure 4 is adjusted to produce zero optical path difference for the two component principal rays. Shear is obtained by a relative rotation of the prisms, exactly as described for the prism of figure 3.

The two components, 1 and 2, of the principal ray will emerge in planes that are perpendicular to the plane of figure 4 and separated by the distance d. The component rays will make equal angles, but of opposite sign, with the plane of figure 4. When the prism is moved along the axis of the lens the two images of the source move along the two rays 1 and 2, remaining equally distant from both the source and the observer. The zero-order fringe remains at the center of the field. The separation and relative direction of the two images of the source change with this movement of prism. Consequently, the fringes rotate and change width during this movement.

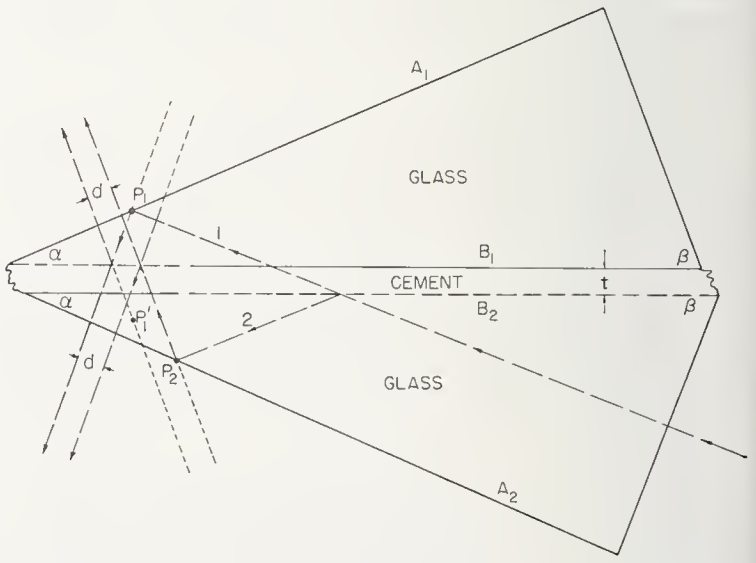


FIGURE 4. *The two components of the prism are identical.*

The reflecting film on P_2 extends from the vertex of β to the normal projection of P_2 onto P. The reflecting film on P_1 extends from the vertex of α to the normal projection of P_2 on P_1.

The uncoated parts of B_1 and B_2 virtually disappear if the index of the glass equals that of the cement.

Figure 2C indicates the two pencils of rays about the component rays 1 and 2. The direction and width of fringes for several positions of the prism, relative to the lens, are shown.

When the fringes are alined in the direction of shear they are of maximum width. Since the reference points must be on a line parallel to the direction of shear, the zero-order fringe can be alined with the line of reference points (assuming a good lens is being tested) and the zero-order fringe can be shifted laterally for most favorable reading of orders.

3. Symbolic Designation of Prism Types

The original prism interferometer was found to be a special case of a large variety of forms. Also, each form may be constructed and adjusted to provide a variety of properties, angles of shear, and other special adaptations. A description of one or two of these prisms would not adequately cover the development.

There are two geometrical forms, a quadrilateral and a hexagonal prism (see fig. 5). The usable angles of each of these two forms cover a considerable but finite range. Either form of prism with any combination of usable angles may be adjusted to have any angle of shear over a considerable range. There are two distinctive ways of producing shears which are normal to each other relative to the prism. Each of these types of shear is associated with adjustments of the interference fringes that have advantages and disadvantages for various types of measurements.

In order to describe these prisms adequately we will describe their common properties, method of adjustment, construction, etc., as a group. To do this an abbreviation will be used for designation of groups, and the abbreviations are so chosen that for future use definite specifications may be indicated for all types and forms. The chosen abbreviation is a compound letter-number symbol that indicates several of the prism's significant characteristics. The generalized symbol is *NEWS* (N=Number, E=Entrance face, W=Width of fringes, and S=Shear).

The first symbol, N, represents the number of sides or faces of the compound prism. When the numerical value of N is significant its value will appear as a subscript (N_4 for quadrilateral and N_6 for hexagonal prisms); otherwise N appears without the subscript.

The second symbol, E, indicates the form of the entrance face of the prism. When E appears without a subscript, the shape of the entrance face is not significant to the statement in which it appears. When the shape is significant it will be indicated by subscripts (E_p for plane and E_s for spherical entrance face).

The third symbol, W, represents the maximum width fringes that can be obtained, with a perfect wave front, by moving the prism along the lens axis. The value of W is finite if the direction of shear is parallel to the refracting edges of the prism (fig. 3) and the beam-dividing area is divided between the two component prisms (fig. 4). The value of W is infinite if the direction of shear is perpendicular to the refracting edges of the prism, regardless of whether the beam-dividing areas are confined to one prism or divided between the components.

The lateral separation of the components of the principal ray, obtained by separating the planes of the two beam-dividing areas, corresponds to the introduction of a fixed wedge between the two component wave fronts. Similarily, the wedge between the wave fronts that is caused by a separation of the two coherent images of the source in the direction of shear may be called the variable wedge. When the directions of these wedges are the same they cannot be distinguished and the fringes act exactly as they do in a prism without a fixed wedge, When the direction of the fixed wedge is perpendicular to that of the variable wedge, the widths of the fringes in these two directions are proportional to the magnitude of the two wedges, respectively. The width in one direction (horizontally in fig. 2) is invariant and the width in the other direction (vertically in fig. 2) varies with the position of the prism relative to the lens (or point of convergence).

A prism with a built-in wedge that is perpendicular to the direction of shear will be designated by W_s (s stands for German *senkrecht*, meaning "perpendicular"). If the numerical value of W is to be indicated also it will appear as a subscript of W and expressed in units of milliradians. Thus a prism with a built-in wedge of 2.1 mrad will be indicated by $NEW_{2.1s}S$.

A prism with a built-in wedge that is parallel to the direction of shear will be designated by W_p (p for "parallel"). Similarily, a prism with no built-in wedge (aluminum on one surface only) will be designated by W_∞ because the magnitude of W is given by this subscript.

It will be noted that the numerical value of W in a NEW_pS-type prism is also infinity. However, it is desirable to discriminate between the three types and these symbols seem to be most appropriate.

The fourth symbol indicates the magnitude and direction of the shear. If the value and direction are insignificant to a statement in which it appears, the symbol S will suffice. If shear is produced by a relative rotation of the prism component, to be described later, its direction will be parallel to the refracting edges of the prism and the corresponding symbol is S_p. If shear is produced by using prisms of different angles, the direction of shear will be perpendicular to the refracting edges of the prism and the corresponding symbol is S_s. When the numerical value of the shear is to be indicated it will appear in units of milliradians as a part of the subscript of S. Thus a shear of 2.1 mrad in a prism having components of different angles will be indicated by $S_{2.1s}$. Similarily, a prism having a shear of 2.1 mrad that is introduced by a relative rotation of similar prism components will be designated by $S_{2.1p}$.

4. Geometry of the Prism Interferometer

The angles and size of the prisms for the shearing interferometer are not critical. The choice of size and angles of prisms that can be used to produce a workable instrument is so diverse that only optimum forms will be considered. These forms are chosen because they fulfill significant favorable operational performances that are unique for the different forms.

4.1. Size of the Prism

An optimum size prism (length of face B in fig. 1) has a dividing plane of 1.5 in. A 1.0 in. or a 1.7 in. dividing plane would be almost equally satisfactory. Larger prisms require better glass and introduce larger aberrations because of thickness, unless a spherical entrance face is used. Small prisms are more difficult to polish and adjust than larger ones.

4.2. Surfaces

All reflecting surfaces of the prism interferometer are planes. The flatness of the beam-dividing plane is critical. The outer reflecting surfaces are less critical because the cones of light have small cross sections (being near the vertex of the cone) at the points of reflection. The entrance and exit faces may be planes if the angle of the cone of light does not exceed 1/8 rad (as when testing an $f/8$ lens with one conjugate relatively large). When larger cones are used, the spherical aberration of the prism introduces appreciable deformation of the wave front and consequently error into the results of measurement. This error is introduced at the entrance face because of the difference in refraction of any two rays that are combined to produce interference. Deformation of the wave front by the entrance face is practically eliminated if a planoconvex lens (figs. 5E and 5J) having the center of curvature of its spherical surface located near the point of internal reflection, P_1, is cemented to the entrance face.

Since any two rays that are combined to produce interference emerge parallel and nearly coincident, the shape of the exit face has little effect on the order of interference observed at any point in the aperture of the lens.

Distortion by a plane exit face does not contribute error to the measurements if they are made visually or photographically, provided reference points are established in the aperature of the objective being tested. If photography is used without previously established reference points (reference points inserted in the photograph), the distortion by the exit face will be appreciable for large cones of light but negligible for small cones.

4.3. Angles of the Prism

The angles of the prism are not critical. Since the magnitude of the angles does not affect the difficulty of construction, but does affect the performance, the choice of angles will be based on certain optimum constraints or requirements of performance. Thus, the angles will be so chosen that when the principal ray of light from a circular lens (circular cone of rays) enters the plane entrance face normally and at the center of this face (thus permitting a maximum size cone of rays to be accepted) the components of this principal ray will emerge from one exit face of the prism with minimum refraction. This condition provides minimum refraction for the entire cone of light and, consequently, minimum error in the results of measurement caused by aberration in the prism.

The prisms shown in figure 5 represent the several forms that are recommended. Other forms are possible but these are adequate and sufficient for all possible applications. For purposes of discussion the characteristic angles of the several prisms are represented as follows: the N_6EWS_p-type prisms by α', β', and ω' (fig. 5A); the N_4EWS_p-type prisms by α' and β' (fig. 5F); the N_6EWS_s-type prisms by α_1, α_2, β, and ω (fig. 5B, the angles β and ω may or may not be equal for the two component prisms); the N_4EWS_s-type prisms by α_1, α_2, and β (fig. 5G). The angles shown here for all $NEWS_p$-type prisms represent the components of the prism angles in the plane of this drawing. Since the elements of these prisms are rotated (about an axis lying in this plane) to produce shear the refracting edges are no longer perpendicular to the figure. If the angle of rotation of any two prism components are $+\Omega$ and $-\Omega$, respectively, all refracting edges are at angles of $(90°\pm\Omega)$ to the plane of figure 5. The angles of the prisms shown here are α, β, and ω. The corresponding components of the prism angles in the plane of figure 5 are α', β', and ω'. The relationship between any prism angle (unprimed) and its component in the plane of this drawing (primed) is

$$\text{Tan } X=\cos\ \Omega \tan X',\ (X=\alpha,\ \beta,\ \text{or}\ \omega). \qquad (1)$$

The deviation of the components of the principal ray out of the plane of figure 3, at the points of internal reflection, P_1 and P_2, depends upon the angles α and Ω. Since the deviations are equal and of opposite signs, and since the angle between the two emerging components from either exit face is the angle of shear, ϕ, the approximate relationship between ϕ, Ω, and α is given by

$$\tan 1/2\ \phi=n \sin\ \Omega \tan\ \alpha, \qquad (2)$$

where n is the refractive index of the glass.

The angle of shear in the $NEWS_s$ prism is, approximately,

$$\phi=2n(\alpha_1-\alpha_2). \qquad (3)$$

When the angle α, of the N_6EWS_p prism is small and the components are rotated π radians the prism becomes a N_6EWS_s type. Such a prism, shown in

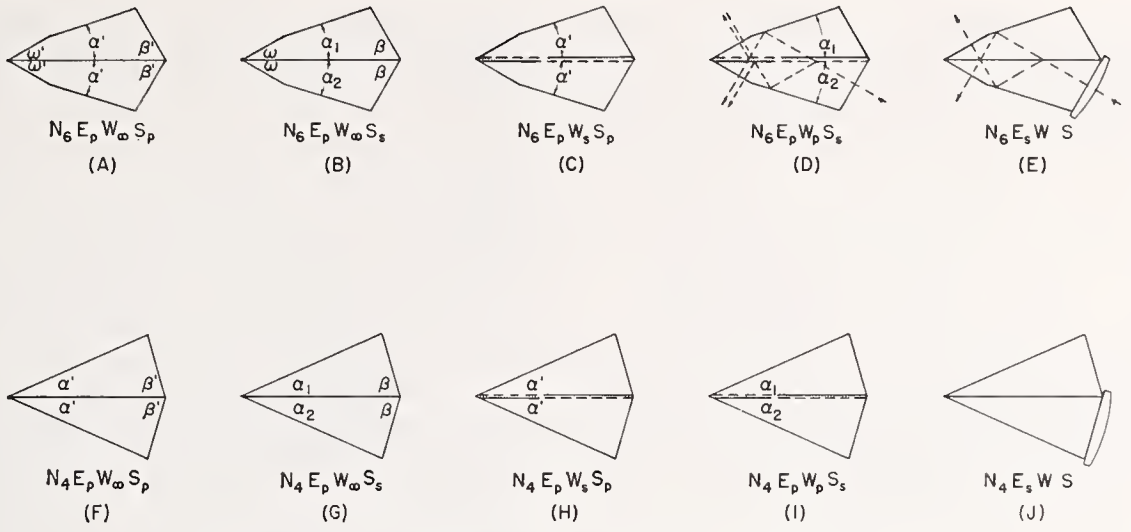


FIGURE 5. *Several different types of prism interferometers.*
The letter-number abbreviations, described in section 3, describe these prisms.

figure 6A can be made with adequate shear and to accept an $f/1$ cone of light. However, for testing such small f-value systems, spherical faces (fig. 6B) should be used on the prism to avoid spherical aberrations. When the angle of rotation, Ω, equals π radians, the angle, α, of one prism is reversed and the shear, ϕ, equals $4n\alpha$. Consequently, the angle, α, is usually small for such prisms.

Minimum refraction is obtained in plane face prisms when the principal ray of the entrance cone is normal to the plane entrance face and the two components of this principal ray, that emerge from each of the exit faces, make equal angles, but of opposite sign, with the exit face. If the prism is of the $NEWS_p$ type, the angles of emergence of the two pairs of beams are equal. However, the angles of emergence for the $NEWS_s$-type prisms, because $\alpha_1 \neq \alpha_2$, cannot be a minimum for both pairs simultaneously. Consequently the pair of beams that emerges from face A_2 (fig. 1) will be ignored when considering the constraints of angles. When rays, $1''$ and $2'$, (fig. 1A), that emerge from face A_1 suffer minimum refraction, their angles of emergence are $+(1/2)\phi$ and $-(1/2)\phi$, respectively, and the rays $1'$ and $2''$, that emerge from face A_2 will be $0°$ and ϕ, respectively. The reason for minimizing refraction of the beams that emerge from face A_1, instead of those from face A_2, is due to their greater equality

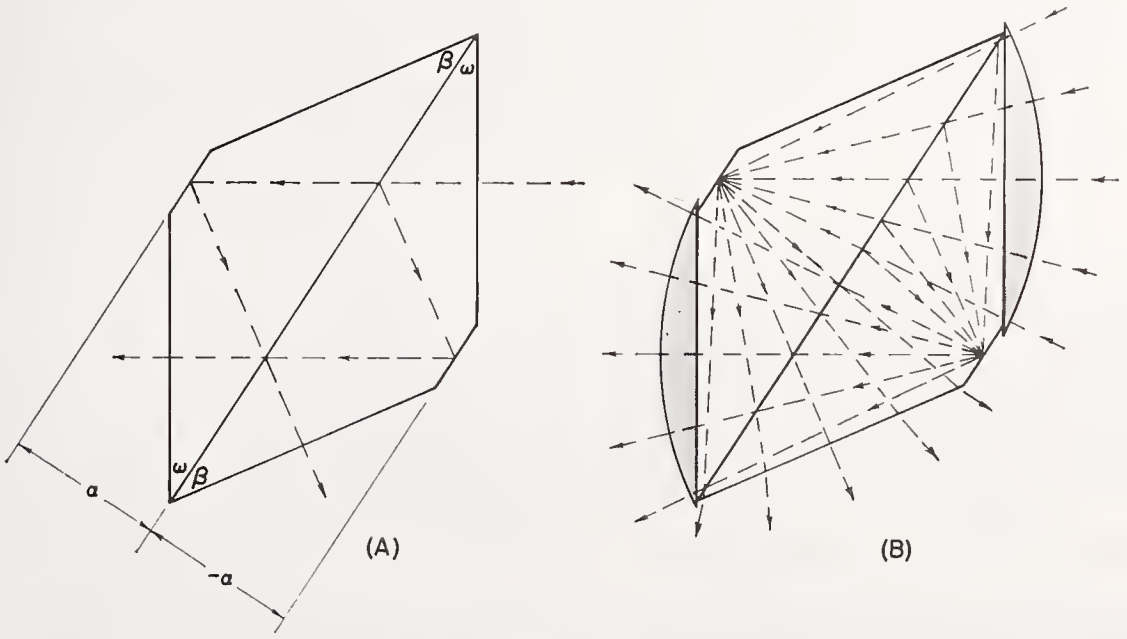


FIGURE 6A. *Prism interferometer for testing very large cones of light.*
B. *Spherical entrance and exit faces eliminate error due to refraction of the prism surfaces.*

of intensity and, consequently, better visibility of fringes. This pair of beams will usually be used for making measurements.

It can be shown by simple triangulation that minimum refraction is obtained in plane face prisms if

$$\left.\begin{aligned} &\beta=3\alpha' && \text{for all } N_4E_pWS_p\text{-type prisms,}\\ &\beta=2\alpha'+\omega' && \text{for all } N_6E_pWS_p\text{-type prisms,}\\ &\beta=2\alpha_1+\alpha_2 && \text{for all } NE_pWS_s\text{-type prisms.}\end{aligned}\right\} \quad (4)$$

We will, therefore, impose these constraints on the design of all prisms described here. Since the face of prism A_1B (fig. 1), that corresponds to the entrance face of prism A_2B, is not used it may be left unpolished. If, however, one wishes to be able to use either prism for receiving the light this surface should be polished and the angle β_1 should be equal to β_2.

The angles may have a variety of values consistent with eq (4). The range of these values depends upon factors such as the magnitude of the angle of shear (defined by ϕ), the form of prism to be used, the refractive index of the glass, the maximum size cone of light that is to be used, and whether or not the outer reflecting surfaces are to be coated for increased reflection.

The range of usable values for α' in the N_6EWS_p prism is from 0° to ω'. Obviously, as α' approaches ω' the N_6EWS_p prism becomes N_4EWS_p in form. The upper limit for α' in the N_4EWS_p prism is $\pi/7$ rad (=25.7°, approximately). Figure 7 illustrates the limitations of α' for this type of prism. If α' ($=\omega'$) is greater than $\pi/7$ rad (assuming the principal ray enters the center of the entrance face and normal to it), the beam will not suffer two divisions at the beam dividing plane, as is shown in figure 7A.

The lower limit of $\omega'(=\alpha')$ for the N_4EWS_p prism depends upon whether the outer reflecting faces are coated to prevent transmission of light or, if not coated, on the refractive index of the glass. As shown in figure 7H, the internal angle of incidence, ψ, increases with decreasing values of α'. Also the angle ψ is different for different rays of the cone. When ψ exceeds the critical angle of internal reflection, some of the light is transmitted at the internal reflecting surfaces and becomes lost. Consequently, if the principal ray is incident at the critical angle, then one-half of the observed image of the lens will appear poorly illuminated. If the reflecting areas of the prism do not appear in the emerging beam, these areas can be optically coated for increased reflection (as shown in fig. 8). The above discussion also applies to hexagonal prisms (fig. 8B). In this case, the entire areas of the reflecting surfaces may be coated.

A N_4EWS_p prism having a value of 24° for α' will provide total internal reflection to a 15° cone of light (corresponding to an $f/3.8$ cone) if it is made of borosilicate crown glass ($n=1.517$). The same shape prism, if made from a dense flint glass ($n=1.72$), will provide total internal reflection to a 25° cone, which corresponds to an $f/2.3$ beam.

We have described in section 2 above how to obtain and control the angle of shear by rotating one prism relative to the other (all $NEWS_p$ prisms), We have also described how shear may be obtained by using component prisms with different angles. Shear may also be obtained by purposely introducing (by grinding and polishing) pyramidal error into the prisms. Obviously this has the same effect on deviating the beams out of the plane of figure 1A that a relative rotation of the components has when the two components are cut from a single original prism.

In general, when two surfaces are cemented together, as described above, they will not be parallel. Also the vertex of the wedge between them may have any direction. Consequently, neither component of this wedge, perpendicular or parallel to the edges of either prism component, will be zero.

If the dividing plane is confined to one surface, the cement effectively becomes a part of the component

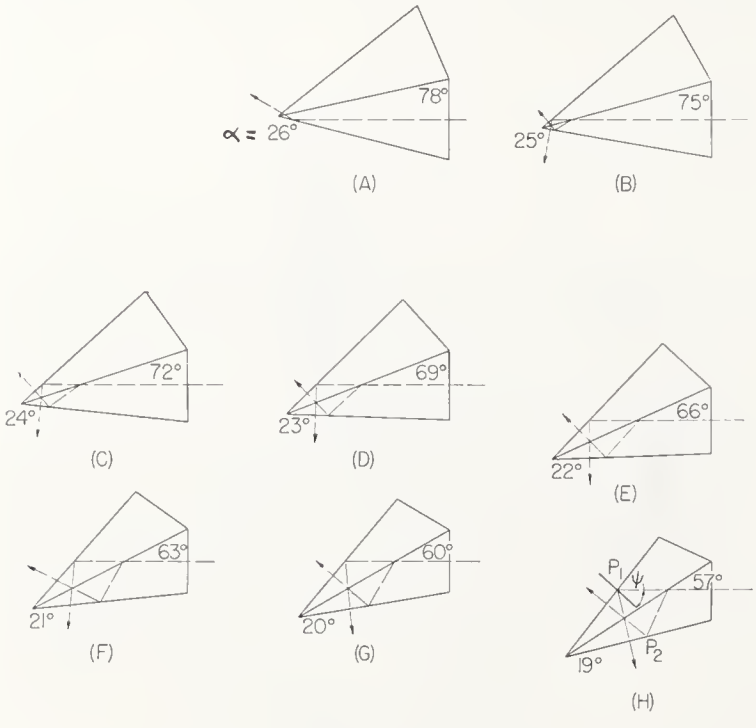


FIGURE 7. *Several prism interferometers showing the effect on the rays, from above the upper limit of α (=26°) to the lower practical limit (=19°).*

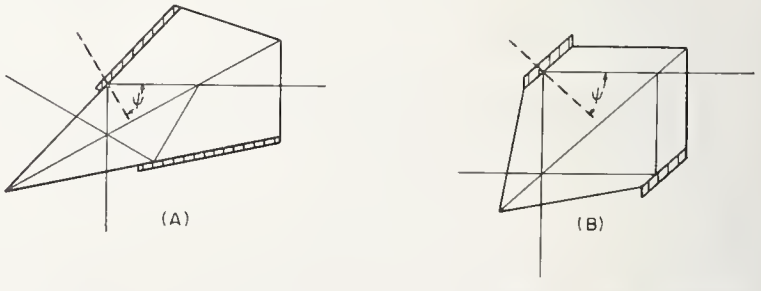


FIGURE 8. *Prism interferometers with opaque reflecting surfaces to prevent transmission of light when the angle of incidence, ψ, is larger than the critical angle.*

prism that is uncoated and we have dissimilar prism components even though the original components were identical. In general, two of the angles are different and one prism has pyramidal error.

5. Construction of the Prism Interferometer

Only prisms of optimum form will be described. All prisms are assumed to be free from pyramidal error, i.e., all refracting edges of a prism component are assumed to be parallel.

5.1. $N_6E_pW_\infty S_p$-Type Prisms

Using the symbolic designation described in section 3, the original *WSI* prism illustrated in figure 5A, was a hexagonal prism (N_6), with a plane entrance face (E_p). The two beam-dividing areas were produced by evaporating a semireflecting layer of aluminum on one of the inner surfaces of the prism (W_∞, dividing planes coplanar), and shear was introduced by a relative rotation of the two prism components.

Figure 9 illustrates how this prism was constructed from a single original prism (figure 9A) having angles β, γ, and ω. None of these angles is critical but their values may be advantageously chosen, as described above. The angle γ is removed by grinding and polishing a fourth face, A, that makes an angle, α, between it and the opposite face, B, as shown in figure 9B. The prism AB is then cut into two equal prisms, A_1B_1 and A_2B_2, by sawing it parallel to the plane of figure 9 This produces two quadrilateral prisms (figure 9C) that have identical angles, each of these component prisms being one-half as high as the original (minus the thickness of the saw cut).

A semireflecting coat of aluminum is put on surface B_2, and the two component prisms are cemented, together as shown in figure 9D. The thickness and variations of thickness of the cement are not critical. Optimum conditions, however, favor a thin parallel film of cement that has optical properties approximating those of the glass prisms. The reflectivity of surface B_1 is negligible and the cement effectively becomes a part of prism A_1B_1.

5.2. $N_6E_pW_\infty S_s$-Type Prisms

This prism is formed from two dissimilar prisms that have a difference in the two angles α_1 and α_2 (fig. 5B) that equals approximately one-half of the angle of shear divided by the index of refraction of the glass.

A practical procedure is to make two prisms, as described in section 5.1. above, with different values for α. After sawing the prisms into equal parts the pairs are crossed by cementing one component of an original prism to a component of the other. Thus, two prisms of this type are obtained.

The reason for making prisms long enough to be sawed into two components is because two small surfaces are more difficult to polish than one larger surface.

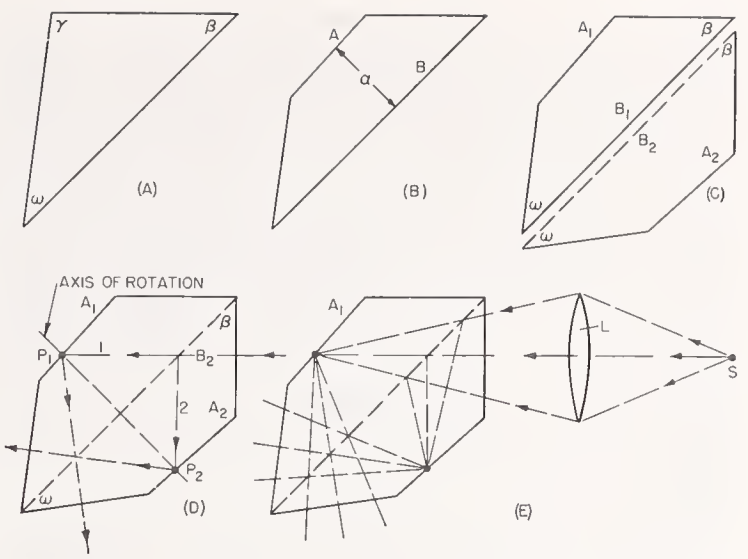


FIGURE 9. *Several forms of a prism interferometer during stages of its construction and adjustment.*

Figure 9 E shows how the prism acts upon a converging beam of light to produce two coherent images of the source.

5.3. $N_6E_pW_sS_p$ and $N_6E_pW_pS_s$-Type Prisms

These prisms differ from the $N_6E_pW_\infty S$ (either $N_6E_pW_\infty S_p$ or $N_6E_pW_\infty S_s$), described above, by having semireflecting dividing planes on each of the two surfaces B_1 and B_2 of figure 9C. The dividing plane of first incidence, B_2, extends from the vertex β of prism A_2B_2 to a vertical line in this plane that passes through the normal projection of P_2 onto it. The dividing plane of second incidence, B_1, extends from the vertex ω of prism A_1B_1 to a vertical line in this plane that passes through the normal projection of P_1 onto it.

5.4. NE_sWS-Type Prisms

The only difference between this prism and the NE_pWS described above is that a planoconvex lens is cemented to the entrance face. Best results are obtained if the lens and prisms are made of similar glass. The center of curvature of the convex surface should coincide approximately with the point P_1, where the principal ray suffers reflection from face A_1 of the prism.

5.5. N_4EWS-Type Prisms

The only difference between the N_4EWS prisms and the corresponding N_6EWS types described above is that the angle α is equal to ω. This changes the component prisms from quadrilateral to triangular in form. They are easier to construct, easier to adjust, and in some respects, have significant advantages over the N_6EWS types.

6. The Beam-Dividing Plane

The recommended optical coating for reflection and transmission of the beam-dividing plane is aluminum. Dielectric coatings that have been tried produced unfavorable polarization effects.

The ratio of reflection to transmission is more critical in this prism interferometer than in many other types of interferometers. If we assume R_1 and T_1 to be the reflection and transmission, respectively, of the beam divider at the first incidence and R_2 and T_2 the same at the second incidence, then the relative intensities of the several component beams may be designated by these quantities before and after each incidence on the beam divider. If we assume unit intensity for the beam before its first incidence, then the transmitted and reflected components are T_1 and R_1, respectively (see fig. 10). The T_1-beam is again divided at the second incidence into a $T_1\ T_2$-beam and a $T_1\ R_2$-beam. The R_1-beam is divided into a $R_1\ T_1$-beam and a $R_1\ R_2$-beam.

If the beam divider is coated with a uniform film, the difference between R_1 and R_2 (or T_1 and T_2) is due only to the difference in angle of incidence. The difference in angle of incidence varies from one type prism to another with variation in the prism angles, and also from one ray to another in the converging beam.

If the first division of the beam occurs at the inner surface of one prism (see fig. 4) and the second division at the inner surface of the other prism, then the difference between R_1 and R_2 (or T_1 and T_2) may be decreased by placing one prism closer to the source during the aluminizing process.

The contrast in the fringe pattern depends upon the ratio of the intensities of the two beams that combine to produce interference. Since the difference between R_1 and R_2 (or T_1 and T_2) is usually much smaller than the difference (T_1-R_1 or T_2-R_2), the pair of beams $R_1\ T_2$ and $T_1\ R_2$ are nearer equal than the pair $R_1\ R_2$ and $T_1\ T_2$. Consequently, the former pair will produce the best contrast and are, therefore, the beams usually used for making observations. When the fringes are to be photographed the $R_1\ R_2$-$T_1\ T_2$ beams will be used for visual adjustments for and during exposure.

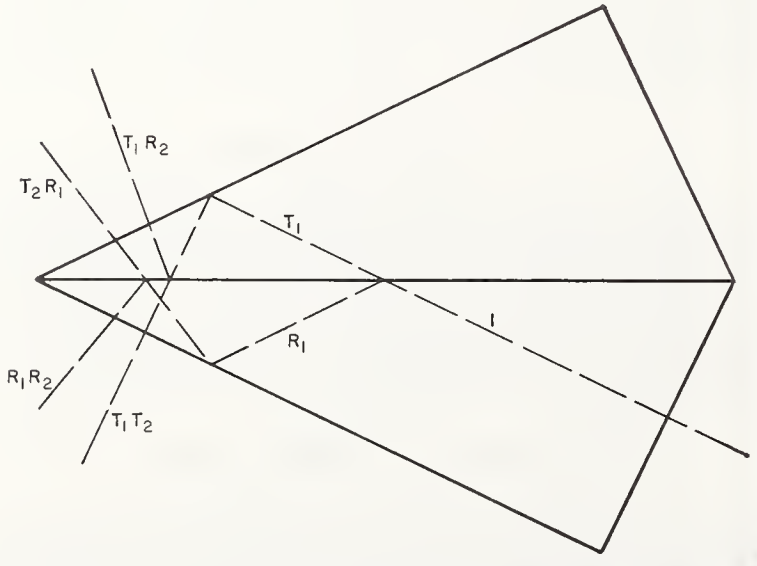


FIGURE 10. *Ray trace showing relative intensities of the four emerging beams*

An adequate test for the ratio of reflection to transmission of the beam divider is to observe visually the two images of a small source by means of the $R_1\ R_2$-$T_1\ T_2$ beams. If the images appear to differ in intensity by no more than a factor of about 2, the contrast and intensity of fringes produced by the $R_1\ T_2$-$T_1\ R_2$ beams will be adequate.

7. Optical Cement

Any optical cement that can be worked as a liquid, becoming hard at ambient temperatures, can be used in these prisms. Canada balsam has been used quite successfully. Its hardness is adjusted, by distillation of volatile constituents, to a consistency considered proper by an experienced optician for cementing lenses.

The advantages of Canada balsam over other cements is that its viscosity may be varied at will with temperature, and almost unlimited time is available for making the adjustments. This cement should be cooked until it is quite hard at ambient temperatures. Other cements are more stable than Canada balsam, do not require elevated temperatures, and do not permit unwanted changes in the adjustments after the prism is put to use.

8. Prisms Support and Adjusting Screws

The adjustment of the *WSI* prism is quite similar to that for the Kösters prism described in reference [8]. The two type prisms, N_6EWS and N_4EWS, require different blocks for adjusting the prisms.

A photograph of the block used for adjusting N_6EWS-type prisms is shown in figure 11. It consists of a compartment that approximates the shape and is slightly larger than the prism. Four vertical adjustable screws contact the bottom surface with

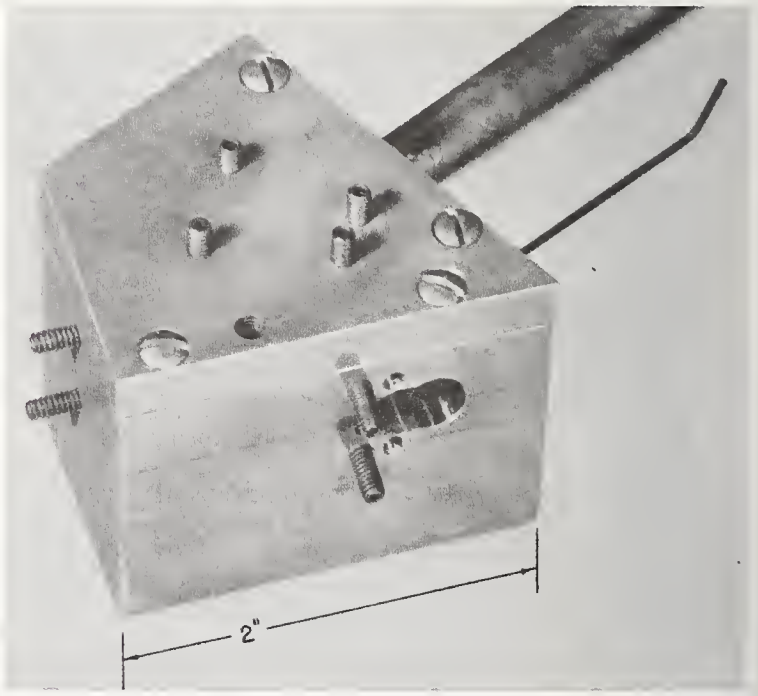


FIGURE 11. *Photograph of adjustable block for adjusting N_6EWS-type prisms.*

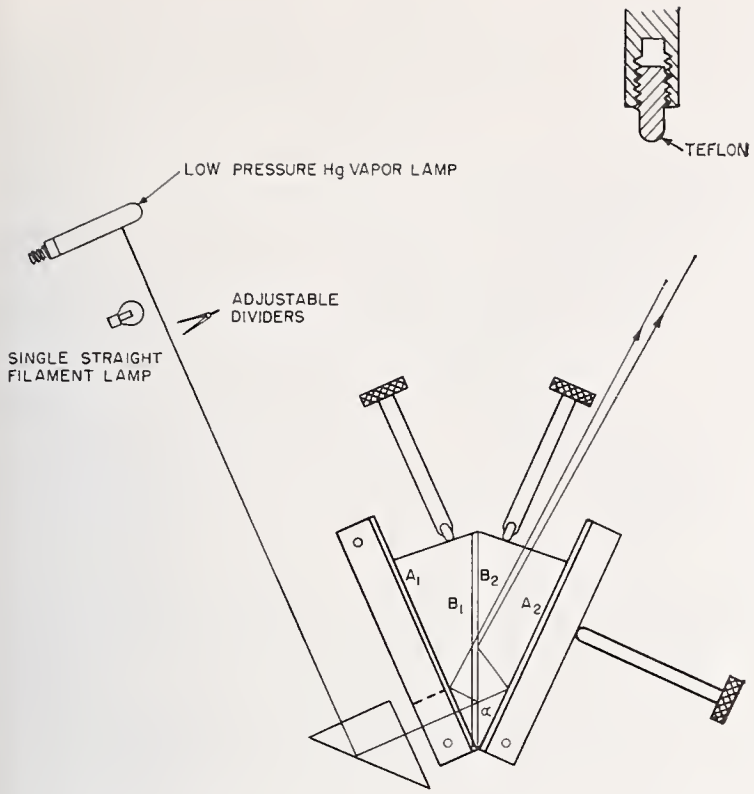


FIGURE 12. *Schematic diagram of V-block for adjusting N_4EWS-type prisms.*

The inset shows how the contacting ends of the screws are modified to avoid injuries to the glass surfaces. The adjustable dividers serve to measure the angle of shear.

two screws supporting each component prism. The top of the block has four vertical screws placed directly above those in the base of the block.

Two horizontal screws contact the entrance face of the prism and two others contact the exit face of the same component prism. Also, four similarly placed screws contact the corresponding faces of the other component prism. An entrance aperture, not shown in the photograph, and an exit aperture permit the light to pass through the prism to the observer. A supporting rod is shown for holding the block steady during the adjustment. This block may be heated on a simple hotplate.

The *V*-block for adjusting the N_4EWS prisms illustrated in figure 12 is very similar to that used for the Kösters prism. Additional detail that applies to this adjustment appears in reference [8], which should be consulted.

9. Illumination for Adjusting the Prism

Monochromatic fringes of interference are obtained from a low-pressure mercury vapor lamp. This lamp is cylindrical in form, is approximately 3 cm in diameter, and 16 cm long. White light fringes are obtained from a galvanometer lamp that has a spherical globe 2.5 cm in diameter and a single straight filament. A right-angle prism is used to reflect the light into the prism.

The mercury lamp is located, as shown in figure 12, approximately 120 cm from the prism. The galvanometer lamp is in line with the prism and mercury lamp and approximately 10 cm closer to the prism. This permits simultaneous use of both sources. The light passes through the prism in the reverse direction of that for normal use. This reversal of direction may be applied to almost any interferometer.

A low power telescope that permits a broad range of focusing is quite helpful for viewing and counting the interference fringes that are formed in the prism. These fringes are localized as in the conventional Mach-Zehnder interferometer.

10. Adjustment of the Prism Interferometer

The adjustment of quadrilateral prisms (N_4EWS) is no different from that of hexagonal prisms (N_6EWS). Consequently, a description of one type is sufficient. Since the former type is simpler in structure and easier to adjust because of differences in the mechanism for performing the adjustments, the N_4EWS type will be used for this description.

If the entrance face is to be spherical (NE_sWS), a plane entrance face prism is constructed and adjusted before the planoconvex lens is cemented to the plane entrance face.

A description of the adjustments of all types of prisms is provided by describing those for the four types shown in figure 13. Figures 13A and 13B show prisms with dissimilar components ($\alpha_1 \neq \alpha_2$). Figures 13C and 13D show prisms with similar components. The two beam divisions occur at a common dividing plane in figures 13A and 13C and at different planes in figures 13B and 13D.

There are four adjustments to be performed on a prism. Some of these are important and some are relatively unimportant to the performance of the prism. These adjustments are as follows: (1) to adjust the optical path of the two components of the

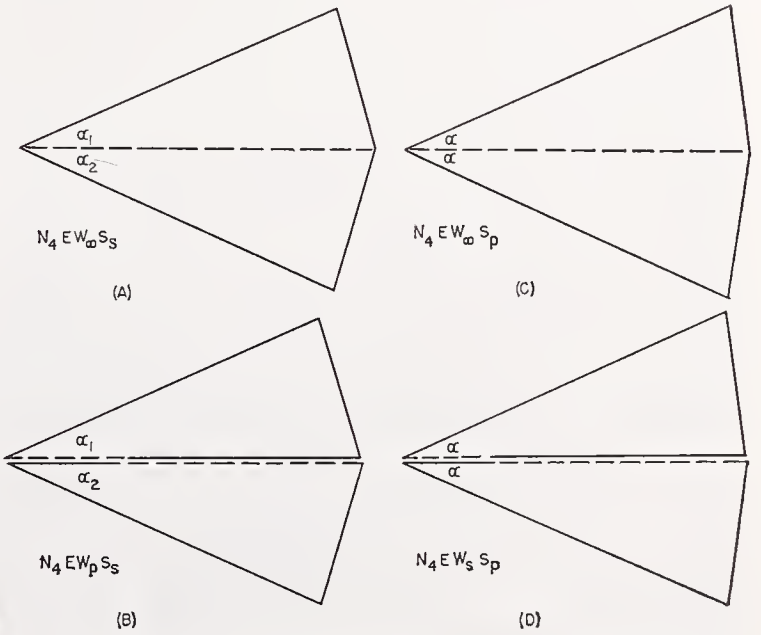


FIGURE 13. *Four types of prism interferometers.*

principal ray 1 and 2 (fig. 9D) to equality so that the zero order of interference is centered in the fringe pattern; (2) to adjust the thickness, t, of the cement; (3) to parallel the film of cement so that surfaces B_1 and B_2 (fig. 9C) are approximately parallel to each other; and (4) to adjust the magnitude of the angle of shear, ϕ.

There is no definite order for making these four adjustments except that number (4) should be fourth and last. The other three can be performed in any order. However, by first making adjustment number (1), and maintaining it, fringes of good contrast are available for making the other adjustments.

10.1. Interferometric Fringes for Observing Prism Adjustments

There are two types of interference fringes that are used for making prism adjustments. One of these sets of fringes is nonlocalized and the other is localized with a high degree of variability in the position of the plane of localization. The nonlocalized fringes are produced by two images of a small source. A lens or mirror must be used to produce an image of the source and the prism forms two images of it by amplitude division of the beam. The arrangement shown in figures 1 and 9 with a suitable source will fulfill the requirements. The direction and angular width of the nonlocalized fringes, mentioned in section 2 above, depend, respectively, upon the direction of the line joining the two coherent images of the source and on the angle subtended by this line at the point where the fringes are observed. This source must be small [1] in the direction of shear. The nonlocalized set of fringes is unnecessary for making adjustments except when a large angle of shear is used. When ϕ is large the localized fringes cannot be seen without the use of a long-focus microscope because they are so closely spaced.

The localized fringes may be used for making all necessary adjustments if the value of ϕ is not too large (approximately 0.02 rad). These fringes are, in all respects, similar to the fringes produced by a Mach-Zehnder [6] interferometer. They may be localized anywhere from $+\infty$ to $-\infty$. The plane of localization shifts so much during initial adjustments that a low-power telescope with a large focusing range is quite useful. These fringes may also be seen with the unaided eye. An extended source may be used when the fringes have moderate spacing and the order is small, but a source that is narrow in the direction of shear improves the contrast and is necessary when the fringes are very closely spaced.

When the dividing plane is confined to one plane as in the $NEW_{\infty}S$-type prisms (figs. 13A and 13C) and an extended or narrow source is used without a lens, as shown in figure 12, only one set of fringes is seen. When the dividing planes are divided between the two prism components (figs. 13B and 13D), three sets of fringes are seen. Figure 14 illustrates how the paths of the light rays differ for the three sets.

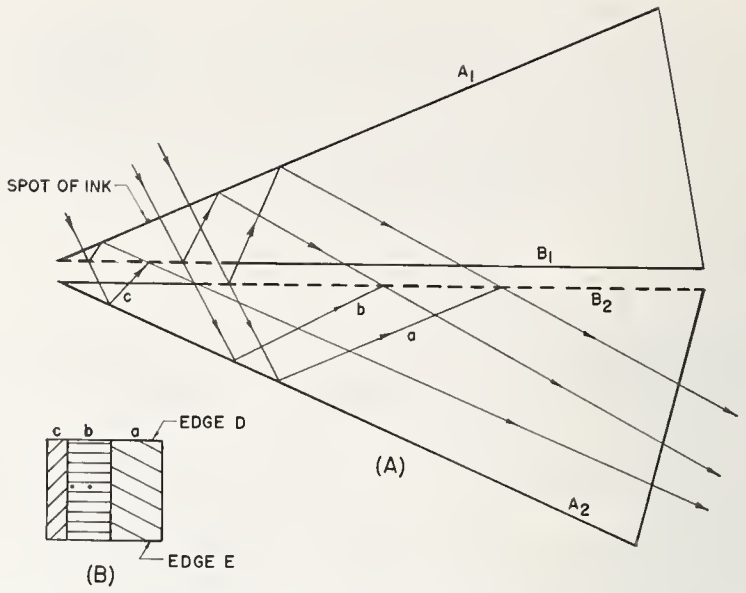


FIGURE 14A. *Ray trace showing how three sets of interference fringes are obtained when the dividing planes are on the two surfaces B_1 and B_2.*

B. *Appearance of fringes when illuminated as illustrated in figure 12. The edges, D and E, are the intersections of face A_2 with the base and top surfaces of the prism.*

10.2. Preparation of Prism for Adjustments

The prism components are cemented together initially in much the same manner as described for the Kösters [8] prism. The cement is assumed to be Canada balsam. The faces of the prism are cleaned. A small spot of ink is then placed on the exit face, A (which becomes the entrance face during the adjustment), approximately 2 mm from the vertex of angle α and midway between the edges D and E (fig. 14B) of the prism. The compound prism is placed in the adjustable V-block (fig. 12) and the assembly in an oven. See reference [8] for oven characteristics. Using the mercury lamp, illustrated in figure 12, and observing the spot of ink one sees three spots. One is a direct image of the inkspot and the other two are images in silhouette produced by light reflected from faces A_1 and A_2. The images of this spot aid the observer in finding fringes and in placement of the zero-order fringe when adjustments are complete.

10.3. Adjustment Procedure

The separation of the two images of the spot of ink, described above, is a coarse (or rough) measure of the prism adjustment. A low-power telescope or a long-focus microscope is convenient for observing both the spot and the interference fringes. A lateral separation (in the direction perpendicular to edges D and E, fig. 14B) of the images indicates either a relative rotation of the prism components or a lateral wedge in the cement. A separation in the direction perpendicular to the vertex of α indicates a difference in the two light paths. In general, if one prism is moved relative to the other in this latter direction, the two images change their relative separation in this direction and, consequently, can

be made to approach coincidence. When they're close together, fringes of interference can usually be seen in the section (a, b, or c of fig. 14B) in which the images are seen.

If the prism is of the NEW_pS_s or NEW_sS_p type, the original inkspot should be located at a distance from the vertex, α, such that the two images of it appear in the central section, b, of figure 14B, because this section transmits the light when the prism is used later for making measurements. The relative movement of the images of the inkspot is performed by sliding one prism on the other with the cement as a fluid lubricant. Heating the cement permits the viscosity to be altered at will.

After interference fringes are obtained with the mercury lamp their visibility can usually be improved by reducing the order of interference. When the order approaches zero, white light fringes may be seen with the tungsten lamp. Both lamps may be in use simultaneously. The order of interference is altered by a relative movement of the prisms perpendicular to their refracting edges (vertices of the angles).

When the direction of the wedge is known the adjustable V-block and its screws can be adjusted to correct the cement wedge. Reducing the angle of the block and applying pressure to the tops of the prisms causes the cement wedge to become thinner at the top. Conversely, by opening the V-block wedge until it is larger than the sum of the prism angles, 2α or $(\alpha_1+\alpha_2)$, the cement wedge becomes thinner at the bottom. These adjustments are described in reference [8] and apply equally well to this prism. A zero wedge in the cement is desirable for all prism types $NEWS_p$.

The wedge in the cement of prism types NEW_pS_s (fig. 13B) is unimportant. This prism is relatively easy to adjust if the angle $(\alpha_2-\alpha_1)$ is not large. The fringes are rotated parallel to the refracting edges of the prism and the zero order is adjusted to center (on the inkspot shown in fig. 14A). Best compensation is obtained for white light use if the thickness of the cement is small. The effect of cement thickness on the nature of the interference fringes is insignificant in prism types $NEWS_s$ (fig. 13B) but is important in prism types NEW_sS_p (fig. 13D).

The localized fringes in prism types $N_4E\ W_\infty S_s$ (fig. 13A) will always be relatively narrow, since the angle between the two emerging wave fronts approximates the angle of shear and is usually too large to yield broad fringes. The wedge in the cement adds to this angle and is difficult to distinguish from the component due to the prism angles. However, the cement wedge can always be reduced to insignificance by heating it to produce low viscosity and applying a little pressure on the prisms. Even if a wedge remains in the cement of this prism, it can be counteracted or compensated for by rotating one prism relative to the other until the fringes are parallel to the vertices of the prism angles. Thus, a residual wedge in this prism does little harm.

The localized fringes in the $N_4EW_\infty S_p$-type prism (fig. 13C) will be very broad if there is no wedge in the cement and no relative rotation of the component prisms. If the fringes become parallel to the prism's refracting edges when one prism is rotated about an axis normal to the dividing plane, the vertex of the cement wedge is parallel to that of the prism. The maximum fringe width, during this rotation of prism components, is then a measure of the magnitude of the wedge. The direction can be determined by the usual test for increasing orders. If the reflecting film is on surface B_2 (fig. 12) and the cement is thicker at the top than at the bottom, the fringes will move away from the vertex of angles α when prism A_2B_2 is moved downward. Also, they will move toward the vertex of α when prism A_1B_1 is moved downward or prism A_2B_2 is moved upward.

The light that passes from a source, arranged as indicated in figure 12, through a NEW_sS_p-type prism produces three sets of interference fringes. These are indicated in figure 14B and labeled a, b, and c. One of the two interfering beams in each set is reflected from A_1. The second beam in each set travels the following paths: beam a is reflected from B_2 to A_2 and finally from B_2 again; beam b is reflected from B_2 to A_2 and finally from B_1; beam c is reflected from B_1 to A_2 and finally from B_1 again. The set a is identical to the fringes seen in prism type $NEWS_p$ (fig. 13C). Set c is the same as would be seen in this prism should it be rotated 180° in the adjustable V-block. The fringes in set b are independent of the wedge in the cement (assuming plane surfaces and homogeneous media). Consequently, these fringes will always be normal to the refracting edges of the prism. A relative rotation of the prisms will cause these fringes to become very broad. This occurs when Ω (defined above) becomes equal to 0. When this condition exists the fringes in set a are parallel to the vertex of the wedge formed by the two surfaces B_1 and B_2. Also, the width of these fringes is a measure of the magnitude of this wedge. The direction of increasing thickness of the cement wedge is the same as the direction of movement of the fringes (in set a) when prism A_2B_2 (fig. 12) is moved downward. When the direction and magnitude of the cement wedge are known, corrective measures may be taken to reduce it. When the surfaces B_2 and B_1 become parallel, all three sets of fringes will remain equally wide as Ω is varied for introducing shear. The angle of relative rotation of the prisms, Ω, may be held to a value that produces fringes of a convenient width for counting. By adjusting the zero-order fringe in set b to a reference point on the boundary between b and c the difference in order between sets b and c can be counted while component prism A_1B_1 is moved upward relative to component A_2B_2. This difference in order multiplied by the wavelength of the light is equal to $2t\sqrt{n_2^2-n_1^2\sin^2 3\alpha}$, where t is the thickness of the cement at the bottom of the prism (fig. 12), n_2 and n_1 are the refractive indices of the cement and glass, respectively. Thus the thickness of the cement wedge may be measured.

A similar count of the difference in order between sets a and b will also give a value for t. The change

of zero order in set b to zero order in set a is obtained by moving component prism A_2B_2 downward relative to component A_1B_1. This difference in order multiplied by the wavelength equals $2t\sqrt{n_2^2-n_1^2sin^2\alpha}$. Since this latter quantity is larger than that for sets b and c by a factor of 2 or 3, and since this count is not critical, the smaller value will usually be used for obtaining t.

If t becomes too small it can only be increased by recementing the prisms. Spacers in the form of paper strips placed along the margins (edges D and E, fig. 14B) will serve to control the thickness of the cement. If t is found to be too large, it can be readily reduced by prolonged pressure on the prisms which causes the cement to be forced from between the prism components. Usually a value of t that is within 50 percent of the optimum value will be satisfactory.

When the required value of t is obtained the prism components are rotated to produce the desired shear, the zero-order fringe is adjusted to the center of the prism aperture, and the temperature is reduced sufficiently slowly (approximately 25° per hour or less) to avoid excessive gradients during cooling.

11. Angular Deviation by the Prism

The mean deviation of a pair of interfering beams is the angle between the entrance beam and the bisector of the angle between any two emerging beams that combine to produce interference.

The range of deviation of interfering beams that may be obtained with this prism interferometer is considerable. There may be occasions when it is desirable to have a mean deviation of a pair of interfering beams equal to 0°, 90°, or both. The N_4EWS prism cannot be made to produce zero deviation. A 90° mean deviation, however, is obtained with the N_4EWS_p prism by making $\alpha=22.5°$. The N_6EWS_p prism will produce a 90° mean deviation of two interfering beams if $(\omega+\alpha)=(\beta-\alpha)=45°$. The N_6EWS_s prism will produce a mean deviation of 0° for one pair of beams and 90° for the other pair of interfering beams if $\alpha_1=-\alpha_2$ and $\beta=\omega=45+2\alpha_1$.

12. Summary

Many types of prisms have been constructed and tested. They are found to be quite useful for measuring the shapes of wave fronts and changes in wave fronts with time. Prisms with large shear angles yield higher accuracy when testing high-quality (telescopic) optical systems, and prisms with small shear angles yield relatively high accuracy when testing systems that have large aberrations (large OPD variations).

Important advantages of the several prism types are: N_6EWS prisms can be most economically made by grinding and polishing one surface, A, (fig. 9B) on right-angle reflecting prisms, which are readily available commercially. N_4EWS prisms are easier to make from unworked glass because only three faces have to be polished, instead of the four required for the N_6EWS prisms. The range of deviations for the N_6EWS prisms is different from that of the N_4EWS type, and this is sometimes important. The vertex of the cone of light, received by an N_4EWS prism (image S, fig. 1), can be much closer to the exit face than for the N_6EWS prism, which is important for observation of large cones of light from high-power optics, because the size of the pupil is limited. The NEW_sS prism permits rotation of the fringes whereas the NEW_pS prisms do not. The N_6EWS prism can be made to accept a larger cone of light than that for the N_4EWS type. The N_4EWS prisms requires a simpler fixture for making adjustments (fig. 12) than does the N_6EWS type (fig. 11). These and other differences permit the choice of a best type for any given application.

All prisms used in this development were made by Mr. John V. McDermott.

13. Appendix A

Reference Points

The analysis of data taken with the WSI requires that the reference points be chosen in a manner that permits the application of the operations described in reference [5]. The reference points should be equally spaced on a straight line (usually a diameter of the lens or mirror being tested) that is parallel to the direction of shear. The angular spacing of the points is equal to the angle of shear. Such a set of points is defined as a "family" of reference points.

If the reference points are labeled (see fig. 15A) P_0, P_1, P_2 . . . P_N, where $(N+1)$ is the number of points in the family, the corresponding images will be labeled (fig. 15B) P_0' and P_0'', P_1' and P_1'', P_2' and P_2'', . . . P_N' and P_N''. One image of each point will coincide with images of the adjacent points. One image of P_2 (i.e. P_2') will coincide with image P_1'' and the other image of P_2 (i.e. P_2'') will coincide with image P_3''. If one point (such as P_2 in fig. 15C) is displaced from its family position the images of it will not concide with those of the neighboring points (fig. 15D). If the line of reference points (fig. 15E) is not parallel to the direction of shear the images will not coincide in pairs.

The number of points in a family depends upon the angle of shear and the angular aperture of the lens or mirror being tested. The ratio of aperture angle to shear angle is equal to the number of spaces between the points plus the excess spaces between the border and end points. The excess fraction of spaces allows for a possible change of one in the number of members of a family. For instance, if there are 3½ spaces and one point is chosen at the center, then only 3 points can be included in the family. However, if one point is chosen near the margin, there will be 4 points in the family.

The position of the family of points on the line is not critical. When one point is chosen, however, the positions of all members of the family are known.

The chosen point may be at the center, at the margin, at any chosen distance from the margin or center, and on any line parallel to the direction of shear. The chosen line is usually a diameter of the circular aperture (if it is circular) but may be any chord of this circle that is parallel to the direction of shear and of sufficient length.

It sometimes happens that the number of points in a family is too small and, consequently, the spacings are too large for sufficient data. For example, if an $f/11$ (5.2° aperture angle) telescope objective is to be tested with a prism whose angle of shear equals one degree (see fig. 16), the largest number of reference points that may be included in a family is six. This is obtained by choosing one point (whose images, P_0' and P_0'', appear as dots in the figure) near the margin. If one point (whose images P_5' and P_5'' appear as small circles in fig. 16) is chosen at the center of the lens, it will be found that the family consists of only five points; and since the number of reference points in the fringe pattern is one less than the number in the family, there will only be four points for reading data. Five or six points may not provide data at sufficiently close intervals along the line of reference points—especially if the optics that are being tested have zonal errors.

The number of usable points along the line of reference points may be increased by using two or more families. Thus, two families of points (labeled family No. 1 and family No. 2, in fig. 16) may be combined to decrease the angular spacing to 0.5°. A third family may be added to provide additional points. Any integral number of families may be added to obtain data at as many points as desired.

Reference [5] describes how to compute for one family of points. We will show here how to apply more than one family. When two or more families of points are combined, it is desirable to use a different unit for ν and to choose equal spacings for the points. This permits the use of integers for subscripts with no added complications. When two families are to be used, the separations of the points equal one-half the shear. When three families are used, the spacing is one-third the shear, etc.

We will use the fringe pattern of reference [5] (shown in fig. 17A) to show how to apply two families of points. The second family is added (fig. 17B) as small circles located midway between adjacent points of the original family of points. The orders of interference, for all points in the fringe pattern, are listed in column 2 of table 1.

The equations of observation (No. 5, ref. [5]) for family No. 1 are

$$\left.\begin{aligned}
\delta_2&=\delta_0+2.74+r-2\epsilon\\
\delta_4&=\delta_2+2.10+r-4\epsilon\\
\delta_6&=\delta_4+1.70+r-6\epsilon\\
\delta_8&=\delta_6+1.45+r-8\epsilon\\
\delta_{10}&=\delta_8+1.30+r-10\epsilon\\
\delta_{12}&=\delta_{10}+1.00+r-12\epsilon\\
\delta_{14}&=\delta_{12}+0.30+r-14\epsilon
\end{aligned}\right\} \qquad (10)$$

and the set for family No. 2 are

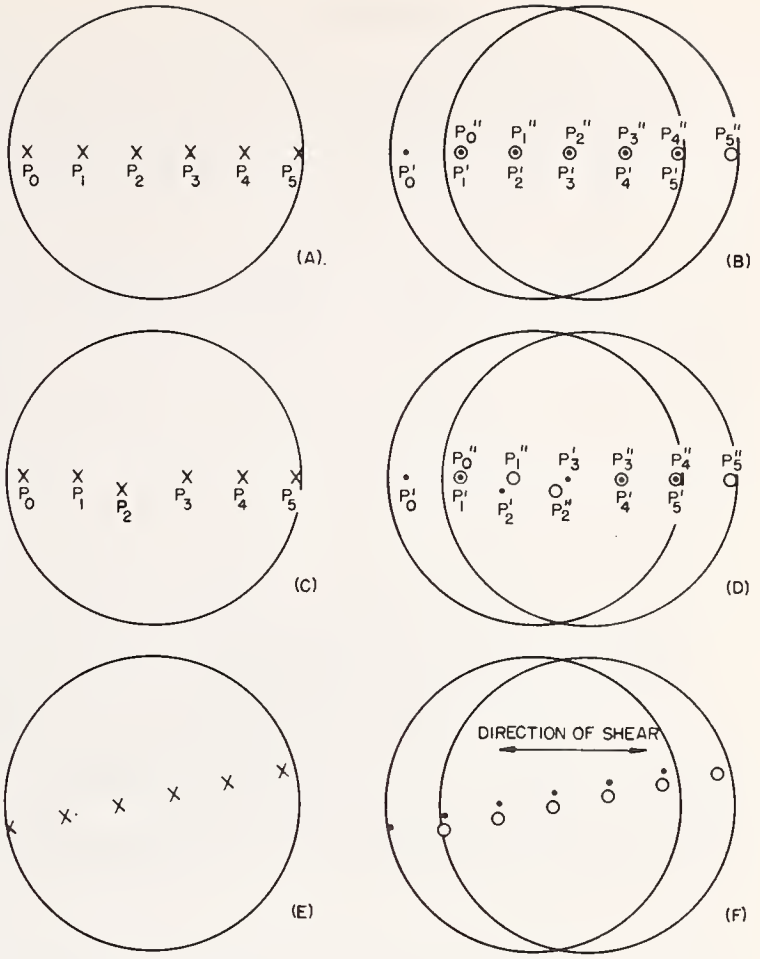


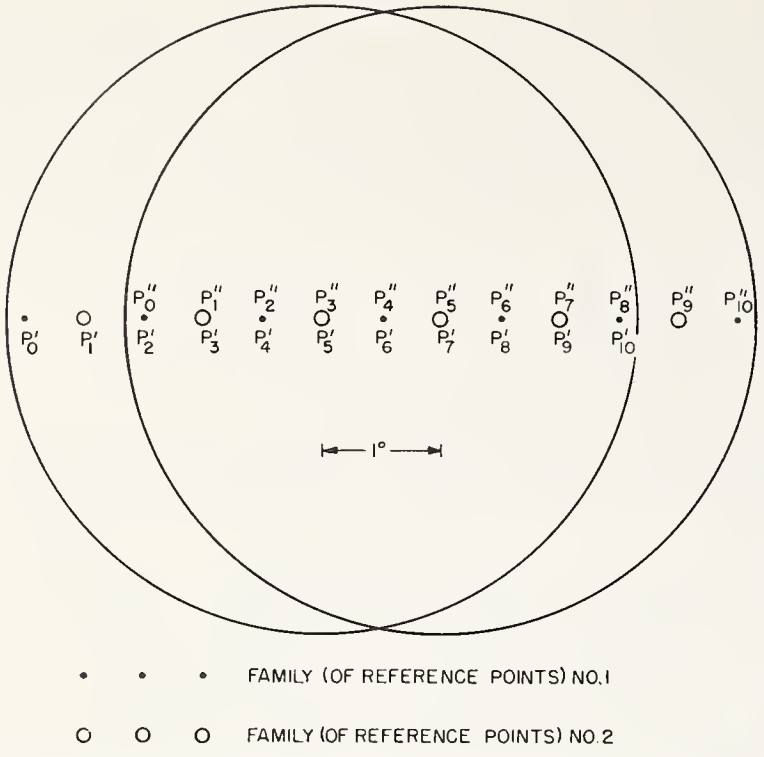
FIGURE 15. *Illustration of the characteristics of a family of reference points.*

$$\left.\begin{aligned}
\delta_3&=\delta_1+2.40+r-3\epsilon\\
\delta_5&=\delta_3+1.88+r-5\epsilon\\
\delta_7&=\delta_5+1.56+r-7\epsilon\\
\delta_9&=\delta_7+1.37+r-9\epsilon\\
\delta_{11}&=\delta_9+1.18+r-11\epsilon\\
\delta_{13}&=\delta_{11}+0.71+r-13\epsilon
\end{aligned}\right\} \qquad (11)$$

When using two or more families of reference points, the methods of solution described in reference [5] are impractical. The deviations (δ's) of one family will not, in general, fall on the same curve (which represents the shape of the wave front) that passes through the other family of deviations. The two families may be required to fall on one and the same wave front shaped curve by applying Newton's interpolation formula [9] and requiring that it be a polynomial of degree $(N-1)$. This is obtained by requiring that,

$$\delta_0+(1-N)\delta_1+\frac{(1-N)(2-N)\delta_2}{2!}+\frac{(1-N)(2-N)(3-N)\delta_3}{3!}+\cdots+\frac{(1-N)(2-N)\ldots(-1)\delta_{(N-1)}}{(N-1)!}=0 \qquad (12)$$

where N is the total number of points in the two families.



• • • FAMILY (OF REFERENCE POINTS) NO.1

○ ○ ○ FAMILY (OF REFERENCE POINTS) NO.2

FIGURE 16. *Illustration of how two families of points may be combined to increase detail in the computed wavefront.*

The dots, labeled P_0, P_2, P_4, P_6, P_8 and P_{10} form a family of points that are defined as family No. 1. The circles, labeled P_1, P_3, P_5, P_7, and P_9 form another family of points that are defined as family No. 2.

The two families of deviations are referred to a common reference sphere by requiring that

$$\left.\begin{aligned}\delta_0+\delta_1+\delta_2+\delta_3+\delta_4=0\\ \delta_5+\delta_6+\delta_7+\delta_8+\delta_9=0\\ \delta_{10}+\delta_{11}+\delta_{12}+\delta_{13}+\delta_{14}=0\end{aligned}\right\} \tag{13}$$

and the deviations are required to fit the same wave front curve by requiring ($N=15$ in equation (12)) that

$$\delta_0-14\delta_1+91\delta_2-\ldots\delta_{14}=0 \tag{14}$$

On combining eqs (10), (11), (13), and (14) and solving them as a set of simultaneous equations, we obtain the deviations shown in table 1, column 3.

Equation (14) may be replaced by the equivalent simpler equation,

$$\begin{aligned}15(\delta_1+\delta_{13})-90(\delta_2+\delta_{12})+365(\delta_3+\delta_{11})\\ -1000(\delta_4+\delta_{10})+2003(\delta_5+\delta_9)\\ +3002(\delta_6+\delta_8)-3432\delta_7=0,\end{aligned}$$

which is obtained by subtracting $\sum_{\nu=0}^{14}\delta_\nu$ (which from 13, equals 0) from 14.

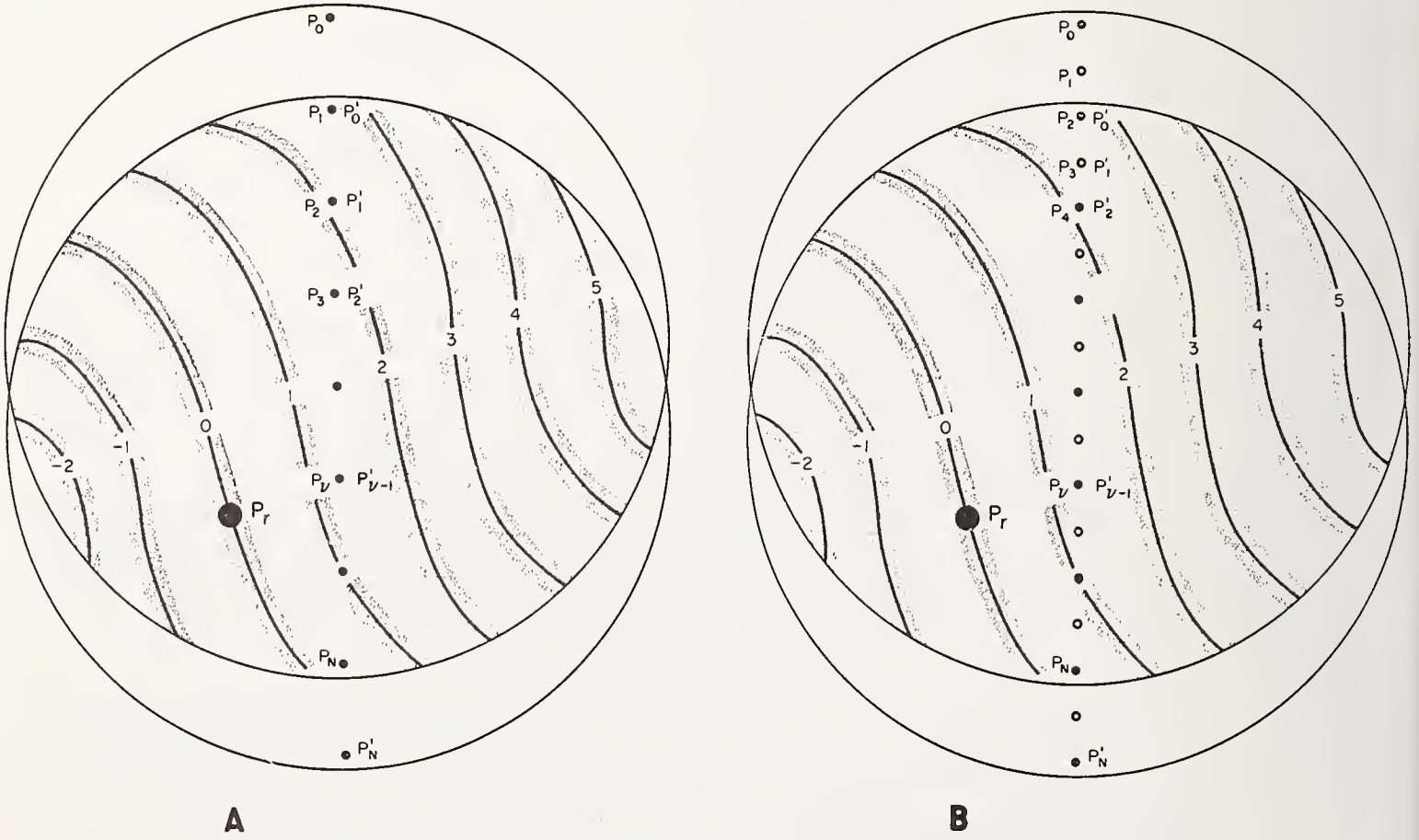


FIGURE 17A. *A fringe pattern with one family of points.*

B. *A fringe pattern with two families of points.*

Table 1. *Observed orders of interference, q_ν, and computed deviations, δ_ν, from figure 17*

ν	q_ν	δ_ν
0	----	−0.28
1	----	−.05
2	2.74	.09
3	2.40	.13
4	2.10	.11
5	1.88	.07
6	1.70	.02
7	1.56	−.02
8	1.45	−.04
9	1.37	−.02
10	1.30	.04
11	1.18	.09
12	1.00	.11
13	0.71	.01
14	.30	−.24

(Paper 68C3–160)

14. References

[1] Bates, W. J., Proc. Phys. Soc. (London), **59,** 940 (1947).
[2] Drew, R. L., Proc. Phys. Soc. (London), B, **64,** 1005 (1951).
[3] Brown, D., Proc. Phys. Soc. (London), B, **67,** 232 (1954).
[4] Brown, D. S., J. Sci. Instr., **32,** 137 (April 1955).
[5] Saunders, J. B., J. Res. NBS **65B** (Mathematics and Mathematical Physics), No. 4 (Oct–Dec 1961).

Errata-Reference [5]

Change the term $\sum_{\sigma=0}^{\nu-1}\sigma$ in each equation, (8) and (11), to read $\sum_{\sigma=1}^{\nu}\sigma$. Change the third member of eq (12) to read $\delta_1+3\delta_2+6\delta_3+10\delta_4+15\delta_5+21\delta_6+28\delta_7=0$. Change the second member of eq (16) to read $14\delta_0+11\delta_1+9\delta_2+8\delta_3=0$.

[6] Candler, C., Modern Interferometers, p. 487, Hilger & Watts Ltd. (1951).
[7] Jenkins, F. A., and White, H. E. Fundamentals of Optics, p. 241, McGraw-Hill Book Co., Inc. (1957).
[8] Saunders, J. B., Construction of a Koesters double-image prism, J. Res. NBS **58,** 21 (1957) RP2729.
[9] Milne-Thompson, L. M., The Calculus of Finite Differences, p. 2, Macmillan and Co., Ltd., St. Martin's Street, London (1951).

Reprinted from **APPLIED OPTICS,** Vol. 6, page 1581, September 1967

A Simple, Inexpensive Wavefront Shearing Interferometer

J. B. Saunders

National Bureau of Standards, Washington, D. C. 20234.
Received 8 May 1967.

A beam divider made by cementing together the hypotenuse faces of two right-angle prisms, one of which is half silvered, is potentially a wavefront shearing interferometer. This beam di-

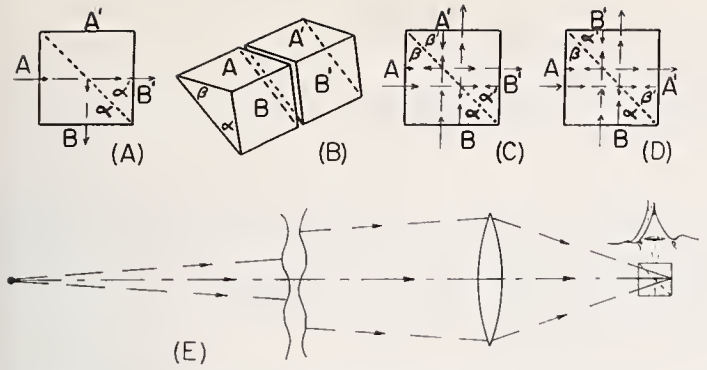


Fig. 1. Wavefront shearing prism interferometer. (a) Conventional prism beam divider; (b) two identical prisms may be obtained by cutting one prism; (c) and (d) four possible combinations of two different prisms; (e) optics for testing a lens with the cube interferometer.

vider may be adjusted for perfect chromatic compensation if the prisms are made of the same glass and one pair of adjacent angles [α and α', Fig. 1(a)] are equal. If the two prisms are cut from one longer prism [Fig. 1(b)] equality of angles and glass is automatically achieved.

Ideally the refractive index and the dispersion of the cement should closely match those of the glass. However, the layer of cement may be made so thin that a small difference in those properties will result in only imperceptible deviation from perfect compensation. The cement should be sufficiently slow setting to permit adequate time for making the adjustments. In both these respects epoxy cement is satisfactory.

Only simple equipment is needed for making the adjustment. A 90° V block is convenient for holding one prism with the hypotenuse face horizontal. Beeswax is used to stabilize the adjustment during the setting period of the cement. An incandescent lamp is a suitable light source for use when making the adjustment for chromatic compensation. However, a monochromatic source is helpful for locating the fringes originally.

The adjustment is readily performed by moving the upper prism, with the fingers, over the horizontal hypotenuse face of the lower prism, mounted in the V block. Chromatic compensation is achieved by translating the upper prism, in a direction that equalizes the distances from a point on the center of the beam divider surface to the two reflecting base surfaces. Perfect compensation is achieved when the prisms are oriented so that the fringes are quite broad and the zero-order fringe is centered in the aperture of the prism. The two component beams that emerge from face A' [Fig. 1(c)] are then parallel and coincident. When satisfactory adjustment has been accomplished the prisms are allowed to rest in that position until the cement has set.

The motivation for white light compensation is to permit the use of relatively broad band spectral filters. The limitation of source size, that may be used with this interferometer, is given by Bates.[1]

The angle between the two emerging beams can be adjusted at will by rotating one prism relative to the other about an axis normal to the beam dividing face. This operation converts the beam divider into a wavefront shearing interferometer. It also produces a narrowing of the localized fringes so that perfect compensation is achieved only along the narrow zero-order fringes. A magnifier may be useful for observing the fringes during the adjusting to ensure that the zero-order fringe remains at the center. The direction of shear will be approximately parallel to the vertex edges of the prisms.

If a dark background is used, the beams of light reflected from the uncoated surfaces B and B' [Fig. 1(c)] are sufficiently bright that fringes may be observed, located near surface B, while the adjustments are being made. However, the brightness of the fringes is greatly enhanced by applying high reflecting coatings to faces B and B'.

It is sometimes convenient to have the direction of shear perpendicular to the vertex edges of the prism. This requires that the angles α and β be unequal, as is usually the case. When the prisms are combined, as in Fig. 1(d), and the fringes adjusted to maximum width, the angle of shear will equal $2(\alpha - \beta')$. Larger shear angles are obtained, as above, by rotating one prism relative to the other. This rotates the direction of shear to an angle intermediate between the two above-mentioned preferred directions.

If the prisms are made separately, the angles α, β, α', and β' will usually be unequal and there are four possible angles of shear, the direction of shear being normal to the vertex edges of the prism. These four angles are $2(\alpha - \alpha')$ when the light enters face A [Fig. 1(c)], $2(\beta - \beta')$ when the light enters face B [Fig. 1(c)], $2(\alpha - \beta')$ when the light enters face A [Fig. 1(d)], and $2(\alpha' - \beta)$ when the light enters face B [Fig. 1(d)]. Each of these four arrangements requires a different pair of high reflecting surfaces.

Prism angles need not be held to close tolerances. Relatively large pyramidal errors contribute negligible effects to the results. The description of adjusting procedures given by Saunders[2,3] for other prism interferometers apply also to the cube interferometer, with slight changes for direction of illumination.

Figure 1(e) shows an arrangement for testing a camera lens with this interferometer. One sees two images of the aperture of the lens, displaced from each other by the introduced shear. Fringes of interference appear in the overlapping area of the combined image. From this pattern of fringes, as made with a series of wavelengths, the monochromatic and chromatic aberrations of the lens can be determined. If the source is narrow in the direction of shear, the contrast in the fringes will be quite high because the intensity of the two component beams will always be equal. The recommended size of the cube is 10 mm to 15 mm. A small interferometer permits the pupil of the observer's eye to receive a larger cone of light than can be received through a larger interferometer.

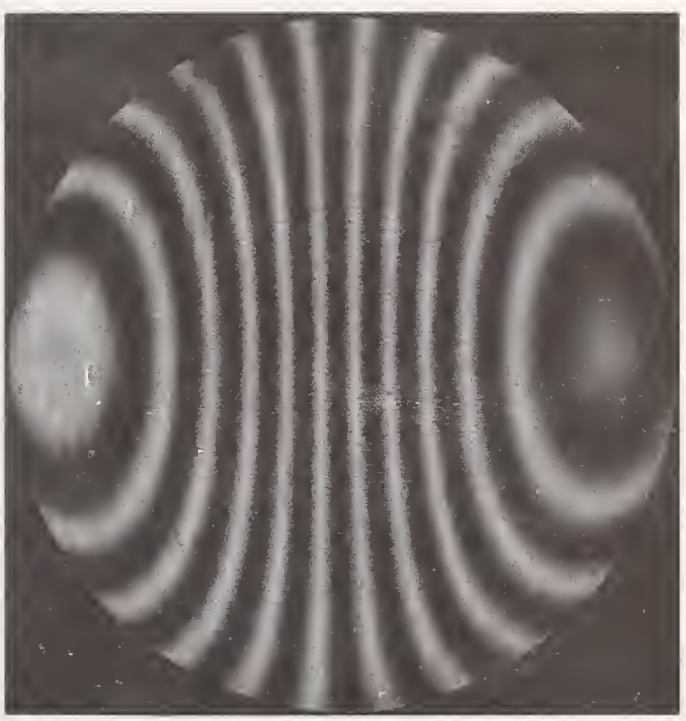


Fig. 2. Photograph of fringes with a 15-mm cube prism and a 9-cm aperture, $f/4.5$ achromatic lens.

A commercially made 15-mm cube beam divider was readjusted and two of its outer faces aluminized to produce the above described interferometer and the photograph shown in Fig. 2.

Information received subsequent to the submission of this manuscript indicates that reference should also be made to Lenouvel and Lenouvel.[4]

References

1. W. J. Bates, Proc. Phys. Soc. **59,** 940 (1947).
2. J. B. Saunders, J. Res. Natl. Bur. Std. **58,** 21 (1957).
3. J. B. Saunders, J. Res. Natl. Bur. Std. **68C** (1964).
4. Léan Lenouvel and François Lenouvel, Rev. Opt. **17,** 350 (1938).

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 73C, Nos. 1 and 2, January–June 1969

An Interferometer for Measuring Gradients in Both Refractive Index and Thickness of Large or Small Optics

J. B. Saunders

Institute for Basic Standards, National Bureau of Standards, Washington, D.C. 20234

(March 24, 1969)

A small aperture prism interferometer is described for measuring refractive index and thickness gradients between pairs of adjacent points in both small and large optics. It is relatively free from vibration problems and thermal gradients are reduced to a minimum by confining the two component beams of light to a relatively small space. Large specimens are tested by scanning them with the light beam.

Key words: Interferometer; optical homogeneity; refractive index.

1. Introduction

The classical instrument for measuring variations in refractive index of glass was described by Twyman and Green [1][1] in 1916. This interferometer is quite suitable for testing small specimens but is impractical for testing specimens larger than 8 to 10 in. because of vibrations, temperature problems, and difficulties in making large aperture interferometers.

The Kösters prism interferometer was suggested for measuring the homogeneity of optical media in 1963 [2] but a description of this application has not been published. This interferometer is quite practical for measuring index gradients in both small and large optics. It is relatively free from vibration problems and thermal gradients are reduced to a minimum by confining the two component beams of light to a relatively small space. Large specimens are tested by scanning them with the light beam.

2. Components of the Interferometer

Figure 1 shows the optics of a modified Kösters prism interferometer. The light source may be either a continuous emission laser or a highly monochromatic spectral source, such as Krypton or Mercury 198. If a laser is used, a short focus (high power), single element, plano-convex lens that has its axis slightly tilted relative to the principal ray, will provide adequate divergence and sufficient freedom from unwanted interference fringes.

The Kösters prism interferometer has its two components cemented together for stabilization of several

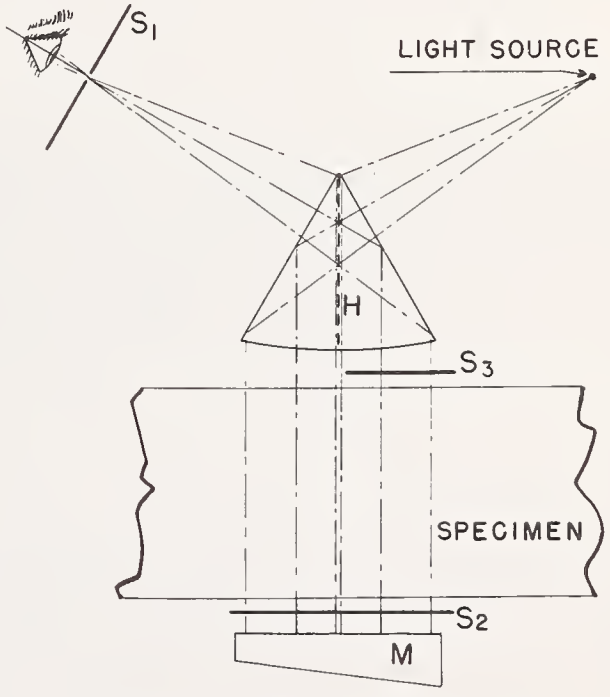


FIGURE 1. *Optics of the interferometer.*
The figure is a section through the center of the prism.

of its interferometric adjustments. These adjustments for a built-in-wedge prism [3] are: (1) one component of the fringe width is adjusted to optimum width, and (2) the zero order of interference is adjusted to the most favorable position (the center) where it remains during further operational adjustments. The adjustments for a prism without a built-in-wedge are to set the zero order of interference at the center of the prism and to maximize its width by polishing the base.

[1] Figures in brackets indicate the literature references at the end of this paper.

The two types of prisms are equally accurate but the integral parts of the orders of interference in the built-in-wedge type are easier to read.

The base surface of the prism is figured to be spherical and convex. The prism serves as: (1) a collimator lens, (2) a beam divider for dividing the beam two times, (3) two mirrors for deflecting the transmitted and reflected beams to a common direction, (4) a near perfect compensator so that white light interference can be used if desired, and (5) a collector lens.

If a laser is used for the source, the mirror, M, may be a small uncoated optical flat. However, if the mirror does not have an opaque reflecting coat, its bottom surface should either be tilted to avoid reflecting light back through the prism or it should be rough ground, or coated with something like petroleum jelly, cedar oil, or a similar fluid to reduce reflection. A sheet of black paper, soaked in the fluid and contacted to the back of the mirror, serves as an excellent nonreflection coat. The remaining components consist of the specimen and three simple screens S_1, S_2, and S_3.

A calibrated glass wedge [4], located between the specimen and prism, enhances the accuracy of reading the orders of interference.

3. Adjustment of the Interferometer

The only adjustments, other than those made invariant by cementing the prism components together, are positioning the light source and mirror relative to the prism. The source should be located near the focal plane of the prismatic lens formed by the base of the prism so as to collimate the two light beams (one from each of the two prism components) in the specimen area of the interferometer. The source should also be adjusted in the plane of fig. 1 so that the two component beams (reflected and transmitted by the beam divider) are parallel to the beam dividing plane.

The distance from the prism to the mirror may be chosen at will, however, adequate space should be allowed for free movement of the thickest specimen that is intended to be tested. The angular position of the mirror about an axis perpendicular to the beam dividing plane is not critical, except that the angle of incidence of the light should remain small. The only critical adjustment that requires attention is the angular position of the mirror about an axis parallel to the vertex of the prism. This adjustment, which can be controlled by a single screw that rotates the mirror relative to the prism assembly, controls the component width of the transmission fringes (defined below). If a prism with no built-in-wedge (fringes parallel to the dividing plane) is used, this adjustment of the mirror only causes the fringes to change separation without rotation. If a built-in-wedge prism is used it causes the fringes to rotate about their points of intersection with the dividing plane.

The fringes produced by light reflected from the two surfaces of the specimen are relatively independent of interferometer adjustments. If the surfaces of the specimen are adjusted to be approximately normal to the incident light, the width of the fringes depends only upon the refractive index and angle between the two reflecting surfaces of the specimen. The angle between the two surfaces of the specimen should form a 3 to 15 s wedge so that the two component beams reflected from them will form interference fringes of suitable widths for accurate measurements. The two surfaces should be flat to within 0.1 μm per inch to avoid excessive variation in fringe width.

4. Opaque Screens for Obstructing Unwanted Light

Three opaque screens (thin, black cardboard is quite adequate) are used to obstruct light which would otherwise interfere with fringe visibility. The single beam of light from the source is divided into two beams by the beam divider, H. When unobstructed by screens S_2 and S_3, each of these beams is incident on the two optical surfaces of the specimen and the mirror, M. Therefore, three components of each of the two beams return to the prism. These six beams are again divided by the beam divider into two components each. One set of six components is transmitted toward the observer and the other set is returned toward the souce. If the observer receives the six beams that are transmitted toward him he sees a complex interference pattern that is useless for the test.

A screen, S_1 (fig. 1), in the focal plane of the source that is dark on the observer's side, white on the source side and having a small round hole (approximately 1 mm in diameter), can be adjusted to transmit the pair of beams from M, if M is adjusted normal to the beam divider. Thus, when the specimen is absent the observer sees a single set of interference fringes. The set of fringes produced by light from M is defined as "transmission fringes" since the light that produces them is transmitted by the specimen during the test. The fringes produced by light reflected from the two surfaces of the specimen are defined as "reflection fringes."

When the specimen is inserted the four beams that are reflected from its two surfaces will each form an image of the source in the plane of S_1. If the surfaces of the specimen are approximately parallel to M the four beams will pass through the aperture in S_1 and interfere with the transmission fringes. These four beams may be adjusted, by slightly tilting the specimen, to fall just outside the aperture in S_1, with no observable effect on the position of the beams from M and no appreciable interference with the transmission fringes.

The reflection fringes are made visible by intercepting the light to M with a screen S_2 (preferably dark on the incident side), intercepting one of the beams from the prism with another screen S_3 and removing S_1 from the beam. This leaves only two beams reflected from the specimen to form the reflection fringes.

5. Movements of the Specimen

The specimen rests on a saddle that can be moved along a straight line normal to the dividing plane, along a set of ways, without rotation. The set of ways should have a scale attached to it that permits the sample to be moved equal intervals. A fixed reference point, P_1 (fig. 2), should be located between the prism and specimen, in the beam that forms the reflection fringes. A second reference point, P_2, symmetrical to P_1, with respect to the beam divider, will insure good visibility of the reference points in the transmission fringes. The separation of these two points equals the interval of movement of the specimen between observations as it traverses the light beam. These two reference points may be small beads on a fine wire attached to the prism cell and stretched normally to the beam dividing plane.

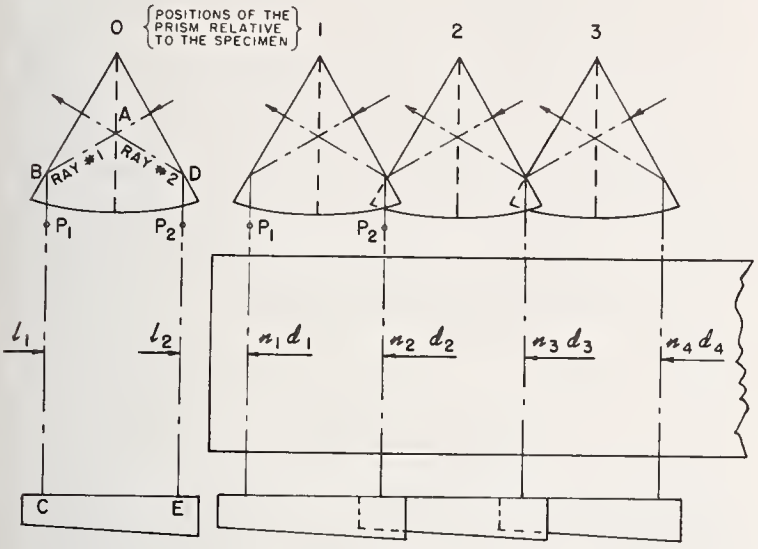


FIGURE 2. *Several positions of the specimen relative to the prism.* Observations are made without the specimen and with the specimen located at each of several positions relative to the prism.

Data can be taken along any straight or curved path across the specimen as long as ray 1 passes through the specimen in exactly the same place that ray 2 passed through before the specimen was moved from the previous position. The number of positions at which measurements are made depends upon the length of the sample and the separation of the reference points. The separation of the reference points may be any value from near zero to slightly less than the aperture of the prism. The prisms used for these tests had diameters of 3 cm.

Figure 2 shows several positions of the prism relative to the sample; although it is the sample that is moved relative to the prism. The prism, light source, reference points, and mirror are rigidly connected to insure interferometric stability. If we let the two optical paths, "ABC" and "ADE", be represented by l_1 and l_2, respectively, the order of interference, Q_0, observed at the superimposed reference points, P_1 and P_2 will be given by the equation,

$$Q_0 = 2(l_2 - l_1)/\lambda.$$

The adjustment for Q_0 is controlled by the single screw adjustment mentioned above. This adjustment should not be altered during the test.

If, when the prism is in any position $\nu (\nu = 1, 2 \ldots)$, we let T_ν and R_ν be observed orders of interference for the transmission and reflection fringes, respectively, and letting n_ν and d_ν be respectively the refractive index and thickness of the sample at the place where ray 1 traverses it, then it can be shown that:

$$T_\nu = Q_0 + [d_{\nu+1}(n_{\nu+1} - 1) - d_\nu(n_\nu - 1)]2/\lambda,$$

$$\Delta R_\nu \equiv R_{\nu+1} - R_\nu = [d_{\nu+1}n_{\nu+1} - d_\nu n_\nu]2/\lambda.$$

It will be noted that the final position of the specimen is required for evaluating R_ν but not T_ν.

If n and d be the nominal refractive index and thickness of the sample, respectively, in the direction traversed by the light, Δn_ν and Δd_ν the variations in n and d between the two interfering beams (rays 1 and 2), it can be shown that [5], to the approximation aimed at in this analysis

$$\Delta n_\nu = [n(T_\nu - Q_0) - (n-1)\Delta R_\nu]\lambda/2d, \tag{1}$$

$$\Delta d_\nu = [\Delta R_\nu - T_\nu + Q_0]\lambda/2. \tag{2}$$

The index gradient and wedge (angle between the two surfaces) of the specimen are obtained by dividing Δn_ν and Δd_ν, respectively, by the separation of the reference points. The total changes in index and thickness are obtained by adding the increments Δn_ν and Δd_ν, respectively. It can be shown that, to the same approximation mentioned above,

$$\sum_{\nu=1}^{\nu} \Delta n_\nu = n_{\nu+1} - n_1$$

$$= \left[n\left(\sum_{\nu=1}^{\nu} T_\nu - \nu Q_0 \right) - (n-1) \sum_{\nu=1}^{\nu} \Delta R_\nu \right] \lambda/2d, \tag{3}$$

$$\sum_{\nu=1}^{\nu} \Delta d_\nu = d_{\nu+1} - d_1$$

$$= \left[\sum_{\nu=1}^{\nu} \Delta R_\nu - \sum_{\nu=1}^{\nu} T_\nu + \nu Q_0 \right] \lambda/2. \tag{4}$$

It is sometimes important for specifications to know the maximum variation in the refractive index and thickness of large optics. When values for $(n_{\nu+1} - n_1)$ and $(d_{\nu+1} - d_1)$ are tabulated they show the maximum and minimum values, from which maximum differences are readily obtainable. Obviously the results apply only to that part of the specimen covered by the selected reference points and may not include the maximum variations of the entire specimen unless the points are chosen to coincide with them.

6. Reduction of Error and Elimination of Q_0

Since the wavelength, λ, the refractive index, n, and thickness, d, of the specimen are usually known more accurately than the orders of interference, the errors in the determined values of Δn_ν and Δd_ν arise almost entirely from the errors in the measured values of the orders of interference. This, of course, is based on the assumption that temperature gradients remain negligible.

It is apparent from eqs (3) and (4) that an error in Q_0 will produce errors in $(n_{\nu+1}-n_1)$ and $(d_{\nu+1}-d_1)$ that are ν times the corresponding errors in Δn_ν and Δd_ν. This error can accumulate to relatively large values when observations are made at many points across a large specimen. On the other hand, the error in ΔR_ν does not accumulate since it can never exceed twice the maximum error of a single observation. It can also be shown that all biased errors in Q_0, R_ν, and T_ν are automatically eliminated by differences. The result of random errors in all observations are obviously reduced by averaging the results of observations taken at the same points during two traversals of the specimen. If the specimen is rotated 180° between the two traversals and observations are made at the same points but in reverse order, the quantity Q_0 does not affect the average and, therefore, may be eliminated from the observations.

If the two traversals are indicated by subscripts of parenthesis, eq (1) becomes, for the first and second traversal, respectively,

$$\left.\begin{aligned}(\Delta n_\nu)_1 &= [n(T_\nu)_1 - n(Q_0)_1 - (n-1)(\Delta R_\nu)_1]\lambda/2d,\\ (\Delta n_\nu)_2 &= [n(T_\nu)_2 - n(Q_0)_2 - (n-1)(\Delta R_\nu)_2]\lambda/2d.\end{aligned}\right\} \tag{5}$$

If no adjustment of the prism, relative to the mirror, is made during the observations then $(Q_0)_1=(Q_0)_2=Q_0$. The constancy of Q_0 is assured by the stability of this instrument. The reversal of the specimen reverses the optical and geometrical wedges between all pairs of adjacent points with the result that values for $(\Delta n_\nu)_1$ and $(\Delta d_\nu)_1$ will be identical in magnitude and of opposite sign (except for errors) to $(\Delta n_\nu)_2$ and $(\Delta d_\nu)_2$, respectively. The best values for Δn_ν and Δd_ν are, therefore,

$$\overline{\Delta n_\nu} = \tfrac{1}{2}[(\Delta n_\nu)_1 - (\Delta n_\nu)_2]$$

$$= \{n[(T_\nu)_1 - (T_\nu)_2] - (n-1)[(\Delta R_\nu)_1 - (\Delta R_\nu)_2]\}\lambda/4d. \tag{6}$$

$$\overline{\Delta d_\nu} = \tfrac{1}{2}[(\Delta d_\nu)_1 - (\Delta d_\nu)_2]$$

$$= \{(\Delta R_\nu)_1 - (\Delta R_\nu)_2 - (T_\nu)_1 + (T_\nu)_2\}\lambda/4. \tag{7}$$

The corresponding values for $(n_{\nu+1}-n_1)$ and $(d_{\nu+1}-d_1)$ are

$$n_{\nu+1} - n_1 = \sum_{\nu=1}^{\nu} \overline{\Delta n_\nu}\,,$$

and

$$d_{\nu+1} - d_1 = \sum_{\nu=1}^{\nu} \overline{\Delta d_\nu}\,.$$

Care must be exercised to insure that the values obtained for T_ν and ΔR_ν in eqs (6) and (7) represent values for the same pairs of points on the specimen. Positive identification is enhanced by the fact that $(\Delta R_\nu)_1$ and $(\Delta R_\nu)_2$ will be equal in magnitude and of opposite sign (except for errors) and the algebraic sum of $(T_\nu)_1$ and $(T_\nu)_2$ will be equal to a constant. The value of this constant is $2Q_0$, which need not be determined.

7. References

[1] Twyman and Green, British Patent 103832.
[2] Saunders, J. B., Large aperture interferometers with small beam dividers, J. Res. NBS **67c** (Eng. and Instr.), No. 3, 205 (1963).
[3] Saunders, J. B., Parallel testing interferometer, J. Res. NBS **61**, 491 (1958) RP2917.
[4] Technical News Bulletin NBS **42**, No. 2, 33 (1958).
[5] Candler, C., Modern Interferometers, p. 143 (Hilger & Watts Ltd., Hilger Div., London, England (1951).

(Paper 73C 1&2–283)

Reprinted from APPLIED OPTICS, Vol. 9, page 1215, May 1970

High-Speed Holographic Interferometry

A. T. Funkhouser and K. D. Mielenz

Metrology Division, National Bureau of Standards, Washington, D.C. 20234.
Received 1 December 1969.

Holographic interferometry has been applied by various authors for vibration analysis, stress–strain analysis, flaw detection, photoelastic studies, microscopy, and so on. Low-speed motion picture recording of fringe motion has been done also.[1] Previous studies of high-speed phenomena, however, have been restricted to a few discrete samples in time of the subject of interest. In the present work, we have combined holographic interferometry and high-speed photography to obtain a continuous, time-resolved record of the event.

The particular subject of our study has been the dynamic distortion of solid state laser rods caused by absorption of optical pump energy. For the ruby rods studied, the distortion occurs and disappears rapidly, so that high-speed recording of the changing interference fringe pattern is required. The experimental arrangement used is diagrammed in Fig. 1. Light from a continuous laser was passed through a passive (not pumped) ruby laser rod. The wave front transmitted by the rod was recorded holographically by a photographic plate situated in a lucite tank, so that it could be developed in place.[2] Thus, when the rod was optically pumped, the new wave front it transmitted passed through the hologram and interfered with the wave front which had been recorded.

The optical pumping unit consisted of a helical flash lamp in a metal housing. The lamp and the rod were cooled by refrigerated, temperature-controlled water flowing through the housing. However, during the interferometric tests, the flow was stopped to avoid vibration. The mirrors that form the pulsed-laser cavity are usually mounted external to the housing; during most tests, they were not in place and the pulsed-laser did not lase.

A number of runs were performed in which the rod was optically pumped at various pumping levels. The results of two of these are summarized in Fig. 2, in which the fringe patterns at several times (measured in milliseconds from the onset of the pumping) are compared for two pump levels. The patterns illustrate the distortion buildup and the subsequent cooling. In each run, the maximum distortion occurred approximately 25 msec after the pumping.

The high-speed camera used to record the fringe pattern was operated without a lens at a nominal framing rate of 4000 frames per second. The photographic film used in the high-speed camera was 16-mm type 2485 high-speed recording film. The framing rate was determined by the film sensitivity and the power output (15 mW) of the continuous He–Ne laser. The optical pumping lasted approximately 1 msec and thus yielded only four frames. The fringe patterns were not resolved and the frames were overexposed due to flash tube light.

With a continuous laser of greater power, the framing rate can be increased, thereby making it possible to record the distortion patterns during pumping and decreasing the effect of the flash light. The method need not be restricted to transparent objects or thermal stress. A reflected wavefront and changes due to other causes can equally well be studied. High speed recording techniques other than photography, such as photoelectric detection of the fringe motion, can be used also.

The advantages of the holographic approach with respect to other interferometric techniques are the following:

(*1*) *Only the change is observed.* Since the hologram compensates for the static distortions of the object under study, only changes in the object made after the hologram is recorded are detected.

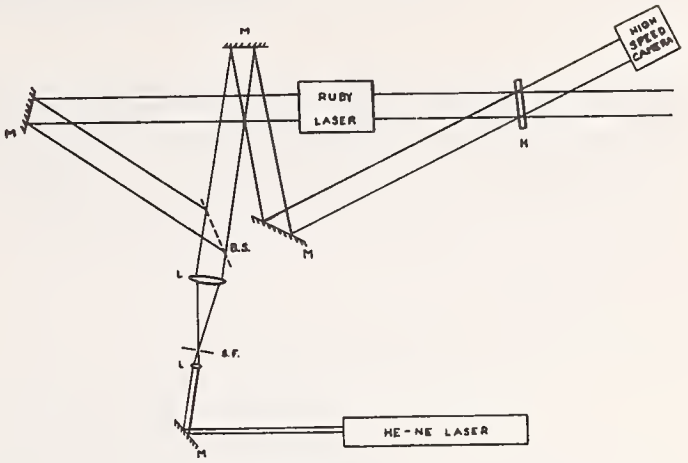


Fig. 1. Hologram interferometer: *M*, mirror; *L*, lens; *S.F.*, lens-pinhole spatial filter; *B.S.*, beam splitter; *H*, hologram.

(*2*) *Ease of adjustment.* Alignment of a hologram interferometer is easily accomplished, especially if the hologram is developed in place as was done in the work reported here.

(*3*) *Low cost.* Because the hologram compensates the wave fronts, relatively large aperture interferometers can be synthesized using inexpensive optics.

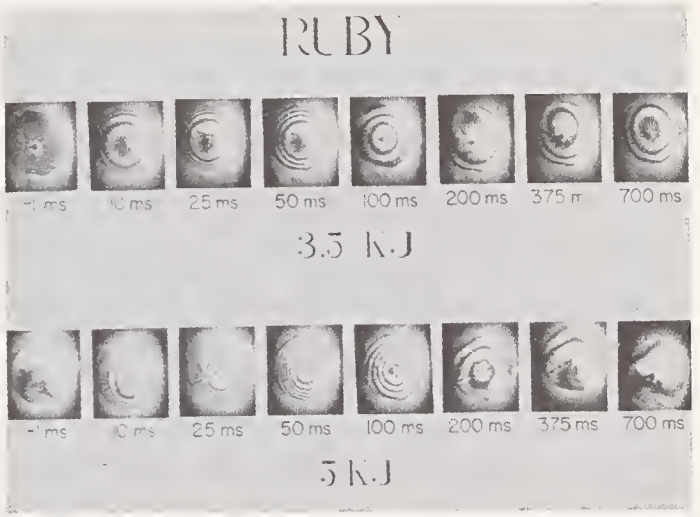


Fig. 2. Selected frames from high-speed recordings for two pump energies, with times elapsed from pumping.

The principal disadvantage of the holographic approach is the low diffraction efficiency (~10%) of the hologram. Thus for good fringe contrast, some light loss is necessary. However, as bleaching methods are improved, blazing of the hologram, thereby increasing the diffraction efficiency, should improve the performance of this method.

References

1. C. Durou and R. Lefevre, Compt. Rend. Acad. Sci. Paris, **267**, B259 (22 July 1968).
2. This method was devised by L. A. Kersch, C. F. Jacobson, and K. R. Porter at G. C. Optronics Corporation.

Reprinted from APPLIED OPTICS, Vol. 10, page 1567, July 1971

Measurement of the Second-Order Degree of Coherence by Means of a Wavefront Shearing Interferometer

D. N. Grimes

A method of measuring the second-order degree of coherence in an arbitrary plane by means of a compact wavefront shearing interferometer is described. The prism interferometer previously described by Saunders produces two sheared images that interfere, allowing measurement of the degree of spatial coherence. If the degree of coherence is a function of coordinate differences only, the visibility is constant; otherwise it varies over the field. Experimental results are given.

Introduction

Because of the increased understanding and use of the theory of partial coherence for describing the behavior of optical systems, it is often desirable to measure the degree of coherence in an arbitrary plane for quasi-monochromatic illumination. Although coherence theory was first formulated in terms of interferometry, the problem of measuring the degree of coherence in an arbitrary plane presents a different problem conceptually. Very few interferometers are suitable for such measurements in the laboratory. Historically, Zernike[1] proposed that the degree of coherence between two points could be determined by a measurement of the visibility of the fringes in a Young's experiment. Thompson and Wolf[2] modified this method by the introduction of two lenses that allowed the interference pattern to be observed at the focus of the second lens. These methods have the disadvantages that the degree of coherence for only two points in the field may be observed at one time and that for small separations of the two points one must use small pinholes which render the light levels too low for practical measurement of the visibility in most cases. More recently, Mallick[3] proposed the use of a polarization interferometer which is especially useful for measuring small separations. A discussion and review of these methods will be found in Ref. 4. The method discussed here utilizes a small compact shearing interferometer described by Saunders.[5] The sheared wavefronts produce fringes with visibility equal to the degree of coherence.

The author is with the National Bureau of Standards, Washington, D.C. 20234.

Received 7 December 1970.

Theoretical Discussion

The interferometer is a cube consisting of two identical right-angle prisms cemented together along their hypotenuse faces, one of which is coated to transmit and reflect equal amounts of light. The angle between the two emerging beams is adjusted to some fixed value by rotating one prism relative to the other about an axis normal to the beam dividing face. Such an interferometer cube is shown schematically in Fig. 1. When the device is used for measuring the degree of coherence in a plane S, a lens (L_1) of convenient focal length is inserted at S as shown in Fig. 2, with the cube placed near the focus of the beam. The interfering wavefronts are then observed with the lens L_2 through the prism.

The general problem of two-beam interferometry has been considered by Steel.[6] However, it is instructive here to particularize the development and consider only the pertinent parameters related to this wavefront shearing interferometer, namely, lateral shear and tilt.

It is convenient to analyze this system in terms of the analogous system shown in Fig. 3. The observed luminance distribution in the plane of the lens L_1 is the sum of the two wavefronts and may be described in the notation of coherence theory[7] as the time-averaged sum of the axial analytic signal and its displaced counterpart taken in complex quadrature. Using vector notation, where $\bar{x}_1$ is the vector[8] associated with a point on the axial wavefront and $\bar{x}_2$ that associated with a point on the displaced wavefront, the luminance in $\bar{x}$ is given by

$$\begin{aligned} L(\bar{x}_1,\bar{x}_2) = \langle (V(\bar{x}_1,t_1)\exp\{-ik[(\bar{x}_1^2/2z_2) - \varphi(\bar{x}_1)]\} \\ + V(\bar{x}_2,t_2)\exp\{-ik[(\bar{x}_2^2/2z_2) - \varphi(\bar{x}_2)]\}) \\ \cdot (V^*(\bar{x}_1,t_1)\exp\{ik[(\bar{x}_1^2/2z_2) - \varphi(\bar{x}_1)]\} \\ + V^*(\bar{x}_2,t_2)\exp\{ik[(\bar{x}_2^2/2z_2) - \varphi(x_2)]\}\rangle, \quad (1) \end{aligned}$$

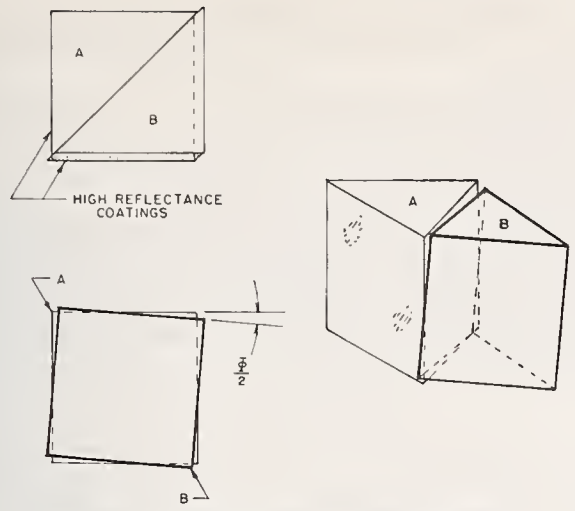


Fig. 1. Prism interferometer details.

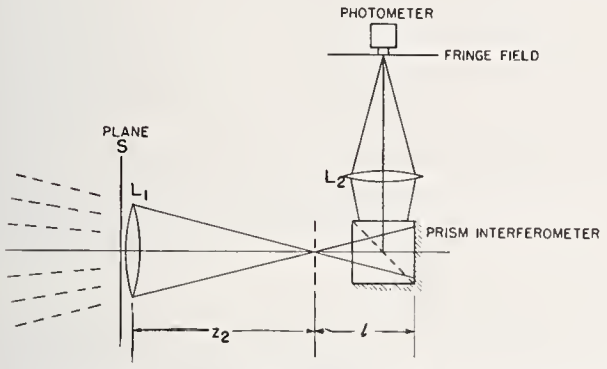


Fig. 2. Optical system for measurement of the degree of coherence.

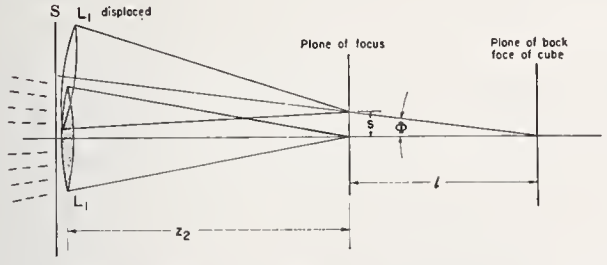


Fig. 3. Analog system for measurement of the degree of coherence.

where $V(\bar{x},t)$ is the analytic signal in the plane S, $\langle\ \rangle$ denotes the time average, and k is the mean wavenumber of the radiation. (The quasi-monochromatic approximation has been assumed.) The exponential term represents the convergent effects of the lens whose pupil function is $\varphi(\bar{x})$ and where z_2 is the distance from the plane of the lens to the focus. Using the definition of the second-order mutual coherence function

$$\langle V(\bar{x}_1,t_1)V^*(\bar{x}_2,t_2)\rangle = \langle V(\bar{x}_2,t_2)V^*(\bar{x}_1,t_1)\rangle = \Gamma(\bar{x}_1,\bar{x}_2;\tau) \quad (2)$$

and the relationships

$$\langle V(\bar{x}_1,t_1)V^*(\bar{x}_1,t_1)\rangle = I(\bar{x}_1),$$

$$\langle V(\bar{x}_2,t_2)V^*(\bar{x}_2,t_2)\rangle = I(\bar{x}_2),$$

where I is the luminance, $\Gamma(\bar{x}_1,\bar{x}_2;\tau)$ is the mutual coherence function, and τ is a time delay, the expression for the observed luminance distribution reduces to

$$L(\bar{x}_1,\bar{x}_2) = I(\bar{x}_1) + I(\bar{x}_2) + 2\Gamma(\bar{x}_1,\bar{x}_2;\tau) \times \cos\left\{k\left[\frac{\bar{x}_1^2 - \bar{x}_2^2}{2z_2} + \varphi(\bar{x}_1) - \varphi(\bar{x}_2)\right]\right\}. \quad (3)$$

If the two beams are of equal intensity so that $I(\bar{x}_1) = I(\bar{x}_2) = I_0$, we may use the relationship for the complex degree of coherence

$$\gamma(\bar{x}_1,\bar{x}_2;\tau) = \Gamma(\bar{x}_1,\bar{x}_2;\tau)/[I(\bar{x}_1)]^{\frac{1}{2}}[I(\bar{x}_2)]^{\frac{1}{2}}. \quad (4)$$

The observed luminance distribution of Eq. (3) may then be simplified and

$$L(\bar{x}_1,\bar{x}_2) = 2I_0 \times \left(1 + \gamma(\bar{x}_1,\bar{x}_2;\tau)\cos\left\{k\left[\frac{\bar{x}_1^2 - \bar{x}_2^2}{2z_2} + \varphi(\bar{x}_1) - \varphi(\bar{x}_2)\right]\right\}\right). \quad (5)$$

The displaced wavefront as shown in Fig. 3 has undergone a rotation through the angle Φ (the fixed shear angle of the cube), as well as a displacement s which is dependent on the focus position with respect to the back face of the cube. Assuming a shear in the x direction, we may write the $\bar{x}_2$ coordinates in terms of the $\bar{x}_1$ coordinates as follows:

$$\begin{aligned} x_2 &= (x_1 - s)\cos\Phi - z\sin\Phi, \\ y_2 &= y_1, \\ z_2 &= z\cos\Phi + (x_1 - s)\sin\Phi. \end{aligned} \quad (6)$$

Then

$$(\bar{x}_1^2 - \bar{x}_2^2)/2z_2 = 1/z_2[xs - (s^2/2)], \quad (7)$$

and since Φ is very small (usually a few milliradians), so that $\cos\Phi \approx 1$ and $\sin\Phi \approx \Phi$,

$$\bar{x}_1 - \bar{x}_2 = x - s - z_2\Phi. \quad (8)$$

Hence, $s + z_2\Phi$ is the displacement of one wavefront with respect to the other as seen through the prism.

Substituting Eqs. (7) and (8) into Eq. (5), the observed luminance distribution becomes

$$L(x) = 2I_0\left(1 + \gamma(x_1,x_2;\tau)\cos\left\{k\left[\frac{1}{z_2}\left(xs - \frac{s^2}{2}\right)\right] + k\varphi(x) - k\varphi(x - s - z_2\Phi)\right\}\right). \quad (9)$$

It is clear from the first term in the argument of the cosine that the observed luminance distribution consists of straight-line fringes of period $\lambda z_2/s$ displaced a constant amount $ks^2/2z_2$ over the field. The fringes are also shifted locally an amount $\varphi(x) - \varphi(x - s - z_2\Phi)$ which is the difference between the values of the pupil function at (x,y) and $(x - s - z_2\Phi, y)$. From measurements of the fringe displacement, one may determine the pupil function of the lens L_1 and consequently its transfer function. This application has been explored at the National Bureau of Standards, and a paper is currently in preparation.

However, of interest here is the visibility or contrast of the fringes which is equal to the degree of coherence

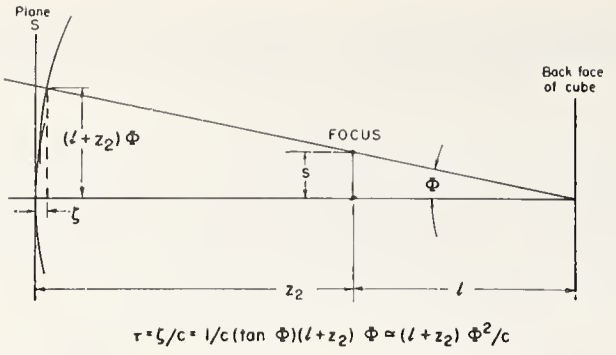


Fig. 4. Parameters for the calculation of the time delay τ.

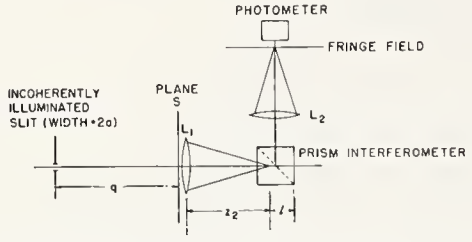


Fig. 5. Experimental setup for measurement of the degree of coherence.

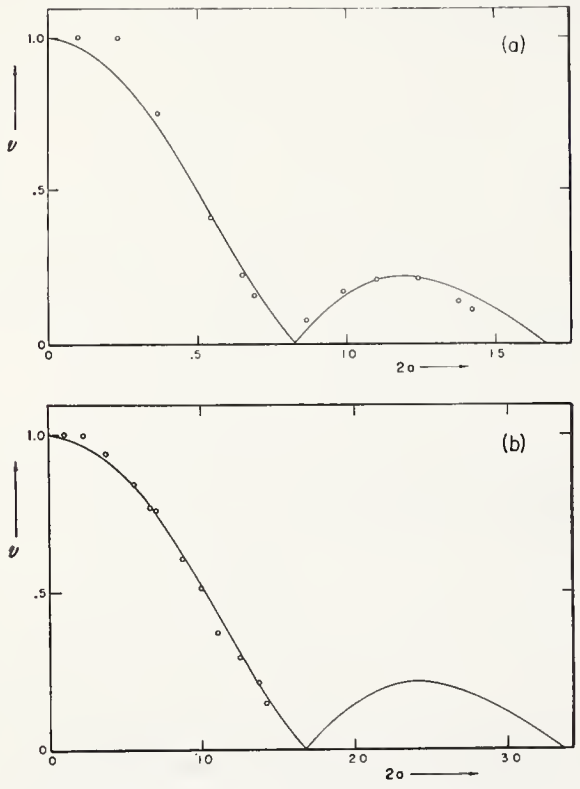


Fig. 6. Experimental data points from the measurement of the visibility ($\mathcal{V}$) equal to the degree of coherence $|\mu_{12}|$ together with the theoretical curve calculated from Eq. (12); (a) $\Phi = 2.27$; (b) $\Phi = 1.02$.

between the points $\bar{x}_1$ and $\bar{x}_2$ separated an amount $s + z_2\Phi$. Hence, the modulation at any point in the fringe pattern depends on the degree of coherence between the points (x,y) and $(x - s - z_2\Phi,y)$ in the plane S. If the degree of coherence is a function of coordinate differences only, i.e., $\gamma(\bar{x}_1,\bar{x}_2;\tau) = \gamma(\bar{x}_1 - \bar{x}_2;\tau)$, the contrast or visibility of the fringes will be constant over the field. Otherwise, the visibility will vary from point to point.

In general, the measured degree of coherence is time-independent, that is,

$$\gamma(\bar{x}_1,\bar{x}_2;\tau) = \mu_{12}. \tag{10}$$

This may be shown by examining the value of τ introduced by the system. (Because the quasi-monochromatic approximation has been assumed, there is no time delay associated with the incident field.) As shown in Fig. 4,

$$\tau = (l + z_2)\Phi^2/c. \tag{11}$$

For a 1-cm cube of 10-mrad shear angle and a 50-mm focal length lens, the value of τ is approximately 2×10^{-16} s, which is about $\frac{1}{50}$ of the period associated with the green mercury line. This is negligible, and we may therefore consider the visibility to be a measure of the time-independent degree of coherence.

To determine the degree of coherence for varying point separations, one may vary the shear angle Φ of the cube, the focal length of the lens L_1 that will change z_2, or the focus position of the beam relative to the back face of the cube that will change s. In the latter case, the frequency of the fringes will change and may pose a measurement problem for extreme values.

Experimental Results

Experimental measurements were made utilizing the system shown in Fig. 5. In the plane S which is located in the far field of an incoherently illuminated slit aperture, the degree of coherence is calculated by the Van Cittert-Zernike theorem to be the normalized fourier transform of the luminance function of the aperture, or

$$|\mu_{12}| = |\sin(ka\overline{P_1P_2}/q)/(ka\overline{P_1P_2}/q)|, \tag{12}$$

where $2a$ is the slit width and $\overline{P_1P_2}$ is the separation of the two points. The degree of coherence in the plane S was determined experimentally by measurement of the visibility ($\mathcal{V}$) of the fringes[6] produced by the wavefront shearing interferometer. With an infinite slit source, the cube must be critically aligned with the shear in the direction of measurement. The lens L_2 was used to image the plane S onto the photometer slit as shown in Fig. 5. The observed fringe frequency was maintained at less than 0.10 cycle/mm so that the effects attributable to the imaging lens were negligible. The measured visibility data for a cube with a shear angle of 2.27 mrad is shown in Fig. 6(a) together with the curve calculated from Eq. (12). Additionally, the shear angle of the cube was changed to 1.02 mrad and produced the data shown in Fig. 6(b). Again the calculated coherence function is shown for comparison.

In general, internal reflections, cube alignment, slit alignment, etc. will produce measured visibility values lower than the actual values. While most such sources of error may be eliminated, some internal reflections are unavoidable in actual practice. This probably accounts for the experimental error in Fig. 6(a) and (b). The error producing higher values for smaller slit

widths is due to lack of sensitivity in the photometer used for measurement.

Conclusions

This compact shearing interferometer offers an alternative to existing methods of measurement of the degree of coherence in an arbitrary plane. The light efficiency of the system (approximately 50%) allows measurements to be made. Although the smallest point separations measured were about 80 μm with a shear angle of 2.27 mrad and a focal length of 32 mm, it is feasible to measure separations as small as 5 μm with a shear angle of 0.1–0.2 mrad.

References

1. F. Zernike, Physica **5**, 785 (1938).
2. B. J. Thompson and E. Wolf, J. Opt. Soc. Amer. **47**, 895 (1957).
3. S. Mallick, Appl. Opt. **6**, 1403 (1967).
4. M. Françon and S. Mallick, in *Progress in Optics*, E. Wolf, Ed. (North-Holland, Amsterdam, 1967), Vol. 6.
5. J. B. Saunders, Appl. Opt. **6**, 1581 (1967).
6. W. H. Steel, *Interferometry* (Cambridge U.P., London, 1967).
7. M. J. Beran and G. B. Parrent, Jr., *Theory of Partial Coherence* (Prentice-Hall, Englewood Cliffs, N. J., 1964).
8. The vector $\bar{x} = x\hat{i} + yj + z\hat{k}$.

3. Optical Design and Image Evaluation

Foreword

The first reprint in this section is the proceedings of a symposium on optical image evaluation, one of twelve symposia held at the National Bureau of Standards in 1951, the fiftieth anniversary of the founding of the Bureau. It was an exciting time, a time of change, as is reflected in the many highly original contributions in the proceedings. Unfortunately, the proceedings were soon out of print and unavailable to many workers. These papers on the beginnings of modern image evaluation are often referred to, so it is a pleasure to see them once again readily available.

The other papers range from the precise measurement of the metric characteristics of aerial mapping cameras, by highly developed techniques of classical geometrical optics, to the physical optics of partial coherence in microdensitometry and the use of optical autocorrelation for the measurement of MTF.

UNITED STATES DEPARTMENT OF COMMERCE • Sinclair Weeks, *Secretary*
NATIONAL BUREAU OF STANDARDS • A. V. Astin, *Director*

Optical Image Evaluation

Proceedings of the NBS Semicentennial Symposium on Optical Image Evaluation Held at the NBS on October 18, 19, and 20, 1951

National Bureau of Standards Circular 526
Issued April 29, 1954

Foreword

The Symposium on the Evaluation of Optical Imagery was one of the twelve symposia conducted as part of the scientific program of the National Bureau of Standards in the year 1951 which marked the fiftieth anniversary of its establishment. The subject for this symposium was chosen because of its current interest and because it is one in which the Bureau has been active for many years.

In the field of applied optics it is the generally accepted practice to evaluate optical designs on the basis of geometrical optics, and the performance of optical systems has often been based upon measurements of the geometric aberrations. This practice is justified when the aberrations are so large that diffraction plays but a small part in determining the quality of imagery. Now, however, better optical systems are being produced, automatic computing machines make it feasible to completely test an optical design by computation, the interferometer enables the wave front emergent from an optical system to be completely mapped, and integrating devices enable the diffraction effects to be readily and completely determined. With these contemporary developments it seemed timely to reexamine and compare the different methods of image evaluation with the purpose of placing them on a sound engineering basis and utilizing the principles of physical optics when justified.

The scientific excellence of the symposium derived from the quality of the papers and discussions presented by the participants. The generosity of the speakers in making their material available for publication in this volume is sincerely appreciated.

Acknowledgment is made to Dr. Wallace R. Brode, Dr. H. R. J. Grosch, Dr. A. Maréchal, Dr. Stanley S. Ballard, and Dr. Brian O'Brien who served as chairmen at the different sessions. The generous cooperation of the Office of Naval Research in making this symposium possible is also gratefully acknowledged.

A. V. Astin, *Director*,
National Bureau of Standards.

Contents

Introduction

By Irvine C. Gardner

Geometric optics is one of the older branches of optics, but the parts most directly related to optical design were for a long time not developed into a continuous body of scientific knowledge. This situation arose partly because the earlier designers were expert artisans or craftsmen who did not have the same facilities for communicating with each other as did the contemporary scientists and, also, partly because methods of design were viewed as trade secrets. Optical design as we know it begins with Fraunhofer (1787–1826) because prior to the discovery of the Fraunhofer lines there could be no precise measures of index of refraction. However, it is only in the last few decades that books have attempted to explain in detail how an optical system is designed. "A System of Applied Optics" by H. Dennis Taylor, is one of the earliest notable books that attempted conscientiously to set forth a method of designing an optical system. Even now the method of designing an optical system is not a well-established regular procedure. Successful designers are very often self-taught, each to a large extent has his own methods; and consequently optical design still remains to a considerable extent an art rather than a science.

If the design of an optical system is such that large aberrations play a predominant part in determining the distribution of energy in an image spot, the problem may lie entirely within the domain of geometrical optics. On the other hand for the microscope, the astronomical telescope, or the modern airplane-camera lens with small field angle and long focus, where the distribution of energy in the image spot is largely determined by diffraction, then the final judgment of the result should involve an appeal to physical optics. Reliance upon physical optics has not been general partly because of tradition and partly because in many instances it was not practicable to make the required extensive computations.

Designing an optical system by an established procedure is an engineering task. It has been customary in requesting the design of a new optical system, such as a camera objective, to specify performance characteristics as, for example, the aperture ratio or relative aperture, the focal length, the field of view, and resolving power. However, there is no generally accepted and established method of interpreting the quality of the image, as indicated by the results of ray tracing, or the computed diffraction effects in terms of resolving power. In most instances if good resolving power is required, the designer simply perfects the design as much as possible within the limits of time and cost and hopes that the resulting lens will perform as well as required. This laissez-faire attitude has persisted even up to the present time because usually a system designed solely on the basis of geometric considerations performs better than anticipated. This occurs because of compensations that arise from the physical nature of light. Although

this is the preferable direction in which an error of judgment should lie, it will be understood that too large a factor of safety may be expected to result in optical designs that are unnecessarily complicated or may deter one from designing systems giving a more spectacular performance than is considered possible on the basis of geometric optics. These considerations illustrate the need for a more completely engineered method of image evaluation. The need for this is further indicated when it is noted that recently serious doubt has arisen as to whether or not resolving power is always a correct criterion of lens performance. As is shown by some of the articles in this volume, there are instances in which the image plane of a photographic lens that produces the most desirable photograph is not in the plane of maximum resolving power but in a plane differently defined, which may be designated in a general way as the plane of maximum contrast, or the plane that produces photographs with the steepest density gradient.

Not all of these apparent lacks should be charged against the lens designer. The design of a complicated lens system requires an extremely large program of tedious and time-consuming computations and it has in most cases not been possible to extend this work beyond the stage at which a satisfactory lens is assured. Furthermore, experimental tools, by which the distribution of energy within an image spot could be accurately determined, have not been available. The recent availability of new experimental and computational tools has greatly changed the outlook. The photomultiplier tube, with its spectacular increase of sensitivity beyond that of previous devices used for the same purpose, makes possible measurements of image quality that were previously unattainable. The programmed electronic computing machine not only makes possible much more elaborate algebraic computations or ray tracings in order to arrive at more nearly perfect designs but it also offers possibilities of making extended calculations in connection with the evaluation of an image that were previously impracticable. It is now practicable to trace skew rays, a task that the lens designer has previously avoided, so far as permissible, because of the excessive labor. It is also possible to compute diffraction patterns, thereby evaluating the image quality in terms of physical optics. Thus optical design, one of the older and traditional branches of optics, has again been brought into a nascent state.

In connection with the recognition of the fiftieth anniversary of the establishment of the National Bureau of Standards it was considered appropriate to conduct a series of symposiums dealing with scientific subjects of current interest. In view of the indicated state of flux existing in the theory of optical design it was considered desirable to select some phase of this subject, and accordingly the Symposium on Optical Image Evaluation was planned. The fact that nearly 250 people from all parts of the United States and Europe fully occupied 3 days with the formal presentation of papers and with informal discussion indicates that the choice of subject was a fortunate one and that the renewed vitality of the subject as presented in the preceeding paragraphs is generally realized.

This interest in optical image evaluation is abundantly justified by its importance and complexity. In order to show briefly the different possible courses that image evaluation may take, reference may be made to figure 1 where two main branches are shown. The evalua-

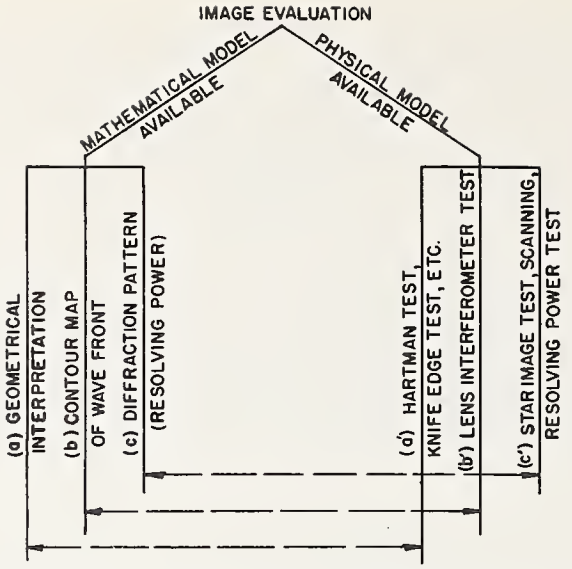


FIGURE 1. *Diagram showing relation between different methods of image evaluation*

tion may be based upon a lens design before a prototype has been actually constructed. In this case one has a constructional specification of the lens embodying such parameters as radii of curvature, thickness, separation, and index of refraction etc. Following Dr. Herzberger, this will be referred to as a mathematical model of a lens. The course of rays through the lens may be determined by computations either by the older or the newer methods. It is now practicable to trace rays sufficient in number for any desired purpose. There are at least three possible evaluations: (a) Purely geometrical interpretation may be based on the distribution of the rays in the neighborhood of an image point; (b) the computation may be made in such a manner that a contour may result showing the distance measured in wavelengths between the emergent wavefront and the desired emergent wavefront, which is usually either plane or spherical; (c) the distribution of intensity in the diffraction pattern may be computed. This gives an interpretation of the image quality in terms of physical optics and usually involves a numerical integration that can be performed by a digital computer or an integration made by a mechanical analog machine. Having determined the diffraction pattern an estimate may be made of the resolving power.

Now follow the other of the two main stems. In this instance one has a prototype or a production sample of the lens and evaluates the quality of the image by an experimental method. The experimental method (a′) most closely analogous to the method (a) of the preceding paragraph is the Hartman test and its variations or some variation of the Schlieren test by which the geometric paths of the rays are determined. The lens interferometer (method b′) may be used to give information that is directly comparable with that afforded by the method (b). According to method (c′) the star image may be photographed on a large scale to show the diffraction pattern, the distribution of energy within the diffraction pattern may be measured by one of the special scanning methods using the photomultiplier tube, or rather scanty information regarding the diffraction pattern may be

obtained by means of the resolving power test. Any of these procedures (method c′) offers results that may be compared with the results of paragraph (c).

The methods outlined permit the characteristics of the imagery to be determined. However, the evaluation, in the full sense of the word, is not completed until these measured characteristics are compared with the desired characteristics. As has been mentioned there are still differences of opinion concerning the characteristics that are desired. And when this subject is more thoroughly studied it will doubtless be found that different performance characteristics of imagery are desired for different purposes. By different purposes one refers not only to such manifest differences as those of visual instruments, photographic instruments, and projection instruments but to divisions of application that exist within these grand divisions. In photography, for example, one has airplane photography in which the object is conspicuously low in contrast, pictorial photography, and process work. In projection one has motion-picture projection where high magnification and good resolving power are required, and television projection for which the detail on the screen is necessarily limited.

This brief résumé gives an idea of the large field of engineering knowledge concerning image evaluation that remains to be filled in. The papers of this symposium touch upon most aspects of the problem in more or less detail but, like most useful scientific work, the papers also suggest the large amount of work that remains to be done.

1. The Diffraction Theory of Aberrations

By F. Zernike [1]

My starting point has been a new series development of the aberration function. Let V represent the optical path from an object point P through any optical system to its Gaussian image P', expressed as a function of the rectangular coordinates y and z of the point of intersection Q with the exit pupil (y in the axial plane). Customarily this is developed into a power series, general term $\alpha_{pq}y^p z^q$, leaving out the dependence on the axial distance of P, with which we are not specially concerned here. Of course y and z may as well be replaced by polar coordinates R and φ, making the general term $a_{kl}R^k \cos^l \varphi$. Opticians are quite accustomed to this development, so much so that specific names have been given to each of the lower terms, each being thought of as a *single aberration*. From a somewhat different problem I came in 1934 to consider a different development, rearranging the powers into polynomials, the general term $b_{nm}S_n^m(y, z)$ being homogeneous of the nth degree and also expressible as $b_{nm}R_n^m(r) \cos m\varphi$, in which $r=R/R_m$, where R_m is the radius of the circular exit pupil. These "circle polynomials" are perfectly determined by one condition, that of *orthogonality*, i. e.,

$$\int_0^1 R_n^m(r)R_{n'}^m(r)rdr=0 \qquad \text{for } n\neq n' \tag{1}$$

together with the normalizing condition that the maximum value of each shall be equal to one. Rather than give a table of these polynomials (see literature) I show their properties by figures 1.1 and 1.2, which give their course within the area of the pupil.

With the new development a *single aberration* shall mean an aberration characterized by a single term, i. e., by a circle polynomial. The remarkable advantages of this will appear presently.

Calculating the diffracted amplitude at a point in the receiving plane specified by polar coordinates ρ, ψ measured from P' as origin, one obtains, apart from irrelevant constant factors,

$$\int_0^1 R_n^m(r)rdr \int_0^{2\pi} \cos(\rho r \cos\varphi) \cos m(\varphi+\psi)d\varphi=$$
$$\cos m\psi \int_0^1 R_n^m J_m(\rho r)rdr=\pm\rho^{-1}J_{n+1}(\rho) \cos m\psi.$$

From this simple result a further development led to a new formula for the amplitude in the neighborhood of an aberrationless focus

$$u_0=\sum_{n=0}^{\infty}(-i)^n(2n+1)x^{-\frac{1}{2}}J_{n+\frac{1}{2}}(x)\rho^{-1}J_{2n+1}(\rho), \tag{2}$$

[1] Natuurkundig Laboratorium, Groningen, Netherlands.

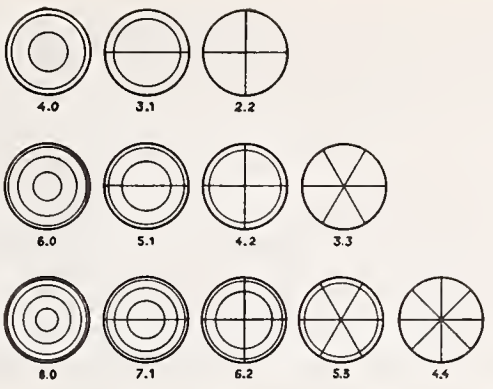


FIGURE 1.1. *Nodal lines for the complete circle polynomials.*

The first row belongs to the third order aberrations, respectively, spherical aberration, coma, astigmatism; the second row to the fifth order aberrations; the third row to the seventh-order aberrations.

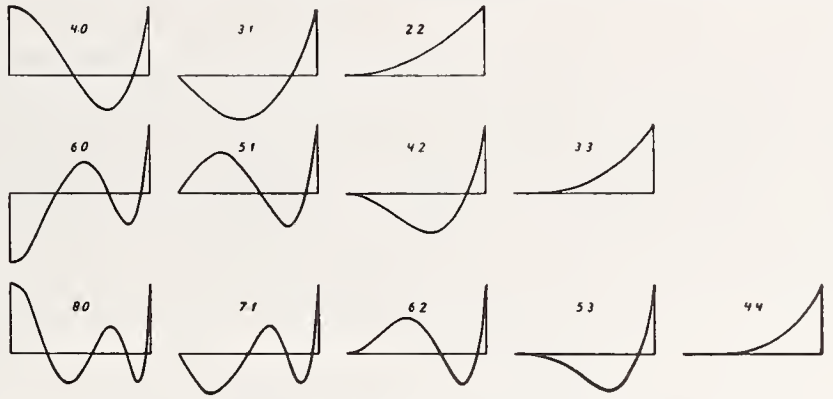


FIGURE 1.2. *Graphical representation of the radial parts, same arrangement as in figure 1.1.*

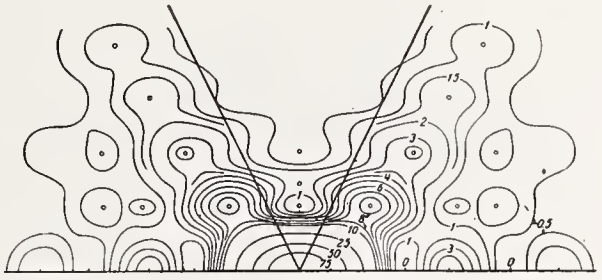


FIGURE 1.3. *Contour lines of equal intensity near an aberrationless focus.*

The numbers give the percentage of the value at the focus. The straight lines indicate the boundary of the geometric cone.

in which x measures the axial distance from the focal plane. This formula was checked numerically with those of Lommel (1886), which contain power series, different for points inside and outside the geometrical light cone. Figure 1.3 gives the resulting surfaces of equal intensity, the complete three-dimensional figure being generated by revolving the figure around the horizontal axis. A figure of this kind was first given by Berek in a rough but correct sketch from Lommel's results and afterwards repeatedly copied in the German literature with such errors that it finally appeared with marginal rays under an angle four times too small and with rectangular axes interchanged.

The next step is to introduce ordinary spherical aberration by adding a term βR_4^0 to V. Developing the diffraction integral into powers of the coefficient β, the first approximation is found to be

$$u_0 - i\beta \sum (-i)^n (2n+1) x^{-1/2} J_{n+1/2}(x) \rho^{-1} [a_n J_{2n+5} + b_n J_{2n+1} + c_n J_{2n-3}],$$

where a_n, b_n, c_n are simple expressions in n. Representing the bracket expression as the result of an operation O_4 on J_{2n+1}, the same operation must be applied repeatedly to obtain the higher terms and the final series is represented symbolically by

$$u = \sum (-i)^n (2n+1) x^{-1/2} J_{n+1/2}(x) \rho^{-1} e^{-i\beta} O_4 J_{2n+1}(\rho), \tag{3}$$

a result that is not only elegant but very useful. For higher order spherical aberrations one has only to substitute a different operator. Figures 1.4 and 1.5 give some results.

In order to judge the quality of an optical image, Strehl as early as 1895 introduced the brightness in the center of the diffraction image as compared to the same without aberrations. This is an efficient way of expressing the deterioration of the image through small aberrations by a single number, which we shall call the "Strehl Definition." For the best performance this should be 90 percent or more. Suppose there are a number of single aberrations present, with coefficients respectively, B_1, B_2, B_3, etc. The Strehl Definition SD will be

$$SD = 1 - B_1^2 - B_2^2 - B_3^2 - \ . \ . \ .,$$

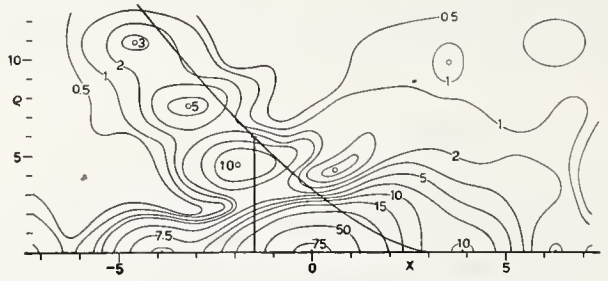


FIGURE 1.4. *Contour lines for third-order spherical aberration, $\beta = 1$, Strehl Definition 0.81.*

The geometric caustic has been inserted, also the radius of the geometric circle of greatest constriction.

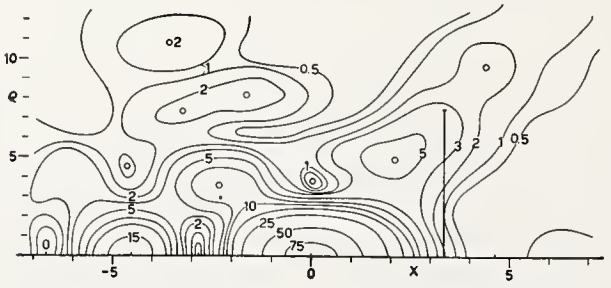


FIGURE 1.5. *Contour lines for fifth order-spherical aberration, $\beta = 1$, Strehl Definition 0.86.*

The vertical line shows location and size of the circle of greatest constriction in the case of optimum balancing according to geometric optics.

where there are no cross products because of the orthogonality. This means that each separate aberration diminishes the *SD* on its own account, uninfluenced by others. In other words, there is no use in leaving a residual of lower aberrations in order to balance a higher one. Or expressed in the old way: Each circle polynomial contains next to the highest power of r a number of lower powers which balance it in the best way. It is worth while to dwell somewhat longer on the question of balancing and on the tolerances that result from these calculations.

The quarter-wave limit was announced in 1878 by Lord Rayleigh, who was then so much ahead of his time that his result was considered for at least 50 years as unsurpassed, and worse, as self-explanatory in the following way.

The waves that come together to a single focus should reinforce each other by being in the same phase. Because of aberrations the phases will differ, but not much harm will be done if the deviations from the mean are less than one quarter, because there will be no opposite components in that case.

However, Rayleigh did much better than this; he calculated the intensity in the paraxial focus for two types of aberration and remarked that it should at least be 80 percent. He omitted the effect of refocusing, which is very marked. As an instance I take the eighth-order spherical aberration, expressed in units of $\lambda/2\pi$ by $V=\beta_8 r^8$. This cannot well be corrected, even in a high-power microscope objective, whereas fourth- and sixth-power terms can be adjusted more or less at will by the designer. If he cancels the fourth and sixth powers the decrease in *SD*, which equals the integral of V^2 over the aperture, will be

$$\int_0^1 \beta_8^2 \left(r^8-\frac{4}{5}r^2+\frac{1}{5}\right)^2 2rdr=\frac{4}{225}\beta_8^2,$$

where the constant term 1/5 appears because of the necessary adjustment of phase, and the term with r^2 because of the adjustment to best focus. The coefficients 1/5 and $-4/5$ have been determined so as to make the integral a minimum. If now a residual of r^4 aberration is introduced for balancing, the three resulting coefficients will have to be calculated anew, with r^6 and r^4 aberrations four new coefficients. The calculation is more straightforward with the circle polynomials. It is easily found that r^8 can be represented by these functions

$$r^8=\frac{1}{70}R_8+\frac{1}{10}R_6+\frac{2}{7}R_4+\frac{2}{5}R_2+\frac{1}{5}, \tag{4}$$

and the integrated square of this sum can be written down at once, as the cross products vanish, whereas R_n^2 gives $1/(n+1)$. The numerical result, brought to a common denominator, is

$$\frac{1}{210^2}(1+63+720+2352+1764). \tag{5}$$

The sum in parentheses equals 4,900 and the whole expression 1/9. However, as stated above, a constant should in all cases be added to (4) so as to minimize the integral and this clearly should be $-1/5$, thus canceling the term 1,764 in (5). This brings down (5) to 16/225

corresponding to the unchanged (paraxial) focus. By adjusting the focus, the term with R_2 can be cancelled, leaving 784 as the sum in (5), result 4/225, same as stated above. Best balancing with the fourth order will in the same way leave only 64, with fourth and sixth orders only 1. These numbers are the squares of the successive binomial coefficients of the eighth power.

If we allow a tolerance of 10-percent decrease in Strehl Definition, we obtain for the maximum value of β_8, $(16/225)\beta_8^2=0.1$ or $\beta_8=(15/4)0.316$ in the most unfavorable case, no balancing paraxial focus. This is the tolerance expressed in radians, a factor $\lambda/2\pi$ must be added to transform it to the usual linear measure. Calling the new coefficient b_8, we obtain successively

$$b_8=0.189\lambda,\ =0.378\lambda,\ =1.32\lambda,\ =10.6\lambda, \tag{6}$$

respectively for no balancing and paraxial focus, no balancing and best focus, balancing with fourth order, balancing with fourth and sixth orders.

In a sense these results may be too favorable for the fully balanced error. We have started from the power series and as an example cut this off after the r^8 term, supposing the higher terms to be negligible. Then the term b_8r^8 was changed to

$$\frac{1}{70}\,b_8R_8=\frac{1}{70}\,b_8(70r^8-140r^6+90r^4-20r^2+1).$$

It is more consistent to judge the advantage obtained from the circle polynomials by starting from a development $V=\Sigma c_{2n}R_{2n}(r)$ and cutting this off after the term with R_8. I have compared a small number of practical examples (to be treated in a later paper) that show that the ratio of the coefficients $b_8/c_8=70$ is only approximated to if the power series converges very fast; if its successive coefficients decrease in a ratio of 10:1, it is 54, if 4:1 then 35. In this last case therefore, the tolerance for b_8 would still be 5.3 λ when fully balanced.

In view of these results, which are specially favorable for spherical aberration, it would seem preferable to state the tolerance for 10-percent decrease in Strehl Definition in the following general way, applicable to all kinds of aberrations:

$$\textit{maximum rms deviation of path}=\lambda/(2\pi\sqrt{10})=0.0503\ \lambda. \tag{7}$$

It is worth while to compare these numbers with the results of balancing according to geometrical optics. Taking once more the path-aberration β_8r^8, this corresponds to a transverse aberration of the ray equal to its derivative, $8\beta_8r^7$, expressed in diffraction units (in which the first dark ring has a radius of 3.83). Now the designer will endeavor to balance this by leaving lower order aberrations, so as to obtain a minimum "circle of greatest constriction" of the beam. For this well-known problem, one should take

$$\beta_8\left(8r^7-14r^5+7r^3-\frac{7}{8}\,r\right)=\frac{1}{8}\,\beta_8T_7(r), \tag{8}$$

in which the polynomial T_n is defined by $T_n=\cos\,n\Phi$, if $\cos\,\Phi=r$.

Evidently, T_n oscillates between $+1$ and -1 in the interval $-1 \rightarrow +1$ of r. In our case, therefore, the greatest constriction has a radius of $1/8\ \beta_8$. Now, it is hardly possible to find a tolerance for this in a consistent way. The conventional view seems to have been that a disk of this kind must be superimposed on the Airy disk. It could then at most be allowed to have a radius of 2, i. e., equal to the mean radius of the diffraction disk. The tolerance would hence be

$$1/8\ \beta_8 = 2, \quad b_8 = 16\ \lambda/2\pi = 2.5\lambda \quad \text{(geometric optics)},$$

or only one-fourth of the real diffraction tolerance! The practical result of this geometric balancing is remarkably good. To calculate it, we must obtain the path-aberration by integrating (8) giving

$$\beta_8\left(r^8 - \frac{7}{3}\,r^6 + \frac{7}{4}\,r^4 - \frac{7}{16}\,r^2\right) = \beta_8\left(\frac{1}{70}\,R_8 - \frac{1}{60}\,R_6 - \frac{1}{168}\,R_4 - \ldots\right)$$

and find the integrated square of this, which is $(\beta_8/120)^2$. Hence the geometric disk of radius 2, and $\beta_8 = 16$, will in reality give a Strehl Definition of 98.2 percent or the tolerance for 90-percent definition is $b_8 = 6.0\lambda$, i. e., four-sevenths of the optimum.

Results like this, and comparable ones for the lower aberrations, explain why practical designers never had trouble to remain within tolerances, as they were apt to adopt far too strict ones. The great use made of the graphical representation of the *longitudinal* errors, initiated by von Rohr, has even led to balancing so as to make the longitudinal aberration oscillate between narrow limits. Even this gross error does little harm because of the beneficial influence of destructive interference. In short, the peculiar properties of the waves of light are very favorable to the lens designer, even if he ignores them altogether.

Coming now to the off-axis errors, I can only give a brief survey of one of these, astigmatism. The calculations were made by Nijboer, in much the same way as above. Figure 1.6 gives the contour lines of equal intensity for a small amount of astigmatism, which according to geometric optics would give rise to a "circle of least confusion" of radius 2. In reality the Airy disk proper is changed very little, the error having a marked influence only on the first and second dark rings, which are each split up into four dark spots. Figure 1.7 shows the case of a four times greater astigmatism. Here the Airy disk itself is split up into four maxima and the cross-shaped figure is very pronounced. This is the largest amount for which the series development with circle polynomials could be used, the final formula occupying a whole printed page.

Large amounts of various aberrations were studied experimentally by Nienhuis. Figure 1.8 gives one of his results. What puzzled us was that the figure did not seem to approximate to the evenly illuminated circle of geometric optics. A tentative explanation is as follows. At any point Q of the edge of the diaphragm some light will be diffracted, inward and outward, in directions perpendicular to the edge. The direct ray through Q will intersect the receiving plane in a point S on the geometric optical circle. If Q describes the edge of the pupil, S will describe the geometric circle *in the opposite direction*, as

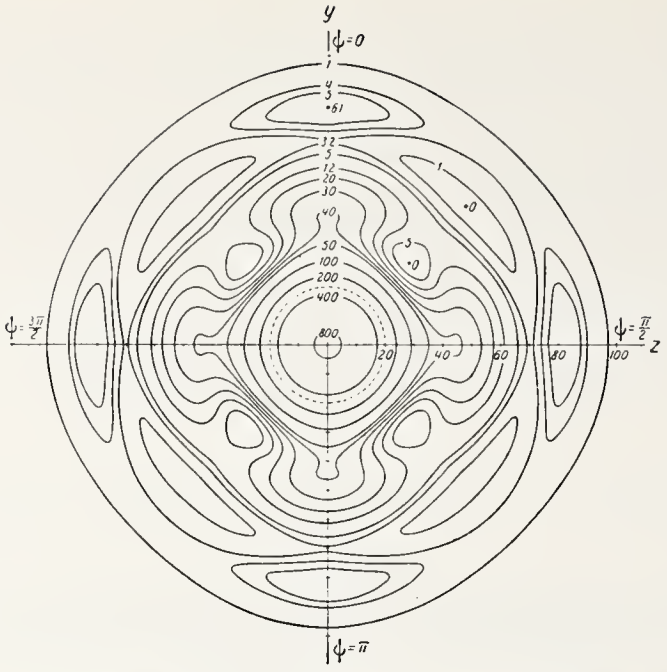


FIGURE 1.6. *Contour lines in the central plane for third-order astigmatism of amount* $\beta = 1$.

The numbers give the intensities corresponding to a value of 1,000 at the center of the Airy disk. The broken circle indicates the geometric disk.

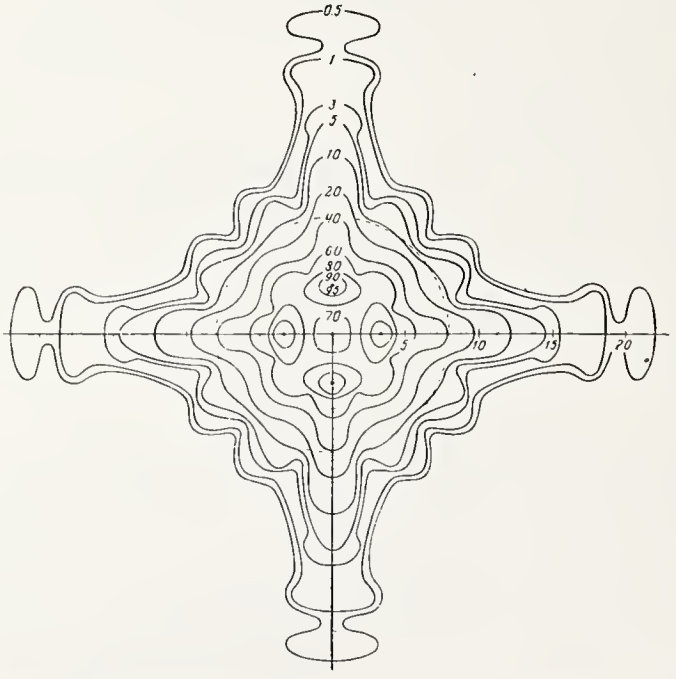


FIGURE 1.7. *Contour lines in the central plane for third-order astigmatism of amount* $\beta = 4$.

Maximum intensities of 0.095 at four off-axis spots. The broken circle should be evenly illuminated according to geometric optics.

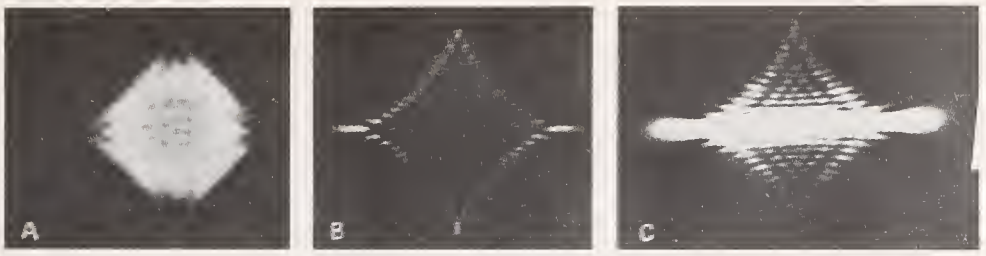


FIGURE 1.8. *Astigmatism, $\beta = 17$.*
A, complete pattern; B, edge effect alone; C, pattern at focal line.

is well known. As S moves, a narrow strip of diffracted light will therefore be seen radiating from S in a direction turning opposite to the radius through S. These strips will all be tangent to a four-pointed curve, an asteroid, which indeed appears in figure 1.8,a, as the boundary of the illuminated pattern. Nienhuis gave an experimental proof of this explanation by covering up most of the lens aperture, leaving only a narrow annulus at the margin. He thus obtained figure 1.8,b, in which the circle of least confusion has disappeared, leaving only an interference pattern caused by the overlapping of the diffracted rays. Practically the same is found in figure 1.8,c, where the geometric optical pattern has contracted into a focal line.

The last development in this matter is by N. G. van Kampen, who found asymptotic formulas for small wavelengths, which confirm these experimental results.

F. Zernike, Physica **1,** 689 (1934).
B. R. A. Nijboer, Thesis Groningen (1942). Physica **10,** 679 (1943); **13,** 605 (1947).
K. Nienhuis, Thesis Groningen (1948).
F. Zernike and B. R. A. Nijboer, Contribution in La Theorie des Images Optiques, p. 227 (Paris, 1949).
K. Nienhuis and B. R. A. Nijboer, Physica **14,** 590 (1949).
N. G. van Kampen, Physica **14,** 575 (1949); **16,** 817 (1950).
E. Wolf, Reports on Progress in Physics **14,** 95 (1951).

2. The Contrast of Optical Images and the Influence of Aberrations

By André Maréchal [1]

Introduction

The estimation of the quality of optical images is mainly based on the knowledge of their contrast, i. e., the relative variation of the illumination in the image. In most cases improving the quality means increasing the contrast. Even in the vicinity of the resolution limit, any gain in contrast is helpful because it may slightly improve the resolution.

To begin with, we shall consider the mechanism of the formation of optical images in order to determine the contrast. Then we shall compute the losses of contrast produced by the presence of small aberrations in the images of some classical objects.

Theoretical Study of the Contrast of Optical Images

The Formation of Images

We generally define contrast as being the ratio $(I-i)/i$, I and i being the maximum and minimum illumination in the image. This definition does not hold for a bright point source or a bright line source, but it is very useful in the case of a dark point or a dark line in an extended background, or in the case of a periodic structure. The distribution of light in the image of a dark line is represented by the solid curve in figure 2.1, and the distribution in the object is shown by the dotted curve.

Consider an instrument where Ox is the axis, Oyz the object plane, and $O'y'z'$ the image plane (fig. 2.2). We can refer the exit pupil to angular coordinates β,γ with origin at O'. We can now determine the laws of formation of the images and study at least the two extreme cases of coherent and incoherent illumination.

Incoherent Illumination

Suppose that every point of the object acts as an independent source (as, for instance, in the use of astronomical instruments). The image of the point source will be a diffraction pattern, sometimes influenced by the presence of aberrations, and the distribution of light in the image will be obtained by adding the illuminations produced by the different points.

Consider a point source at $O(y=z=0)$. The distribution of light in its image is given by the laws of diffraction, i. e., the Huygens-

[1] Institut d'Optique, Paris, France.

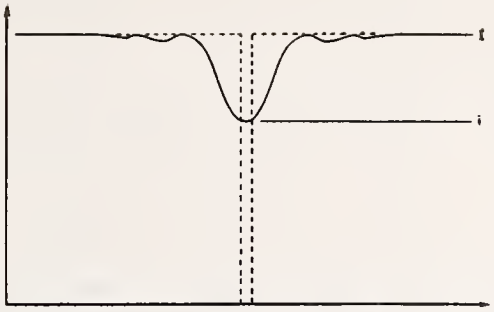


FIGURE 2.1. *Contrast reduction in the image of a dark line.*

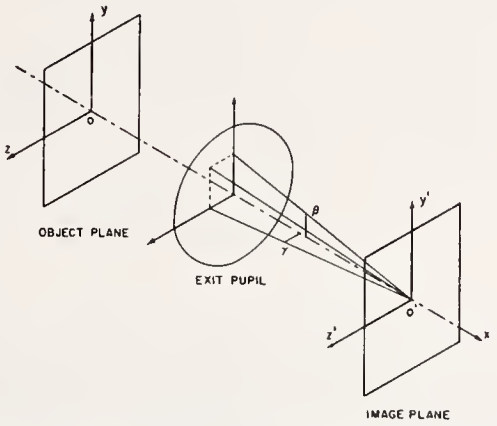


FIGURE 2.2. *Geometrical orientation of the object, lens, and image*

Fresnel principle, which can be expressed as follows: If $\Delta(\beta,\gamma)$ is the wave distortion produced by aberrations, the complex amplitude in y', z' will be

$$A(y',z')=\iint_P \exp\left[j\frac{2\pi}{\lambda}(\Delta+\beta y'+\gamma z')\right]d\beta d\gamma, \tag{1}$$

where λ is the wavelength.

The complex amplitude $A(y', z')$ appears as the Fourier transform of

$$\exp\left[j\frac{2\pi}{\lambda}\Delta\right].$$

If we denote now by

$$D(y',z')=|A(y',z')|^2$$

the distribution of light in the diffraction image, the resulting illumination for the whole object will be

$$I(y',z')=\iint_0 D(y'-y,z'-z)O(y,z)dydz, \tag{2}$$

where $O(y, z)$ represents the distribution of light in the object. Let us now consider some classical cases.

(a) *Dark point in a bright field:* $O(y, z)$ is unity everywhere, except in a small surface s around the origin O. The illumination will be

$$I(y', z') = \iint D(y'-y, z'-z)dydz - sD(y', z') \\ = \iint D(y', z')dy'dz' - sD(y', z'). \tag{3}$$

The first part is obviously a constant proportional to the total energy in the diffraction image of a point source, and we have to subtract from this constant the quantity $sD(y', z')$, as shown in figure 2.1. The contrast varies then as the function $D(y', z')$. This means that any improvement shown in it will be the same as for the case of the image of a single bright point.

(b) *Bright line:* Suppose that a bright line source of width ϵ is extended along the Oz axis. $O(y, z)$ is zero everywhere except for $0<y<\epsilon$; the illumination is then

$$I(y', z') = \epsilon \int_{-\infty}^{+\infty} D(y', z'-z)dz = \epsilon \int_{-\infty}^{+\infty} D(y', z')dz' = \epsilon S(y'), \tag{4}$$

which is a function of y' only.

If we represent by a solid the function $D(y', z')$, the value at any point y' of the illumination $\epsilon S(y')$ would be proportional to the area of the plane section of that "diffraction solid" where it is intersected by a plane through the point and perpendicular to the y' axis.

(c) *Dark line:* The illumination will be obviously complementary to the preceding

$$I(y', z') = \iint D(y', z')dy'dz' - \epsilon \int D(y', z')dz'. \tag{5}$$

(d) *Edge of a bright area:*

$$O(y, z) \begin{cases} 0 \text{ if } y<0 \\ 1 \text{ if } y>0, \end{cases}$$

$$I(y', z') = \int_0^{\infty} dy \int_{-\infty}^{+\infty} D(y'-y, z'-z)dz = \int_0^{\infty} S(y'-y)dy.$$

We will define the contrast as the slope of the curve $I(y')$: $dI/dy' = S(y')$. This signifies that *when the illumination is incoherent, the contrast of the images is related either to $D(y', z')$ when the object is a point, or to $S(y')$ when the object is a line or the edge of an area.*

Coherent Illumination, Phase Contrast

Now suppose that the vibrations from the object are no longer independent of one another, but that they are related, the complex amplitude being a function $F_o(y, z)$. To determine the vibrations in the pupil, we have to compute in any direction from the object the resulting amplitude by applying the Huygens-Fresnel principle. The

mathematical expression is the transformation giving the amplitude in the pupil.

$$G(\beta,\gamma)=\iint_o F_o(y,\,z)\,\exp\left[-j\frac{2\pi}{\lambda}\,(\beta y+\gamma z)\right]dydz,$$

and the amplitude in the image will be

$$F_i\,(y',z')=\iint_P G(\beta,\,\gamma)\,\exp\left[j\,\frac{2\pi}{\lambda}\,(\Delta+\beta y'+\gamma\,z')\right]d\beta d\gamma$$

$$=\iint_P \exp\left[j\,\frac{2\pi}{\lambda}\,(\Delta+\beta y'+\gamma z')\right]d\beta d\gamma$$

$$\iint_o F_o(y,\,z)\,\exp\left[-j\,\frac{2\pi}{\lambda}\,(\beta y+\gamma z)\right]\,dydz.$$

By reversing the order for the integrations, the latter becomes

$$F_i(y',z')=\iint_o F_o(y,z)dydz\iint_P \exp\left[j\frac{2\pi}{\lambda}\left\{\Delta+\beta(y'-y)+\gamma(z'-z)\right\}\right]d\beta d\gamma \tag{6}$$

$$=\iint_o F_o(y,z)A(y'-y,z'-z)dydz.$$

This relation is very similar to (2) where the intensities are now replaced by the complex amplitudes.

Let us now examine various cases, beginning with the case of a uniformly illuminated object, in order to know the amplitude in a coherent background.

(a) *Uniform surface*: Let $F_o(y,z)\equiv 1$ everywhere;

$$F_i(y',z')=\iint A(y'-y,\,z'-z)dydz,$$

which can be written as

$$F_i(y',\,z')=\iint A(y',\,z')dy'dz'.$$

We have already noticed that $A(y',z')$ is the Fourier transform of $\exp j(2\pi/\lambda)\Delta(\beta,\gamma)$. We can write the reverse transform as follows:

$$\exp\left[j\,\frac{2\pi}{\lambda}\Delta(\beta,\gamma)\right]=\iint A(y',\,z')\,\exp\left[-j\,\frac{2\pi}{\lambda}(\beta y'+\gamma z')\right]dy'dz'. \tag{7}$$

If we choose $\beta=\gamma=0$ we obtain the value of

$$F_i(y',z')=\iint A(y',z')dy'dz'=\exp\left[j\,\frac{2\pi}{\lambda}\,\Delta(0,0)\right],$$

which is a constant over the field.

We can now use the convention that $\Delta(0,0)=0$, so that

$$F_i(y',z')\equiv 1.$$

(b) *Dark point in a coherent background:* We have to subtract from the preceding result the contribution of a point located at O and whose area will be $F_i(y',z')=1-sA(y',z')$.

The intensity will be

$$|F_i(y',z')|^2=(1-sA)(1-sA^*)\simeq 1-s(A+A^*)=1-2sR[A(y',z')], \quad (8)$$

where $R(A)$ means the real part of the complex amplitude $A(y',z')$. The contrast of the image will be expressed by $R(A(y',z'))$.

(c) *Bright coherent line:* Suppose that the line is along Oz, ϵ being its width. Then

$$F_i(y',z')=\epsilon\int A(y',z'-z)dz=\epsilon a(y'),$$

where

$$a(y')=\int A(y',z')dz'.$$

The intensity will be:

$$\epsilon^2 a(y')a^*(y'). \quad (9)$$

(d) *Dark line in a coherent background:* We must subtract the amplitude corresponding to the line from the uniform amplitude.

$$F_i(y',z')=1-\epsilon a(y').$$

The intensity will be:

$$1-2\epsilon R[a(y')],$$

and the contrast will be related to $R[a(y')]$.

(e) *Phase contrast objects:* Phase objects observed with phase contrast techniques behave like amplitude objects observed in coherent illumination. The results obtained in b, c, and d can be applied to them.

Expressions of Contrast Implying Various Pupillar Integrals (for Small Aberrations)

It is very useful to express the various quantities that govern the contrast of the image by integrations performed on the pupil. It will then be possible to study the effect of aberrations, whose nature will be given by the wave distortion $\Delta(\beta,\gamma)$. To obtain these expressions, we use general properties of the Fourier transform. Because of Professor Duffieux's work,[2] such transformations are rather easy. Furthermore, if we limit ourselves to small aberrations, we will be able to use the following expansions:

$$\exp\left[j\frac{2\pi}{\lambda}\Delta\right]=1+j\frac{2\pi\Delta}{\lambda}-\frac{2\pi^2}{\lambda^2}\Delta^2; \qquad \cos\frac{2\pi}{\lambda}\Delta=1-\frac{2\pi^2}{\lambda^2}\Delta^2.$$

[2] P. M. Duffieux, L'intégrale de Fourier et ses applications à l'Optique (Société Anonyme des Imprimerie Oberthur, Rennes, 1947).

Points

a. *Bright point or dark point in an incoherent background:* The illumination for a bright point will be

$$|A(y',z')|^2=\left|\iint \exp\left[j\frac{2\pi}{\lambda}(\Delta+\beta y'+\gamma z')\right]d\beta d\gamma\right|^2.$$

We can conventionally introduce the linear terms $\beta\gamma'+\gamma z'$ into $\Delta(\beta,\gamma)$ and limit our study to $A(0,0)$, keeping in mind that Δ has to include a linear form in β and γ.

When Δ is small, and if $d\omega=\frac{d\beta d\gamma}{\iint d\beta d\gamma}$, then

$$|A(0,0)|^2=\left[\iint\left(1+j\frac{2\pi}{\lambda}\Delta-\frac{2\pi^2}{\lambda^2}\Delta^2\right)d\omega\right]\left[\iint\left(1-j\frac{2\pi}{\lambda}\Delta-\frac{2\pi^2}{\lambda^2}\Delta^2\right)d\omega\right],$$

or

$$D(0,0)=|A(0,0)|^2=1-\frac{4\pi^2}{\lambda^2}\left[\iint\Delta^2 d\omega-\left(\iint\Delta d\omega\right)^2\right]. \quad (11)$$

b. *Dark point in a coherent background:* The real part of $A(0,0)$ will be expressed by

$$\iint\cos\frac{2\pi}{\lambda}\Delta d\beta d\gamma.$$

For small aberrations we will write

$$R(A(0,0))=1-\frac{2\pi^2}{\lambda^2}\iint\Delta^2 d\omega.$$

Lines

a. *Incoherent illumination:* We evaluate

$$\int D(0,z')dz'=\int|A(0,z')|^2dz'.$$

We notice that

$$A(0,z')=\int\exp\left[j\frac{2\pi}{\lambda}\gamma z'\right]d\gamma\int\exp\left[j\frac{2\pi}{\lambda}\Delta(\beta,\gamma)\right]d\beta$$

appears as the one-dimensional Fourier transform of

$$\varphi(\gamma)=\int\exp\left[j\frac{2\pi}{\lambda}\Delta(\beta,\gamma)\right]d\beta.$$

Applying the Plancherel theorem (which expresses mathematically the equality of the energy in the pupil and in the image when con-

sidered with two parameters) we can write

$$\int |A(0,z')|^2 dz' = \int |\varphi(\gamma)|^2 d\gamma.$$

For small aberrations this expression can be written

$$\int D(0,z')dz' = \int (\beta_2-\beta_1)^2 \left[1-\frac{4\pi^2}{\lambda^2}\left(\int \frac{\Delta^2 d\beta}{\beta_2-\beta_1} - \left(\int \frac{\Delta d\beta}{\beta_2-\beta_1}\right)^2\right)\right] d\gamma, \tag{12}$$

where $\beta_1(\gamma)$ and $\beta_2(\gamma)$ are the lower and upper limits of the pupil.

(b) *Coherent line:* We evaluate $a(y') = \int A(y', z')\, dz'$.

Using eq 7 and making $\gamma = 0$,

$$\exp\, j\frac{2\pi}{\lambda}\Delta(\beta,0) = \int \exp\left[j\frac{2\pi}{\lambda}\beta y'\right] dy' \int A(y',z')dz',$$

which shows that a(y') is the Fourier transform of

$$\exp\left[j\frac{2\pi}{\lambda}\Delta(\beta,0)\right].$$

The reverse transformation can be written

$$a(y') = \int \exp\left[j\frac{2\pi}{\lambda}\Delta(\beta,0)\right] \exp\left[-j\frac{2\pi}{\lambda}\beta y'\right] d\beta,$$

in which we will take $y' = 0$

$$a(0) = \int \exp\left[j\frac{2\pi}{\lambda}\Delta(\beta,0)\right] d\beta. \tag{13}$$

According to the case to be studied (bright or dark line) we will use one of the two expressions:

$$|a(0)|^2 = \int \exp\left[j\frac{2\pi}{\lambda}\Delta(\beta,0)\right] d\beta \int \exp\left[-j\frac{2\pi}{\lambda}\Delta(\beta,0)\right] d\beta,$$

$$R\,[a(0)] = \int \cos\frac{2\pi\Delta(\beta,0)}{\lambda}\, d\beta.$$

For small aberrations these expressions can be written

$$|a(0)|^2 = (\beta_2-\beta_1)^2\left[1-\frac{4\pi^2}{\lambda^2}\left\{\int \frac{\Delta^2(\beta,0)}{\beta_2-\beta_1}\, d\beta - \left(\int \frac{\Delta(\beta,0)}{\beta_2-\beta_1}\, d\beta\right)^2\right\}\right], \tag{14}$$

$$R\,[a(0)] = (\beta_2-\beta_1)\left[1-\frac{2\pi^2}{\lambda^2}\int \frac{\Delta^2(\beta,0)}{\beta_2-\beta_1}\, d\beta\right]. \tag{15}$$

Application to Determination of Tolerances for Various Aberrations

If the pupil is circular, polar coordinates may be useful:

$$\beta=\alpha\cos\phi$$

$$\gamma=\alpha\sin\phi.$$

Let h be the relative aperture α/α_m Then we may write the distortion of the wave front as follows:

$$\Delta=Dh^2+Sh^4+S'h^6+(kh+ch^3+c'h^5)\cos(\phi-\phi_0)+Ah^2\cos 2(\phi-\phi_0),$$

in which D is the defect caused by errors in focusing, S is the third-order spherical aberration, S' is the fifth-order spherical aberration, k represents the linear expression in β and γ to be introduced in Δ (lateral displacement of the image in the ϕ_0 direction), C is the third-order coma in the ϕ_0 direction, C' is the fifth-order coma in the ϕ_0 direction, A is the astigmatism in the ϕ_0 direction, and h will vary from 0 to 1.

The results of the integrations are quadratic polynomials with respect to the preceding coefficients. The relative losses of contrast are the following:

INCOHERENT ILLUMINATION

(a) *Point:*

$$\frac{4\pi^2}{\lambda^2}\left[\frac{D^2}{12}+\frac{DS}{6}+\frac{4}{45}S^2+\frac{3}{20}DS'+\frac{1}{6}SS'+\frac{9}{112}S'^2+\frac{k^2}{4}+\frac{kc}{3}+\frac{c^2}{8}+\frac{kc'}{4}+\frac{cc'}{5}+\frac{c'^2}{12}+\frac{A^2}{6}\right].$$

(b) *Line:*

$$\frac{3\pi^2}{4\lambda^2}[0.33962S^2+0.30542S'^2+2\times0.31992DS+2\times0.28474DS'+2\times0.31691SS'$$
$$+\sin^2\phi(0.016549c^2+0.038678cc'+0.023919c'^2)$$
$$+\cos^2\phi(0.65160k^2+0.85639kc+0.63653kc'+0.31827c^2+0.50607cc'+0.20991c'^2)$$
$$+0.32508(A\cos 2\phi+D)^2+0.812A^2\sin^2 2\phi].$$

COHERENT ILLUMINATION

(a) *Point:*

$$\frac{2\pi^2}{\lambda_2}\left[\frac{D^3}{3}+\frac{SD}{2}+\frac{S^2+2DS'}{5}+\frac{SS'}{3}+\frac{S'^2}{7}+\frac{k^2}{4}+\frac{kc}{3}+\frac{kc'}{4}+\frac{c^2}{8}+\frac{cc'}{5}+\frac{c'^2}{12}+\frac{A^2}{12}\right].$$

(b) *Dark line:*

$$\frac{2\pi^2}{\lambda^2}\left[\frac{(D+A\cos 2\phi)^2}{5}+\frac{S^2}{9}+\frac{S'^2}{13}+\frac{2S(D+A\cos 2\phi)}{7}+\frac{2S'(D+A\cos 2\phi)}{9}\right.$$
$$\left.+\frac{2SS'}{11}+\cos^2\phi\left(\frac{c^2}{7}+\frac{c'^2}{11}+\frac{k^2}{3}+\frac{2cc'}{9}+\frac{2kc}{5}+\frac{2kc'}{7}\right)\right].$$

(c) *Bright line:*

$$\frac{4\pi^2}{\lambda^2}\left[\frac{4(D+A\cos 2\phi)^2}{45}+\frac{16S(D+A\cos 2\phi)}{15\cdot 7}+\frac{16S^2}{9.25}+\frac{8S'(D+A\cos 2\phi)}{9\cdot 7}\right.$$
$$\left.+\frac{48SS'}{5\cdot 7\cdot 11}+\frac{36S'^2}{13\cdot 49}+\cos^2\phi\left(\frac{k^2}{3}+\frac{2kc}{5}+\frac{c^2}{7}+\frac{2kc'}{7}+\frac{2cc'}{9}+\frac{c'^2}{11}\right)\right].$$

Various problems can now be solved. We have studied the following ones:

a. The effect of a given aberration alone, and the determination of the maximum value of that aberration such that the relative loss of contrast does not exceed the conventional value of 0.2 (average loss of contrast for a point in the case of the Rayleigh limit). If, for instance, we consider the effect of third-order spherical aberration alone (S term) on the contrast of an incoherent line, we write

$$\frac{3\pi^2}{4\lambda^2}\,0.339S^2\leq 0.2 \text{ or } |S|\leq 0.28\lambda.$$

b. For high-order aberrations we have to study the best way of balancing the high-order component with lower order terms. For example, in the case of "corrected" spherical aberration and its effect on the contrast of a coherent bright line, we write

$$\frac{4\pi^2}{\lambda^2}\left[\frac{4}{45}\,D^2+\frac{16}{15\times 7}\,DS+\cdots\right]\leq 0.2,$$

from which we can determine the proper values of D and S leading to the minimum value of the polynomial. These values are $D=(5/11)S'$ and $S=-(15/11)S'$. If we compute the minimum as a function of S' we obtain $S'<3.7\lambda$. The relation $S=-(15/11)S'$ determines the aperture of correction of the curve of longitudinal spherical aberration (fig. 2.3). By using the relations between longitudinal and wave spherical aberration, it can be shown that the aperture h_0 of correction will be $\sqrt{v_0}=\sqrt{2/3\cdot S/S'}$ times the maximum aperture h_m (v_0 is the parameter $(h_0/h_m)^2$).

Let us now consider the case of coma. As is well known, the distortion of the wave front is: Image magnitude times angular aperture times (offense against sine condition minus (longitudinal spherical aberration/distance pupil image). It is interesting to know what maximum value of the fifth-order term will be tolerable and what optimum aperture of correction is to be chosen. In the case of the

TABLE 2.1. *Values of various aberrations leading to a loss of contrast of 0.2*

			Incoherent illumination			Coherent illumination				
			Point	Line		Point	Dark line		Bright line	
Defect of focusing	D	$\leq$	0.25 λ	0.29 λ		0.17 λ	0.22 λ		0.24 λ	
Third-order spherical aberrations: alone	S	$\leq$	0.25 λ	0.28 λ		0.22 λ	0.30 λ		0.27 λ	
Third-order spherical aberrations: plus defect of focusing	D	$\equiv$	$-1.00\,S$	$-0.98\,S$		$-0.75\,S$	$-0.71\,S$		$-0.86\,S$	
	S	$\leq$	0.95 λ	1.04 λ		0.90 λ	1.00 λ		0.92 λ	
Third- plus fifth-order spherical aberrations, plus defect of focusing	v_0	$\equiv$	1.00	1.00		0.89	0.85		0.91	
	S'	$\leq$	3.75 λ	4.00 λ		4.00 λ	4.50 λ		3.70 λ	
Astigmatism: $\phi=0$ or $\phi=\pi/2$	A	$\leq$	0.17 λ	----------		0.25 λ	----------		----------	
Astigmatism: $\phi=0$ or $\phi=\pi/2$	$D\pm A$	$\leq$	----------	0.29 λ		----------	0.22 λ		0.24 λ	
Astigmatism: $\phi=\pi/4$	A	$\leq$	0.17 λ	0.18 λ		0.25 λ	----------		----------	
				$\phi=0$	$\phi=\pi/2$		$\phi=0$	$\phi=\pi/2$	$\phi=0$	$\phi=\pi/2$
Third-order coma: alone	C	$\leq$	0.20 λ	0.20 λ	0.86 λ	0.28 λ	0.27 λ	∞	0.19 λ	∞
Third-order coma: plus lateral displacement	k	$\equiv$	$-0.67\,C$	$-0.67\,C$	?	$-0.67\,C$	$-0.60\,C$	?	$-0.60\,C$	?
	C	$\leq$	0.60 λ	0.58 λ	0.86 λ	0.85 λ	0.66 λ	∞	0.47 λ	∞
Fifth-order plus third-order coma plus lateral displacement	v_0	$\equiv$	1.20	1.20	1.17	1.20	1.11	?	1.11	?
	C'	$\leq$	2.50 λ	2.40 λ	3.10 λ	3.50 λ	2.70 λ	∞	1.85 λ	∞

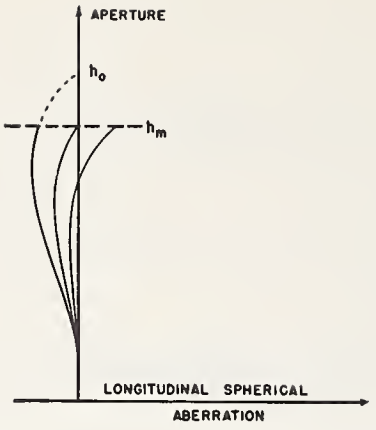


FIGURE 2.3. *Aperture of correction of longitudinal spherical aberration.*

coherent bright line perpendicular to the comatic flare ($\varphi_0=0$) we write

$$\frac{4\pi^2}{\lambda^2}\left(\frac{k^2}{3}+\frac{2kc}{5}+\frac{c^2}{7}+\frac{2kc'}{7}+\frac{2cc'}{9}+\frac{c'^2}{11}\right)<0.2.$$

The minimum value of the polynomial is obtained when $k=(5/21)C'$ and $C=-(10/9)C'(v_0=10/9)$, so that the optimum aperture of correction will be $h_0=(\sqrt{10}/3)h_m$ and the tolerance will be $C'=1.85\lambda$. The results are collected in table 2.1. In the case of coma we have listed the tolerance for the two possible directions of the comatic flare (when $\varphi_0=0$ the flare is perpendicular to the object). In coherent illumination the coma has no effect when the flare is parallel to the line and the tolerance becomes infinite. In the case of astigmatism the best focus is located midway between the focal lines if the object is a point, and the tolerances are given for that position ($D=0$). When the linear object is parallel to one of the focal lines, the best focus is located on that line and then the astigmatism has no effect. The maximum displacement with respect to that line is given by the tolerance for errors in focusing.

Let us now take an example. Suppose that the amount of astigmatism in an instrument is such that $A=\lambda/8$ (the distortion of the wavefront being $\pm 0.125\lambda$ on both sides of the reference sphere; altogether (0.25λ). The losses of contrast will be, for a dark point in a bright field, in incoherent illumination $0.2(0.125/0.17)^2=0.11$ and in coherent illumination $0.2(0.125/0.25)^2=0.05$. The losses for an incoherent line would be (midway between the two focal lines) when the line is parallel to a focal line, $0.2(0.125/0.2)^2=0.04$ and, when the line makes a 45° angle with both focal lines, $0.2(0.125/0.18)^2=0.10$. The rates of losses of contrast can vary on a large scale (here the ratio of the extreme cases is about 3).

The Mechanical Computation of Diffraction Patterns

When the aberrations are either very small or very large, it is possible to compute the distribution of energy in diffraction patterns by using mathematical expansions.[3] We have already shown that the

[3] See paper 1, page 1. in this Symposium.

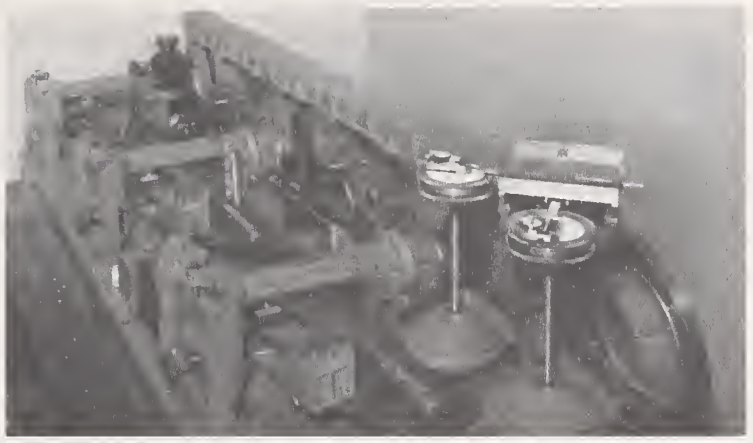


FIGURE 2.4. *Mechanical integrating device for computing diffraction patterns.*

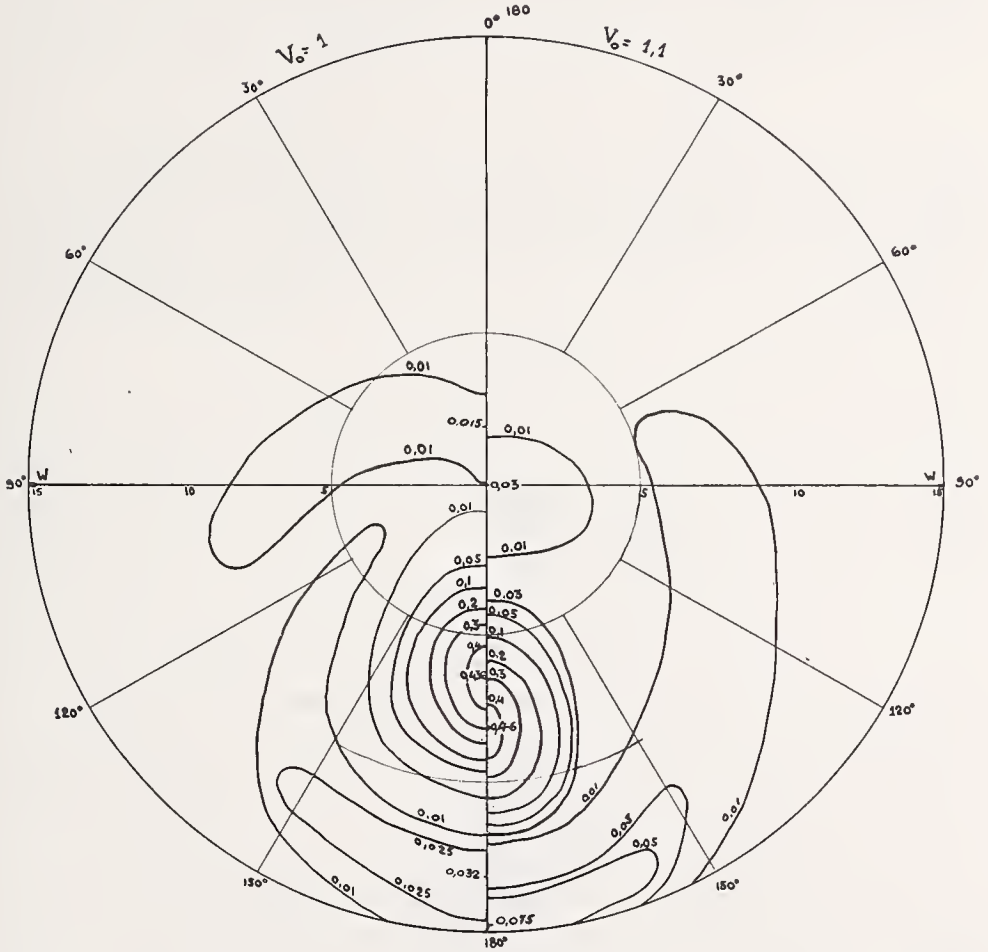


FIGURE 2.5. *Repartition of energy represented by isophote curves, in the presence of "corrected" coma, balanced 3d÷5th order, for four values of the aperture of correction* $(h_o/h_m) = 1^{1}/_{4}, 1.1^{1}/_{4}, 1.2^{1}/_{4}, 1.3^{1}/_{4}$.

losses of contrast can be expressed as functions of various pupillar integrals that can also be easily expressed in the case of small aberrations. When the aberrations are neither very small nor very large (in the transition between diffraction and geometrical optics) these mathematical procedures may fail. It is then useful to perform the integration expressed in eq 1 for a point source by means of a special mechanical device (fig. 2.4).

The principle of such a machine has already been described elsewhere.[4] The real and imaginary parts of the complete integral are developed by two integrating wheels, whereas the distortions of the wavefront (of any order) are given by cams, amplifier levers, adding tape, etc. The machine has been used for solving the following problems:

a. Distribution of energy in the presence of third-order aberrations. The goal of that study was mainly the knowledge of the transition between diffraction patterns and geometrical caustics. It seems that the transition is very rapid in the case of astigmatism when the focus-

[4] A. Maréchal, Thesis (Paris, 1947); Rev. opt. 26, 257–277 (1947), 27, 73–92, 269–287 (1948); J. Opt. Soc. Am. 37, 982 (1946).

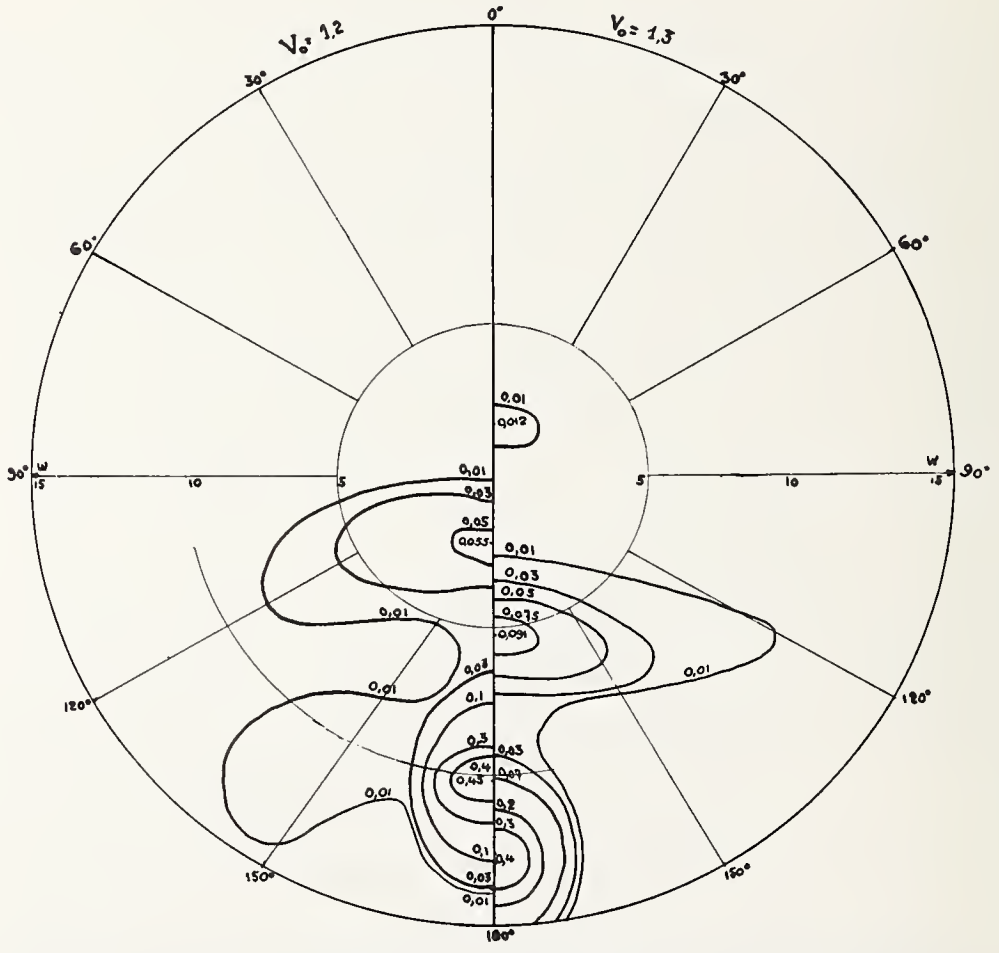


FIGURE 2.5.—Continued.

ing is done on a focal line. The diffraction pattern is already similar to the geometrical pattern even when the aberration is as small as $\lambda/4$. The transition is more gradual for coma, and still more gradual for spherical aberration.

b. Determination of the proper aperture of correction of geometrical aberrations when the aberrations are of the order of two or three times the Rayleigh limit.

c Special studies of practical cases of microscope objectives, Schmidt cameras, etc. The results can be represented by curves of equal illumination in the image plane.

Figure 2.5 shows the example of the study of comatic flares in the presence of corrected coma. The distortion of the wavefront is supposed to be conveniently represented by the superposition of third- and fifth-order terms; the fifth-order term is 5 λ (twice the tolerance for a bright-point source. See table 2.1); the third-order term is varied by steps as well as the aperture of correction $h_0=v_0^{1/2}h_m$; the maximum illumination in the flare is 0.436, 0.46, 0.45, 0.40 for $v_0=1$, 1.1, 1.2, 1.3 so that the optimum aperture of correction is about $1.13^{1/2}$, whereas for a smaller aberration the maximum would be obtained for $1.2^{1/2}$. As a consequence the aperture of correction has to be decreased with increasing aberration. Extensive results will be published in the near future.

The present work has been performed in close collaboration with G. Pieuchard, to whom I express my best thanks. W. G. Steel, working on a generalization of the present work to the case where the pupil involves a central obscuration, has been able to test many of the results obtained.

3. Diffraction Images Produced by Fully Corrected Objectives of High Numerical Aperture

By Harold Osterberg and Robert A. McDonald [1]

Introduction

The theory of diffraction by optical systems has been extended by R. K. Luneberg [1] [2] to include the effect upon the diffraction image of the geometrical coordination of the rays in the object and image space. An explicit form of the primary diffraction integral has been derived from Luneberg's formulation by Osterberg and Wilkins [2] upon the supposition that the objective is free of spherical aberration and coma. The main purposes of the present paper are to integrate and to discuss the primary diffraction integral under the additional supposition that absorption and reflection losses within the objective are negligible. A universal law governing diffraction phenomena with fully corrected objectives can be obtained only by ignoring the variable loss of light in the axial bundle of rays. We shall see that the universal law thus obtained differs in several interesting and important respects from the accepted classical laws governing idealized diffraction images.

Amplitude Variation Over the Converging Wavefront

Wavefronts expand (fig. 3.1. and 3.2) from an unresolvably small area centered about the axial point O and are converged without spherical aberration toward the conjugate point O' of the image space. Let

$$\rho \equiv \sin \vartheta; \qquad \rho_m = \sin \vartheta_m;$$

$$\rho_o \equiv \sin \vartheta_o; \quad \rho_{om} = \sin \vartheta_{om}. \tag{1}$$

The numerical apertures of the objective with respect to its object and image space are $n\rho_m$ and $n_o\rho_{om}$, respectively, where n_o and n are the refractive indices of the object and image space, respectively.

If Abbe's sine condition is obeyed,

$$nM\rho = n_o\rho_o, \tag{2}$$

where M is the magnification ratio. Accordingly,

$$nM\rho_m = n_o\rho_{om} \equiv N.\,A., \tag{3}$$

where $N.\,A.$ denotes the numerical aperture of the system.

[1] American Optical Co., Research Laboratory, Stamford, Conn.
[2] Figures in brackets indicate the literature references on p. 35.

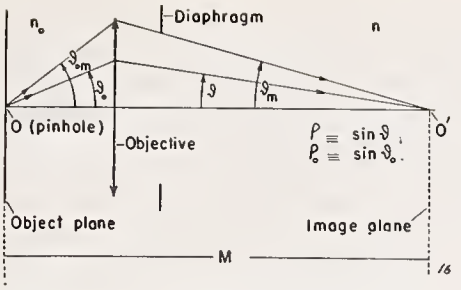


FIGURE 3.1. *Convention with respect to the axial bundle of rays.*

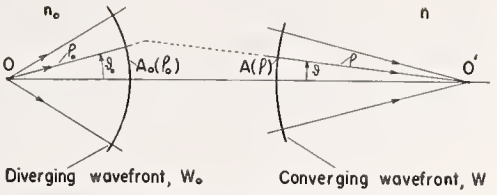


FIGURE 3.2. *The amplitudes of A_o and A of the diverging and converging wavefronts, respectively.*

When the rays of the axial bundle obey (eq 2) and when these rays are transmitted equally from O to O' (fig. 3.2) it can be shown [3] that except for an unimportant factor of proportionality

$$A(\rho)/A_o(\rho_o) \equiv k(\rho) = (1-\rho^2)^{1/4}/(1-\rho_o^2)^{1/4}, \tag{4}$$

in which $A_o(\rho_o)$ and A (ρ) represent amplitudes on the expanding wavefront W_o and on the converging wavefront W, respectively. For unpolarized light one may set

$$A_o(\rho_o) = 1 \tag{5}$$

and

$$A(\rho) = k(\rho) = (1-\rho^2)^{1/4}/(1-\rho_o^2)^{1/4}. \tag{6}$$

Microscope objectives have the property $|M|>1$ and hence that $\rho_o^2 > \rho^2$. Accordingly, we observe from eq 6 that the amplitude $A(\rho)$ increases toward the outer portions of the converging wavefront. This increase becomes substantial when ρ_o can approach unity. The upper practical value of $\rho_{om} = N.A./n_o$ falls in the neighborhood of 0.95. The corresponding ratio $K(\rho_m)$ is approximately 1.8.

On the other hand, camera and telescope objectives have the property $|M|<1$ and hence that $\rho_o^2 < \rho^2$. It follows from eq 6 that the amplitude $A(\rho)$ decreases toward the outer portions of the converging wavefront. This amplitude variation is ordinarily negligible with telescope objectives because they have low numerical aperture with respect to both object and image space.

Let $T(\rho)$ denote the amplitude transmission of the system for rays in the axial bundle. Reflection losses render $T(\rho)$ a decreasing function of ρ. The functions $T(\rho)$ and $k(\rho)$ counteract with microscope objectives but conjoin with camera objectives.

The Primary Diffraction Integral

The primary diffraction integral $U(r)$ for fully corrected objectives obeys [4] the relation

$$U(r) = 2\pi \int_0^{\rho_m} P(\rho) J_0(2\pi r n \rho) \rho d\rho; \tag{7}$$

$$P(\rho) \equiv \frac{A(\rho)T(\rho)}{(1-\rho^2)^{1/2}}. \tag{8}$$

Here, r is distance in number of wavelengths from the diffraction head (center of the diffraction image of an unresolved pinhole), $P(\rho)$ is the pupil function, J_0 is a Bessel function of the first kind and zero order, $A(\rho)$ is given by eq 6, and $T(\rho)$ is the amplitude transmission of the axial bundle of rays. For *lossless* objectives $T(\rho)=1$. The energy density in the diffraction image of a single, unresolvably small pinhole in an opaque slide is proportional to $U^2(r)$.

By combining eq 6, 7 and 8, we obtain explicitly

$$U(r)=2\pi\int_0^{\rho_m}\frac{T(\rho)J_0(2\pi rn\rho)\rho d\rho}{(1-\rho^2)^{1/4}(1-\rho_o^2)^{1/4}} \tag{9}$$

in which ρ and ρ_o obey eq 2.

Airy's diffraction integral is obtained by setting $P(\rho)=1$ in eq 7. The familiar result is

$$U(r)=2\pi\int_0^{\rho_m}J_0(2\pi rn\rho)\rho d\rho=2\pi\rho_m^2\frac{J_1(2\pi rn\rho_m)}{2\pi rn\rho_m}, \tag{10}$$

in which J_1 is a Bessel function of the first kind and first order.

Debye's diffraction integral is obtained by setting $A(\rho)T(\rho)=1$ in eq 8. It is

$$U(r)=2\pi\int_0^{\rho_m}\frac{J_0(2\pi rn\rho)\rho d\rho}{(1-\rho^2)^{1/2}}. \tag{11}$$

The solutions of eq 10 and 11 are practically alike for small ρ_m. However [5], with respect to eq 11

$$\operatorname*{Limit}_{\rho_m=1}U(r)=2\pi\frac{\sin(2\pi nr)}{2\pi nr}. \tag{12}$$

From a physical viewpoint, Debye's diffraction integral does not generate a single type of diffraction image.

Reversibility of the Primary Diffraction Integral for Fully Corrected Objectives

An objective and its primary diffraction integral are said to be reversible with respect to a pair of conjugate object and image planes when the diffraction images formed by interchanging the object point O and the conjugate image point O', figure 3.1, have similar distributions of energy density. Airy's diffraction integral of eq 10 is reversible. Debye's diffraction integral of eq 11 is not reversible.

The reversibility of the diffraction integral of eq 9 will now be demonstrated for the case $T(\rho)=1$, that is, for lossless objectives.

Let r_a denote Airy's limit of resolution measured as number of wavelengths in the image space. Then

$$2\pi r_a n\rho_m=3.8317\equiv\beta, \tag{13}$$

the first root of $J_1(2\pi r_a n\rho_m)=0$. Let x be distance measured in

Airy units r_a from the diffraction head to other points in the diffraction image.

$$x \equiv r/r_a. \tag{14}$$

Let also

$$w \equiv \rho/\rho_m. \tag{15}$$

By introducing the changes of variable of eq 13 to 15 together with ρ_o from eq 2 into eq 9, we obtain the primary diffraction integral for fully corrected, lossless objectives in the universal form

$$U(x) = 2\pi\rho_m^2 \int_0^1 \frac{J_0(\beta x w) w dw}{(1-\rho_m^2 w^2)^{1/4}\left[1-\left(\frac{nM\rho}{n_o}w\right)^2\right]^{1/4}}. \tag{16}$$

$U^2(x)$ describes the distribution of energy density produced about point O', figure 3.1, by the light radiated from an unresolved pinhole centered on the object point O.

If the object point O and the conjugate image point O' are interchanged,

$$U'(x) = 2\pi\rho_m'^2 \int_0^1 \frac{J_0(\beta x' w) w dw}{(1-\rho_m'^2 w^2)^{1/4}\left[1-\left(\frac{n'M'\rho_m'}{n_o'}w\right)^2\right]^{1/4}}, \tag{17}$$

in which the primed quantities refer to the new image space (formerly the object space). The equations of transformation from one space to the other are

$$n' = n_o; \quad n_o' = n; \quad M' = 1/M; \tag{18}$$

$$\rho_m' = \rho_{om} = nM\rho_m/n_o. \tag{19}$$

Therefore,

$$n'M'/n_o' = n_o/nM; \tag{20}$$

$$r_a' = 0.6098/n'\rho_m' = 0.6098/n_o\rho_{om} = r_a/|M|. \tag{21}$$

By introducing eq 18 to 20 into eq 17, one obtains in a straightforward manner the result

$$U'(x') = 2\pi\rho_m'^2 \int_0^1 \frac{J_o(\beta x' w) w dw}{\left[1-\left(\frac{nM\rho_m}{n_o}w\right)^2\right]^{1/4}(1-\rho_m^2 w^2)^{1/4}}, \tag{22}$$

where x' is distance measured in Airy units r_a' in the new image space.

Comparison of the integrals of eq 16 and 22 shows that they are similar with respect to the role of x and x'. The primary diffraction integral of eq 16 is therefore reversible.

It follows from the reversibility of the primary diffraction integral that the diffraction images produced by a fully corrected objective operated as a microscope or as a camera objective between a fixed pair of conjugate planes are similar. This similarity holds in spite of the fact that the amplitude variation on the convergent wavefront is an increasing function of ρ, figure 3.2, with microscope objectives but a decreasing function of ρ with camera objectives.

Repetition of the above argument relative to eq 16 and 22 shows that the more general primary diffraction integral of eq 9 is reversible

provided that the amplitude transmission $T(\rho)$ for the axial bundle is the same whether the light passes from O to O' or from O' to O in figure 3.1. Functions $T(\rho)$ having this property will be called reversible. If, therefore, the reflection and absorption losses within a fully corrected objective lead to reversible amplitude transmission functions $T(\rho)$, the corresponding primary diffraction integral is reversible.

An Approximation with Respect to Microscope Objectives

In integrating eq 9 for microscope objectives, one may set

$$(1-\rho^2)^{-1/4}=1 \tag{23}$$

because $0\leqq\rho\leqq\rho_m$ with $\rho_m<0.04$. The primary diffraction integral for fully corrected microscope objectives may therefore be integrated in the alternate approximate forms

$$U(r)=2\pi\int_0^{\rho_m}\frac{J_0(2\pi r\rho)\rho d\rho}{(1-\rho_o^2)^{1/4}}, \tag{24}$$

$$U(x)=2\pi\rho_m^2\int_0^1\frac{J_0(\beta xw)wdw}{(1-N^2w^2)^{1/4}}, \tag{25}$$

in which $n=1$ and

$$\rho_o=nM\rho/n_o=M\rho/n_o; \tag{26}$$

$$N\equiv\frac{|M|\rho_m}{n_o}=\frac{N.A.}{n_o}. \tag{27}$$

$$r_a\equiv 0.6098/N.A. \tag{28}$$

in number of wavelengths and

$$x\equiv r/r_a. \tag{29}$$

Phenomena at the Diffraction Head

The diffraction head corresponds to the center of the diffraction image of a single particle, that is, to the point $r=x=0$ in eq 24 and 25. From eq 25

$$\frac{U(0)}{\pi\rho_m^2}=2\int_0^1\frac{wdw}{(1-N^2w^2)^{1/4}};$$

$$=\frac{4}{3N^2}[1-(1-N^2)^{3/4}]. \tag{30}$$

N is defined by eq 27.

The energy density $G(0)$ at the diffraction head is given by $U^2(0)$ so that

$$\frac{G(0)}{\pi^2\rho_m^2}=\frac{16}{9}\frac{[1-(1-N^2)^{3/4}]^2}{N^4} \tag{31}$$

$G(0)/\pi^2\rho_m^4$ is called the normalized energy density. From eq 31

$$G(0)/\pi^2\rho_m^4 \to 16/9 \quad \text{as} \quad N \to 1; \tag{31, a}$$

$$G(0)/\pi^2\rho_m^4 \to 1 \quad \text{as} \quad N \to 0; \tag{31, b}$$

$$\therefore 1 \leqq G(0)\pi^2\rho_m^4 \leqq 16/9. \tag{31, c}$$

The usual classical value of $G(0)/\pi^2\rho_m^4$ is obtained from Airy's diffraction integral of eq 10 and is unity. We observe that the normalized energy density at the diffraction head can exceed the classical value appreciably.

It has been concluded by R. K. Luneberg [6] that among all primary diffraction images obeying eq 7 and producing *equal* total energy flux through the plane of the image, the primary diffraction image having the greatest normalized energy density at the diffraction head is the classical case of eq 10, that is, the case in which $P(\rho)$ is equal to a constant. This conclusion remains useful in considering the properties of diffraction integrals. However, one should not construe Luneberg's theorem to mean that the highest possible normalized energy density at the diffraction head is under *all* circumstances the classical value unity.

Primary Diffraction Images with Microscope Objectives

With respect to eq 25, let

$$F(w^2) \equiv (1-N^2w^2)^{-1/4} \tag{32}$$

be approximated by the power series

$$f(w^2) \equiv 1 + a_1w^2 + a_2w^4 + \ \ldots \ a_mw^{2m}, \tag{33}$$

in which the finite number m of coefficients a_j is chosen high enough so that $f(w^2)$ is a satisfactory approximation to the monotonic function $F(w^2)$ in the interval $0 \leqq w \leqq 1$.

In the following calculations the coefficients a_j have been determined from $f(w^2)$ and $F(w^2)$ by the method [7] of least squares.

From eq 25, 32, and 33,

$$\frac{U(x)}{2\pi\rho_m^2} = \sum_{j=0}^{m} a_j \int_0^1 J_0(\beta x w) w^{2j+1} dw. \tag{34}$$

It is convenient to introduce the change of variable $z \equiv \beta x w$ and thus obtain

$$\frac{U(x)}{2\pi\rho_m^2} = \frac{1}{\beta^2x^2} \sum_{j=0}^{m} \frac{a_j}{(\beta x)^{2j}} \int_0^{\beta x} J_0(z) z^{2j+1} dz. \tag{35}$$

Then, because

$$\int_0^{\beta x} J_0(z) z^{2j+1} dz = \sum_{\nu=0}^{j} \frac{(\beta x)^{2\nu}}{(-4)^{\nu-j}} \frac{(j!)^2}{(\nu!)^2} [\beta x J_1(\beta x) + 2\nu J_0(\beta x)], \tag{36}$$

$$\frac{U(x)}{2\pi\rho_m^2}=\frac{J_1(\beta x)}{\beta x}+\sum_{j=1}^{m}\frac{a_j}{(\beta x)^{2j}}\sum_{\nu=0}^{j}\frac{(\beta x)^{2\nu}}{(-4)^{\nu-j}}\frac{(j!)^2}{(\nu!)^2}\left[\frac{J_1(\beta x)}{\beta x}+2\nu\frac{J_0(\beta x)}{(\beta x)^2}\right]. \quad (37)$$

To calculate $U(x)$ for small x-values, we have preferred to expand $J_0(z)$ in eq 35 into the standard Bessel series and to integrate term by term. The resulting expression for $U(x)$ is

$$\frac{U(x)}{2\pi\rho_m^2}=\frac{1}{2}\sum_{j=0}^{m}a_j\sum_{\nu=0}^{\infty}\left(\frac{-1}{4}\right)^{\nu}\frac{(\beta x)^{2\nu}}{(\nu!)^2(j+\nu+1)}. \quad (38)$$

It is pointed out that the method of eq 32 to 38 can be applied to the exact integral of eq 16 by choosing

$$F(w^2)=(1-N^2w^2)^{-1/4}(1-\rho_m^2w^2)^{-1/4}. \quad (39)$$

The use of eq 39 instead of eq 32 is not justified unless one employs a higher number of coefficients than employed by the writers.

The normalized energy densities $U^2(x)/\pi^2\rho_m^4$ have been calculated from eq 37 and 38 for the cases $N\equiv N.A./n_0$ equal to 0.85, 0.92 and 0.95 and are plotted against $x=r/r_a$ as the solid curves in figures 3.4 to 3.6. The case $N=0.95$ corresponds to "dry" microscope objectives having the numerical aperture 0.95 or to oil-immersion objectives having the numerical aperture 1.44.

The number and values of the coefficients a_j are listed in table 3.1. The agreement between $F(w^2)$ and $f(w^2)$ is illustrated in figure 3.3 for the case $N=0.85$. The deviation of f from F is less than 3, 13 and 19 parts per thousand for $N=0.85$, 0.92 and 0.95, respectively. The addition of one coefficient for the case $N=0.95$ is indicated but would not, we estimate, alter appreciably the diffraction curve of figure 3.6.

TABLE 3.1 *Number and Values of the Coefficients* a_j

N	a_1	a_2	a_3	a_4
0.85	0.208353	−0.065553	0.231649	--------
.92	.206122	.203478	−.316991	0.491368
.95	.322624	−.328069	.362459	.400207

The normalized energy densities $U^2(x)/\pi^2\rho_m^4$ belonging to Airy's diffraction integral of eq 10 are included for comparison as the broken curve of figures 3.4 to 3.6. Airy's diffraction curve is unity at the diffraction head where $x=0$ and reaches its first zero at Airy's limit of resolution, where $x=1$.

The most conspicuous property of the solid curves drawn in figures 3.4 to 3.6 for fully corrected, lossless microscope objectives is the rapid increase of energy density at the diffraction head ($x=0$) with increasing N. This property has been explained on page 28.

A second interesting property of the solid curves of figures 3.4 to 3.6 is that the first zero of the energy density occurs at x-values less than unity, that is, within Airy's limit of resolution. The data from which figures 3.4 to 3.6 have been plotted show that the first zeros occur near $x=0.96$, 0.95, and 0.94 for cases $N=0.85$, 0.92, and 0.95, respectively. Let x_0 denote the first root of $U^2(x)=0$. Then with respect to the primary diffraction curves for fully corrected objectives $x_0\leqq 1$. Furthermore, x_0 is a decreasing function of N.

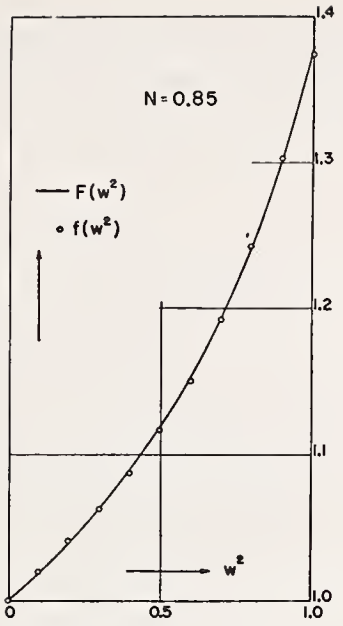


FIGURE 3.3 *The fit of $f(w^2)$ to $F(w^2)$ for the case $N=0.85$.*

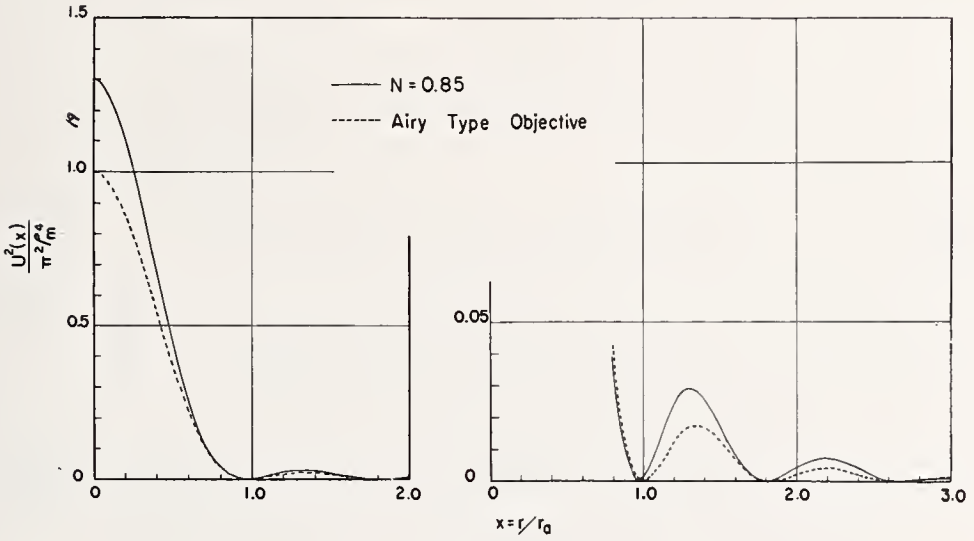


FIGURE 3.4. *The primary diffraction curve for the case $N=0.85$.*

$U^2(x)$ is proportional to the energy density in the diffraction image. Distance x from the center of the diffraction image is measured in Airy units r_a.

It is especially noteworthy that x_0 is a decreasing function of N while the energy density at the diffraction head is an increasing function of N. The energy density at the diffraction head is therefore a *decreasing* function of $x_0(N)$. The reduction of x_0 below $x_0=1$ by artificial means such as coating [8] the exit pupil of the objective is invariably accompanied by marked reduction in the energy density at the diffraction head. In other words, when one attempts to decrease x_0 artificially, he is faced by the damaging fact that the energy density at the diffraction head becomes an *increasing* function of x_0.

Figures 3.4 to 3.6 show that the ratio of the energy densities at the subsidiary maxima to the energy density at the central maximum is higher with the fully corrected objective than with Airy type objec-

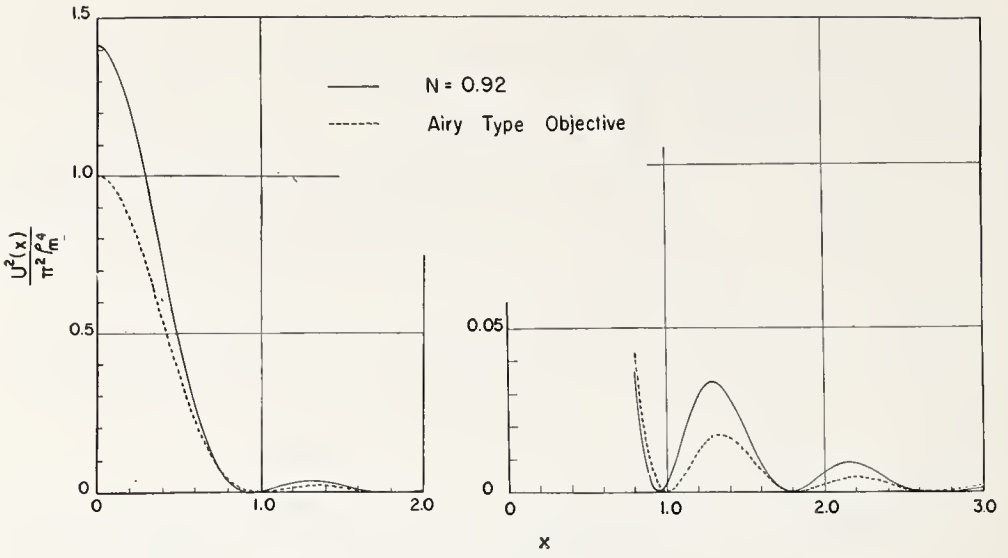


FIGURE 3.5. *The primary diffraction curve for the case $N=0.92$.*

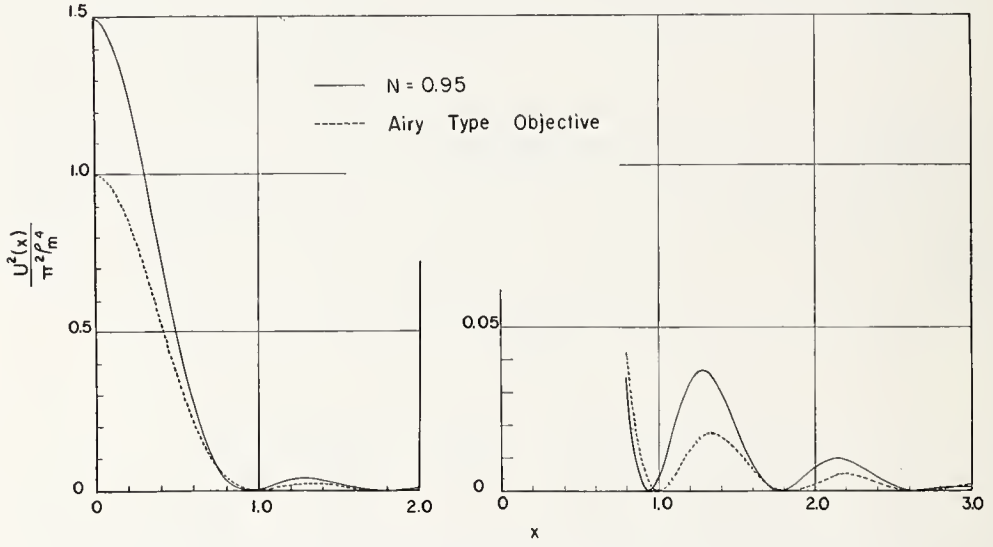


FIGURE 3.6. *The primary diffraction curve for the case $N=0.95$.*

tives (idealized objectives that obey Airy's diffraction integral). The relative increase in the brightness of the diffraction rings is however small.

Resolution of Two Like Pinholes in an Opaque Slide

We have seen from figures 3.4 to 3.6 that as N is increased the energy density at the diffraction head increases and the location x_0 of the first zero in the energy density decreases, that the normalized energy density at the diffraction head is greater than unity and that x_0 is less than unity. These observations lead to the expectation that fully corrected objectives will have greater resolving power than the idealized Airy-type objective. This expectation will now be justified.

Let two like pinholes having unresolvably small area A be located equidistantly from the optical axis and be illuminated as in a microscope by a substage condenser whose numerical aperture with respect to the object space of the objective is variable and is denoted by

$$N.A._{\text{condenser}}=n_o\rho_{cm}. \tag{40}$$

Let

$$s\equiv N.A._{\text{condenser}}/N.A._{\text{objective}}; \tag{41}$$

$$\beta\equiv 3.8317; \tag{42}$$

$x\equiv$distance measured in Airy units from the optical axis in the image space of the objective;

$2L\equiv$separation in Airy units of the centers of the geometrical images of the two pinholes in the image space of the objective;

$$Z_1\equiv\beta(x-L); \tag{43}$$

$$Z_2\equiv\beta(x+L); \tag{44}$$

$$\psi\equiv\frac{2J_1(2\beta sL)}{2\beta sL}; \tag{45}$$

$$u(Z)\equiv U(Z)/2\pi\rho_m^2, \tag{46}$$

where $U(Z)$ is to be calculated from eq 37 and 38 for fully corrected objectives;

and

$G(x)=$the distribution of energy density in the sharply focused diffraction image of the two pinholes.

Then it can be shown by repeating for opaque backgrounds an argument by Osterberg and Wissler [9] that

$$G(x)/4\pi^3A^2\rho_m^4\rho_{cm}^2=u^2(Z_1)+u^2(Z_2)+2\psi u(Z_1)u(Z_2). \tag{47}$$

Equation 47 is identical to an equation derived independently (and almost simultaneously) by Hopkins and Barham [10]. Equation 47 includes the effect of the numerical aperture of the condenser upon the diffraction image of the two pinholes.

The curves of figures 3.7 to 3.9 have been calculated from eq 47 for the case $s=1$, that is, for the case in which the numerical aperture of the substage condenser is equal to the numerical aperture of the objective. The curves $G(x)$ are symmetrical about the point $x=0$ and therefore have been plotted only for $x>0$.

The family of diffraction curves of figure 3.7 belong to fully corrected objectives for which $N=0.85$. A pronounced concentration of energy density occurs at the points $x=\pm 0.5$ Airy unit when the separation $2L$ is 1.0 Airy unit. The two pinholes are easily resolved when $x=\pm 0.5$, are undoubtedly resolved when $L=0.45$ but are probably

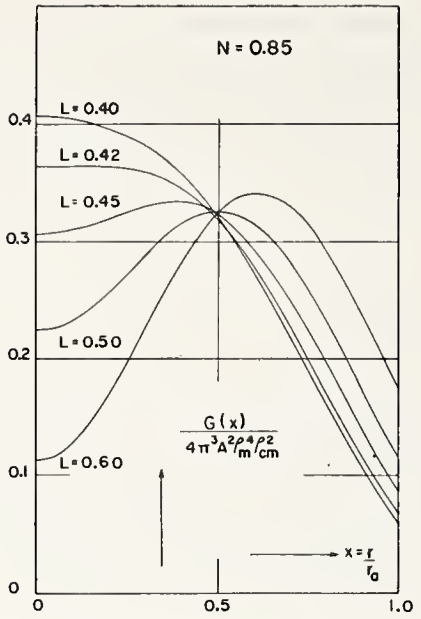


FIGURE 3.7. *Diffraction curves for two pinholes for the case $N=0.85$.*

The pinholes are located equidistantly from the optical axis and x is measured in Airy units from the optical axis.

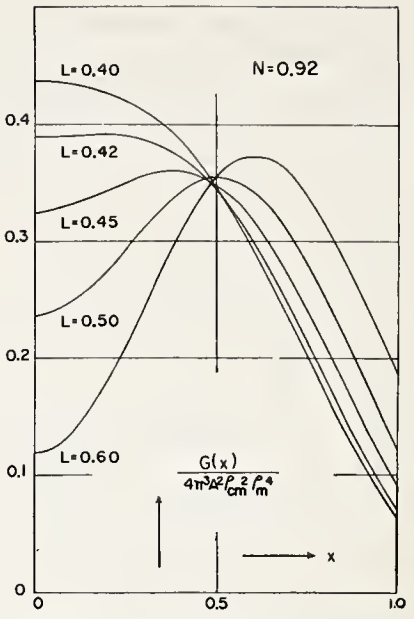


FIGURE 3.8. *Diffraction curves for two pinholes for the case $N=0.92$.*

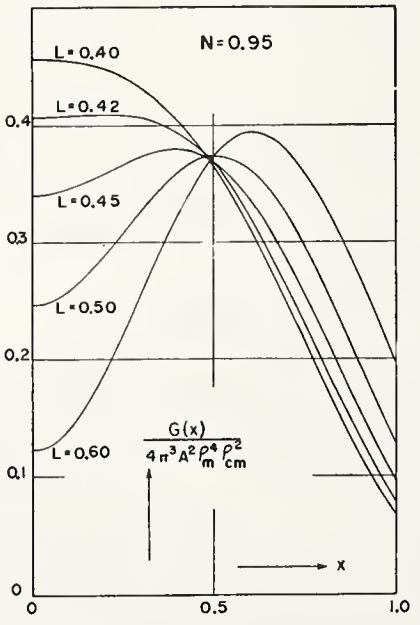


FIGURE 3.9. *Diffraction curves for two pinholes for the case $N=0.95$.*

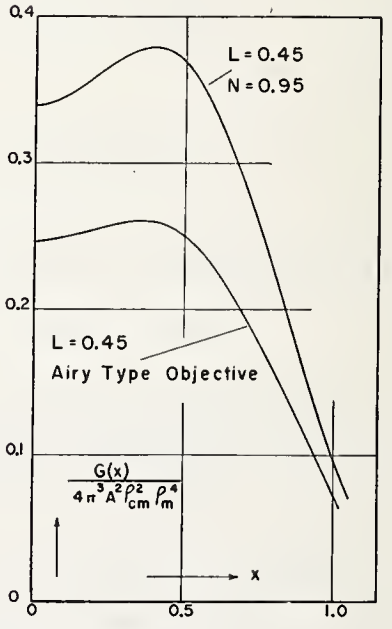


FIGURE 3.10. *Comparison of the diffraction curves for two pinholes as produced by an objective of Airy type and by a fully corrected objective for which $N=0.95$.*

The pinholes are separated by $2L=0.90$ Airy unit in the object space.

not resolvable when $L=0.42$. The *physical limit* [11] of resolution is the separation $2L$ for which

$$\frac{\partial^2 G(x)}{\partial x^2}=0 \quad \text{at} \quad x=0. \tag{48}$$

Evidently, the physical limit of resolution, $2L$, falls slightly below 0.84 when $N=0.85$.

The family of curves of figure 3.8 belong to the case $N=0.92$. A more definite dip in the energy density now occurs at $x=0$ for the half-separation $L=0.42$. The diffraction images described by this family of curves show definitely better contrast and resolution than do the diffraction images for the case $N=0.85$. This trend toward better contrast and resolution with increasing N is continued in figure 3.9 for the case $N=0.95$.

The physical limit $2L$ is estimated to be 0.82 when $N=0.95$. The physical limit $2L$ is [12] 0.843 with Airy-type objectives in viewing two, like, true pinholes in an opaque slide under the condition $s=1$. The physical limit of resolution is therefore lower (more favorable) with fully corrected objectives than with the idealized Airy-type objective.

The diffraction curves of figure 3.10 have been included for comparison of the fully corrected objective with the Airy-type objective. To compute $G(x)$ for objectives of Airy type, one determines $U(Z)$ in eq 46 from Airy's diffraction integral of eq 10. Whereas the half-separation $L=0.45$ is resolved by both objectives, resolution is more certain and contrast is better with the fully corrected objective.

Another interesting point of comparison is afforded by figure 3.10. According to the laws of geometrical optics, the concentration of energy density should occur at $x=\pm L$ Airy units when the separation is $2L$ Airy units. Figures 3.4 to 3.6 display the typical tendency that the concentrations of energy density do occur near $x=\pm L$ when $2L\geqq 1.0$, but that the concentrations depart from the locations $x=\pm L$ when $2L<1.0$. In figure 3.10 the concentration of energy density occurs near $x=0.4$ and $x=0.35$ for the fully corrected objective and the Airy-type objective, respectively.

The concentration of energy density lies closer to the geometrical location $x=0.45$ with the fully corrected objective than with the objective of Airy type.

In summary, a fully corrected objective that has negligible internal losses of light due to reflection and absorption is superior to the idealized objective of Airy type in practically all respects from the viewpoint of physical optics. The amplitude variation introduced over the converging wavefront by bending the rays from the object space into the image space in accordance with Abbe's sine condition is beneficial and should not be counteracted by the deposition of an absorbing coating over the exit pupil of the objective to render the objective of Airy type.

Conclusions

The following advantages over the idealized objectives of Airy type are possessed by fully corrected objectives that have negligible internal losses of light.

(1) The energy density at the diffraction head is higher. This energy density increases with $N \equiv N.A./n_o$ and approaches 16/9 times the classical value as N approaches unity. (2) The first zero of the energy density occurs nearer to the diffraction head. (3) The physical limit of resolution is lower (more favorable) in viewing two pinholes in an opaque slide. (4) Actual resolving power is better in viewing the diffraction images of two illuminated pinholes in an opaque slide. (5) Contrast in the diffraction image of two illuminated pinholes is higher.

The single advantage retained by the classical objective of Airy type is that the brightness of the diffraction rings is lower with respect to the brightness of the central maximum.

The primary diffraction integral for fully corrected objectives is reversible. Not all primary diffraction integrals have this property.

References

[1] R. K. Luneberg, Mathematical theory of optics, p. 344–386 (Brown University, Providence, R. I., 1944).
[2] H. Osterberg and J. E. Wilkins, Jr., J. Opt. Soc. Am. **39**, 553–555 (1949).
[3] H. Osterberg and J. E. Wilkins, Jr., J. Opt. Soc. Am. **39**, 554, eq 6 and 8 (1949).
[4] H. Osterberg and J. E. Wilkins, Jr., J. Opt. Soc. Am. **39**, 554, eq 3, 8, and 9 (1949).
[5] A. Gray, G. B. Mathews, and T. M. Macrobert, Bessel Functions, p. 211, eq 66 (Macmillan & Co., Ltd., London, 1931).
[6] R. K. Luneberg, Mathematical theory of optics, pp. 390–391, eq 50.45 (Brown University, Providence, R. I., 1944).
[7] I. S. Sokolnikoff and E. S. Sokolnikoff, Higher mathematics for engineers and physicists, pp. 536–542 (McGraw-Hill Book Co., New York, N. Y., 1941).
[8] H. Osterberg and J. E. Wilkins, Jr., J. Opt. Soc. Am. **39,** 556, figure 3 (1949).
[9] H. Osterberg and F. C. Wissler, J. Opt. Soc. Am. **39,** 558–561 (1949).
[10] H. H. Hopkins and P. M. Barham, Proc. Phys. Soc. [B] **63,** 737–744, eq 22 (1950).
[11] H. Osterberg, J. Opt. Soc. Am. **40,** 299 (1950).
[12] H. Osterberg, J. Opt. Soc. Am. **40,** 300, figure 5 (1950).

Discussion

Dr. F. Mooney, Bausch & Lomb Optical Co., Rochester, N. Y.: I would like to hear Dr. Osterberg redefine a fully corrected objective.

Dr. Osterberg: Insofar as my purposes are concerned, a fully corrected objective is without spherical aberration and is fully corrected for any design condition that is without color. With monochromatic light, you would not have to correct for color.

Dr. G. Toraldo, Instituto Nazionale di Ottica, Arcetri, Florence, Italy: I would like to remark on Dr. Osterberg's paper. I think that to treat the problem from a purely academic point of view is not sufficient when such large angles are considered. I have already discussed with Dr. Osterberg the question of polarization, but there is still another question. What is the meaning of the curves of the diffraction disks when one has such a large angle? One must specify what is intended. Also, it appears that the conservation-of-energy law is violated because either the other patterns or these are true. The other patterns contain more energy. Which is the correct one?

Dr. Osterberg: You have asked, I believe, two questions. One question you have asked really amounts to this: If we had an illuminated pin hole can we say that the wavefront that expands from it can be regarded as spherical?

The first remark to be made is this: In the classical viewpoint, of course, when considering Airy-type objectives we asked no question in this regard so I may stand my ground and say I take the same privilege. However, there is this to be said. I agree with Toraldo that the only way you can really solve these problems is to start with the atomic radiator in the source, trace your polarization vectors all the way through, and calculate the energy density at the end.

I have done this in connection with the building up of a theory of phase microscopy. I have started with the dipole radiators in the source and have traced the development through to the final energy densities and under certain approximations that are made we get the final result.

Now, the approximations in a theory like this are sometimes very difficult to analyze. Just what do they mean? If one supposes that the microscope illuminator has sufficient numerical aperture to focus the complete diffraction energy in the source then your conclusion would follow. That takes care of one question. What was the second?

Dr. Toraldo: The energy consideration, the diffraction disk.

Dr. Osterberg: Yes. Here is how I view that. Let us take a small cone of rays containing a certain amount of energy. On passing through a microscope objective, this energy flux converges as another, and longer, cone. As the angular aperture increases, the emergent flux is squeezed into a narrower and narrower cone relative to the incident cone, so the amplitude is greater on the outer portions of the wavefront.

Now, when you develop a theory for the Airy-type objective what do you say? You say that the energy on that wavefront is uniform. What have you done? You have thrown away energy. I have only included that energy and so I get the taller diffraction peaks.

Chairman: Dr. Herzberger.

Dr. M. Herzberger, Eastman Kodak Research Laboratory, Rochester, N. Y.: I wish to make one remark with respect to the beautiful paper of Prof. Zernike. I enjoyed very much the mathematical treatment as well as the agreement shown between theory and practice. I wish to say that geometrical optics is not as bad as it has been made to seem. The circle of confusion of which Prof. Zernike spoke is based on an inaccurate approximation to geometrical optics. From more exact computation one gets a picture that is similar to your diagrams and that shows the diamond shape of the image though, of course, it does not show the interference pattern.

Prof. Zernike: I want to make a remark on Dr. Osterberg's paper. I looked into the history of the subject myself about 5 years ago. I did not publish anything about it. A year and a half ago I visited H. H. Hopkins in England who has written a paper on this subject. Then we learned of a paper by "Ignatovsky" a Russian physicist who worked in Germany before 1917, but this paper was written in 1918 in Russian. He probably solved everything about it, including the directed vibrations of light and my impression is that because of that the influence is much less than Dr. Osterberg has shown. The Airy disk remains practically unchanged.

Dr. Osterberg: With respect to Dr. Hopkins, I am under the impression that any of his papers that I have seen had to do with assumed vibrations on the converging wavefront. Generally speaking,

518-743 O - 73 - 17

the amplitude is decreasing. I am not aware that Hopkins has studied the particular variation of this paper. Here we have included the variation of the amplitude produced by bending of the rays in accordance with the Abbe sine condition. Is it your impression that Hopkins has discussed this consideration?

Prof. Zernike: I could not say off-hand.

Prof. Maréchal. Hopkins has discussed the reverse. He considered an incident plane wavefront and studied the formation of the image assuming that the angular aperture was very great. It is exactly the reverse of the case studied by Dr. Osterberg.

In the line of Hopkins' work we have performed the computation for the diffraction image given by an aplanatic system of high numerical aperture (in the image space). The image is no longer circular, but elliptical, as Hopkins has shown. The direction of the longer axis of the ellipse representing the electric energy is also the direction of the shorter axis for the magnetic energy; now if we consider the Poynting vector we find a circular distribution. Experiments have been tried in order to show the ellipticity of the distribution of electric energy, but no conclusion has been obtained as yet.

Dr. Osterberg. Are you certain that he included the bending of the rays in his (Ignatovsky's) work according to the Abbe sine condition? If he did not he would not get this result.

Prof. Zernike. I should have added that he gave me a reprint of the paper to be translated but I have not referred to it for a year and a half.

Dr. Osterberg. Its contents should be interesting.

Dr. M. Golay, Signal Corps Engineering Laboratories, Fort Monmouth, N. J. We have seen many papers in which we are concerned with what happens to an image. We start with a point source, or if you like, from a perfect Airy disk, which I believe will give us the same results, and we arrive at an imperfect image. Of course, we are not interested in one lens. We are interested in many lenses. We are interested in what happens to the imperfect image and how much worse it is in the next step, and one question would be, what is there about these calculations that is additive and would permit one to take these steps individually and know what happens at the end of the system. I have a suspicion that some of the B^2 terms of Prof. Zernike have this additive property. If you look at it from the standpoint of thermodynamics, something is always lost and is never retrievable. Recently in communication engineering there have been beautiful applications of this notion of things that are lost and not recovered. I wonder if there is not an application in this field of that notion of entropy that has a good mathematical and useful significance, and can be applied to what happens in images and give us a good strong concept to discuss not just one lens but to discuss a system of lenses. This is merely a suggestion.

Dr. Herzberger. I would like to make a remark about what has been said. I think there is a law that governs the action of an optical system. The optical image formation can always be described as the result of a wavefront calculated from geometrical optics, diffracted at the exit pupil. There is ample experimental evidence to verify this statement, which is the basis of all calculations of the diffraction image.

Besides, there is some kind of an invariant, discovered by R. Straubel, that governs the connection between contrast and resolution in

optical systems. One can only improve one of these two factors at the expense of the other. Straubel especially investigated the influence of absorption on resolution and contrast, an investigation in which the theoretical conclusions were paralleled by experimental studies.

Dr. R. C. Spencer, Air Force Cambridge Research Laboratories, Cambridge, Mass. I am interested primarily in microwave optics and microwave optical systems. Although we have attempted to obtain some help from existing optical theory, in general we have been forced to attack directly certain problems that differ materially from the usual optical problems. These differences I would like to point out.

(a) The intensity illumination over the aperture of a reflector-type antenna or metal lens is not constant but reduced by as much as a factor of 10 at the edges. (b) The f/d ratios range from unity down to one-third. (c) The field, or the angle over which a horn will pick up the image, ranges in certain cases from 30 through 90° and at times 360°. (d) In some cases the feed is off-axis so that the optical system is not symmetrical. (e) We are working much closer to the limitation of the Rayleigh disk. For instance, a drop of 20 percent (one decibel) in the intensity of the center of the Rayleigh or Airy disk will raise the intensity of the side lobes (diffraction rings) to such an extent that the pattern becomes unusuable.

Operational methods [1] have proved a powerful tool in evaluating the effect of aberrations on diffraction patterns. For a one-dimensional aperture with amplitude illumination $f(x)$, the diffraction pattern is the Fourier transform $g(u)$ of $f(x)$. It can be shown that if $f(x)$ is perturbed by multiplication by a function $H(x)$ then the diffraction pattern is $H(-iD)g(u)$ where $D=d/du$. Thus $H(-iD)$ is an operator that transforms or perturbs the normal diffraction pattern $g(u)$. This device gives a short-cut method for evaluating certain patterns. Thus the pattern for an amplitude illumination $(1-x^2)$ within the range ± 1 and zero without, is $(1+D^2)g_0(u)$ where $g_0(u)$ is the amplitude diffraction pattern for the uniformly illuminated aperture. For an aperture illumination in a series of powers of x over the range ± 1 the diffraction pattern is a series of derivatives of $(\sin \phi)/\phi$.[2]

When applied to the analysis of phase errors over the aperture caused by various aberrations, the amplitude perturbation function is $H(x)=e^{i\phi(x)}$ and the perturbed amplitude diffraction pattern is simply the normal pattern $g(u)$ multiplied by the operator $e^{i\phi(-D)}$, the rule being to replace x by $-iD$. This has been used successfully in evaluating the effects of defocusing and should converge for other aberrations. In general, the operator is expanded in powers of D. For example, in the case of Fresnel diffraction for a uniformly illuminated aperture the aperture amplitude $e^{i\beta x^2}f(x)$ is transformed over to the amplitude diffraction pattern $e^{\beta D^2}g(\phi)=(1+\beta D^2+\beta D^2+\beta^2 D^4/2!+\ \ldots)g(\phi)$ where where β is the phase error at either end of the aperture and $g(\phi)=$

[1] MIT Rad. Lab. Series, 12, 186 (McGraw-Hill Book Co. 1949): (a) Roy C. Spencer, Paraboloid diffraction patterns from the standpoint of physical optics, MIT Radiation Laboratory Report T-7 (Oct. 21, 1942); (b) Roy C. Spencer, Synthesis of microwave diffraction patterns with application to $\csc^2\theta$ patterns, MIT Radiation Laboratory Report 54-24 (June 23, 1943); (c) Roy C. Spencer, et al., Tables of Fourier transforms of Fourier series, power series, and polynomials, MIT Radiation Laboratory Report S-58 (Aug. 30, 1945); (d) Roy C. Spencer, Fourier integral methods of pattern analysis, MIT Radiation Laboratory Report 762-1 (Dec. 5, 1945); (e) Roy C. Spencer and Pauline M. Austin, Tables and methods of calculation for line sources, MIT Radiation Laboratory Report 762-2 (March 30, 1946).

[2] Harvard Computation Laboratory Series, vol. 22, Tables of the function $(\sin \phi)/\phi$ and of its first eleven derivatives, Introduction—Section II "Application" by Roy C. Spencer.

$(\sin \phi)/\phi$ is the amplitude diffraction pattern for a uniformly illuminated aperture.

A special aspect of this problem is the fractional loss in gain (or brightness at the center of the Rayleigh disk) due to phase errors. If the phase front is approximated by a plane so that the square of the error between the phase front and the plane is minimized by the method of least squares, using as a weight function the amplitude illumination $F(x, y)$, then the fractional loss in gain[3] is given by

$$\frac{\Delta G}{G}=\frac{\int \phi^2 F(x, y)dxdy}{\int F(x,y)dxdy}=\overline{\phi^2}.$$

Thus, neglecting powers of ϕ above the second the fractional loss in gain is the *weighted-mean squared phase error*, with ϕ in radians. Needless to say, if the loss in gain is held down to a few percent, this sets an upper limit on both the increase in the width at half power and the energy increase in the side lobes.

Dr. J. G. Baker, Research Associate of Harvard College Observatory, and Optical Consultant to the Perkin-Elmer Corp., Norwalk, Conn.: I should like to ask Dr. Zernike if in his treatment he can include the primary astigmatism in terms of the higher order angular departures from the optical axis. When you go far off-axis, large aberrations have to be considered, and I am wondering whether your expressions will not have to be modified. You have normalized to unity over the aperture in each case and I am wondering if corrective factors are needed when you get to large oblique aberrations? The image may show hydrid symmetrical and unsymmetrical aberrations.

The question is of interest to me because in some systems I have designed, the inclinations run to 30 degrees off-axis and yet the requirements on image quality are comparable to those on-axis. Vignetting is another thing to consider.

Dr. Zernike: We have applied the theory only to circular pupils.

Dr. Maréchal: There is no essential difference in the off-axis theory at moderate angular apertures for a circular pupil.

Dr. Baker: I have one system of 52-degrees total field at f/0.65 where these considerations are of importance. The vignetting arises from film-holder shadows as well as from lens and mirror cells, and the pupil is far from circular.

Mr. A. J. Lipinski, University of California, Los Alamos Scientific Laboratory, Los Almos, N. Mex.: I wish to address a question to Dr. Zernike. How did you consider the geometrical as compared to the diffraction images? You mention four times better just by the diffraction theory and then you later mention this can be made by balancing the aberrations to 28 times what actually appears geometrically. Then you say in a practical case you reduce it to 5.6 times. I am a little uncertain as to what you mean.

Prof. Zernike: I had to abridge my talk somewhat. I had really in my paper a somewhat different thing. The 5.6 or 28 refer to the difference between balancing the error or by fourth and sixth orders or by not doing that. This polynomial (illustrating) would mean

[3] Roy C. Spencer, A least square analysis of the effect of phase errors on antenna gain, AFCRC Report E5025 (January 1949).

that you want the ray to go up and down so that you have this circle of confusion. You have the answer to this in the paper. I find that by doing it geometrically, making a transverse aberration of the rays you get an approximation that is not too far from the diffraction theory treatment, which means that judged by the geometrical optics you are always rather far out. This disk that would be calculated from the rays is much larger than the relative function that you get.

If you look at the diffraction properties, it would be only 1.75 times less than the correct one. I think that answers your question.

Mr. G. H. Conant, Jr., Harvard College Observatory, Cambridge, Mass.: Dr. Zernike, would you restate your ordinate and abscissa on the diagram of the isophotes? Axial sections near the focus were shown.

Prof. Zernike: It was just near the focus of the aberrations. You would have to rotate the entire diagrams in order to get a three-dimensional figure.

Prof. Maréchal: The question of resemblance between the effective diffraction pattern and the geometrical approximation is rather intricate. It seems that it depends entirely upon the nature of the aberrations. In the case of astigmatism if you focus on one focal line the diffraction pattern is very similar to the geometrical focal line even if the aberration is very small (half a wavelength). If you increase the aberration you will rapidly obtain the aspect predicted by geometrical optics.

In the case of coma it seems that you have to go up to three or four wavelengths at least to have some resemblance between diffraction image and the geometrical one, and for spherical aberrations, it seems to be still more. In fact, it is quite difficult to tell exactly how we can define the resemblance.

In the case of coma you have still a high maximum when the aberration is of the order of magnitude of 2 or 3 wavelengths. Instead of rings you have horseshoes.

Dr. H. R. J. Grosch, General Electric Co., Lockland, Ohio: I would just like to comment upon noise. I suppose we could interpret band width, which, after all, is the limitation of the channel over which the information is coming as some function of N, but what are we going to do about the question whether the noise introduced by the imperfect lens system is what the communication boys call "white" or not? I think the trouble is that there are highly selective effects. I have a feeling that we are not anywhere near close enough to information theory from the communication engineering viewpoint yet.

Mr. Golay: As I understand it, we have a loss due to the poor quality of the system. This poor quality may be something that cannot be avoided but it is not noise. We may have a system that is as good as we can make it. But can we have a measure of the poor quality that is additive so that we can predict what we will have at the end of the system?

Dr. Baker: This is a short question. Are these orthogonal functions necessary and sufficient to represent any possible aberrations of the wavefront to all orders?

Prof. Zernike: Certainly.

4. Bases for Testing Photographic Objectives

By L. E. Howlett [1]

General Considerations

It is a matter of more than passing interest that in spite of the fact that optical instruments have for a very long time been widely used as tools in scientific research, there is a great deal of disagreement as to the way in which their image quality should be judged. An explanation for this condition, which is at least plausible, may be found in the historical development of the uses to which lenses have been put. Whatever the real validity of such an explanation it serves to bring out the erroneous trends of thought on optical performance that have been largely responsible for the present diversity of opinion.

There have been three distinct periods in the development of the usefulness of optical instruments to meet scientific needs.

Until the early years of the nineteenth century optical instruments found their major, if not exclusive, scientific use as an aid in making visual observations. Astronomical research combined with other applications of telescopes was probably the most important scientific use.

Shortly after the opening of the nineteenth century the development of photographic science was very rapid. For pictorial reasons a demand arose for lenses that would cover a reasonably large field of view. To attain this end a diminution of image quality below that of the astronomical requirement was considered quite acceptable because the intrinsic resolution of available emulsions was certainly inferior to that obtained visually on the axis of telescopic systems.

By the twentieth century lenses came to be combined not only with photographic emulsions but with many other kinds of energy receivers such as thermopiles, photocells, etc. They became an important element in television where the character of their optical imagery is of great importance in the performance of electrical circuits and the over-all quality of a system as a transmitter of information in the form of electrical signals.

It is a most unhappy circumstance that during the expansion of the use of lenses from the stage where they were an aid to visual observation to where they were used to make observations in combination with photographic emulsions, thermopiles, photocells, etc., designers ignored the necessity of reconsidering and adjusting for validity the design criteria, which had proved wholly acceptable for the design of good objectives for visual use. It seems they had become generally satisfied with an erroneous conception that if a lens performed well in combination with the eye it would perform equally well when combined with any other type of energy receiver. Those who would not expect a microscope objective designed for the infrared to excel in the

[1] Division of Physics, National Research Council, Ottawa, Canada.

ultraviolet nevertheless assumed that a good visual lens must be a good photographic lens. The thought that a lens performing in only an adequate way with the eye might be superior to a first-quality visual lens when combined with a photographic emulsion or other receptor did not receive appropriate consideration. It seems that only currently is the condition underlying such a concept being given any attention by designers. If all lenses over their reputed field of view were limited only by diffraction theory no such confused thinking would have developed in the study of lens-film and lens-electrical combinations. The fact that all lenses with significant fields of view have their imagery limited by aberrations makes possible a very varied performance of the same lenses when combined with different types of observing receptors. Therefore, it cannot be too strongly emphasized or frequently repeated that when the image of a point source is an irregular spot in which the distribution of energy is a function of specific residual aberrations, a variety of interpretations of its form and size can be made by energy receivers of different chromatic and contrast sensitivity and that hence a number of different standards of performance are possible.

In the beginning, astronomy presented the most important demands and the requirements of this field together with limitations of theoretical understanding and available optical materials governed the type of systems developed. In observing stars it is not essential, even though sometimes desirable, that the telescope cover a large field of view. In any case, a very large amount of astronomy can be carried out by instruments having quite limited fields. At the same time it is of prime importance that the quality of the image be of the very best. It so happened that with these conditions it was possible to develop lens systems of quite simple design with available materials that would form on the axis an image limited only by diffraction effects. Because of the use it was natural enough to use simple visual experiments to determine the quality of the image formed by the finished product and to judge in large measure the success of the design by the ability of the system to separate visually the images of two closely placed point sources in the object space. Diffraction theory permitted the mathematical prediction of the intensity distribution in the image of a point source when no aberrations were present and an ultimate standard of performance could thus be set. The ability of an optical system to separate point sources came to be called resolution and Lord Rayleigh's name is associated with the standard mathematical definition of this property when diffraction theory alone determines image quality. Resolution, in the absence of aberrations and except for its dependence on the wavelength of the light employed, came to be considered as an intrinsic property of the geometry of the system. Since the range of wavelengths used in such optical systems is generally not large, the persistent tendency to slip erroneously into thinking of visual resolution as an intrinsic property of optical systems was not too calamitous. However, it was most unfortunate that this mode of thought had become a habit when consideration had to be given to photographic objectives in which images are limited by residual aberrations.

When aberrations are present the distribution of the energy in the image, both as to intensity and wavelength, is dependent on the particular residual aberrations left by the designer and the quality of the light forming the point source. This energy distribution will be a

function of the particular design and it is difficult to define the function completely and accurately on a theoretical basis. The distribution will vary with position in the field. The position of best focus or the position of the smallest size of image will vary with the position in the field of view and the position can be different for different means of observation—eye, photographic emulsion, photocell, etc. When the lens is used in photography, it is practically essential that the field of view be recorded on a flat plane. Since the position of best focus varies with the angular position in the field of view, a problem arises as to how to select as the focal plane for a particular lens-emulsion combination the one that will give the best compromise between the size of spots in different positions in the field of view. The interpretation of the energy in the image plane for a photographic emulsion will not be the same as that of the eye because of different response characteristics. It may see either a larger or smaller spot. In consequence the lens-emulsion combination may be able to separate two nearby point sources with less or more facility than the lens-eye. The position of maximum brightness of the spot may appear in different places to the eye and to the photographic emulsion. Consequently photographic records of distortion might differ from visual ones. There are emulsions with different sensitivities and each will give its own interpretation of the image as to size and apparent energy distribution. Other types of receptors give other interpretations of the same energy distribution because of their own peculiar responses. It becomes immediately apparent from these considerations why a lens of high visual performance is not of necessity equally satisfactory for photographic emulsions or for forming a signal to be transmitted electrically.

The foregoing shows that the essential requirement for judging the quality of images in the presence of aberrations is that the judgment be based, not on any particular type of energy receiver for observing the images, but on that particular energy receiver with which it is proposed to couple the lens for some definite scientific task. It is not to be taken from this that visual observations are to be outlawed in making studies of photographic objectives. It does require, however, that if for convenience a visual method is desirable a series of experiments must first be done to show the equivalence of the visual methods to those of practice. The same applies to the substitution of any other kind of receptor for the observations. The equivalence once established holds only for one lens type.

One can point out that for well-designed photogrammetric objectives no difference has yet been shown between the distortion as measured visually and the distortion as measured photographically. On the other hand, it is reasonable to suppose that it would be possible for a significant difference to exist in the case of lenses of inferior design or poor workmanship. Accordingly, if the purpose of the test is to establish acceptability of the new lens for photogrammetric purposes, it is more reasonable because it is safer to determine photographically the distortion of a lens of unknown performance.

The plane of best photographic focus can differ from the position of best visual focus even in a good lens. Consequently this plane must be initially selected for a particular lens on the basis of a photographic investigation. Such an investigation is tedious, and it is fortunate that for a particular type of lens it seems that the plane of

best photographic focus once determined can be located for other lenses of the same type by reference to the position of best visual focus on the axis. Preferably, several lenses of a type should be used to determine the relationship.

Inasmuch as some earlier visual methods of image assessment are shown to be generally inappropriate for lenses intended for photography and other specialized purposes, it is well to consider whether the long used resolution criterion is a suitable one for grading the performance of different types of photographic objectives. We must make the decision in the light of the use. In the case of lenses used for photography that will be the basis of photogrammetric routines or reconnaissance the collection of information is of prime importance. For this purpose the resolution criterion seems very apt. The ability of the lens to reveal information must be very intimately and directly related to its ability to resolve adjoining fine detail. There may be some question as to whether the relationship is a linear one. Indeed, there are good grounds for thinking that in air photography *useful* information does, to some extent, come in size ranges. Thus, a 10-percent increase in resolution might be relatively unimportant in small scale photographs for forestry interpretation since such an increase does not make possible the recording of another size group of useful detail. On the other hand, at large scale when a built-up area is being photographed an increase of 10 percent in resolution might very materially increase, by a much larger percentage, the amount of useful information available for interpretation. Although accepting the validity of resolution as a method of evaluating a lens for air photography, it can still be conceded that experiments designed to relate such measurements to photographic interpretability may be of considerable general interest. Nevertheless, it seems unlikely that these experiments will either disprove that a lens of higher resolution will give more information than a lens of low resolution or disturb the continuing practical demand for lenses of higher and higher resolution.

An Acceptable Testing Procedure for Aerial Photographic Objectives

Keeping the previous discussion in mind we can propose suitable tests for photographic objectives intended for aerial photography. Different but similarly conceived tests can be proposed for lenses intended for other tasks. The procedures set forth in the following are those currently used for aerial photographic objectives in the National Research Laboratories of Canada.

A good quality collimator having a focal length at least four times that of the objective to be tested is used to place the resolution target at infinity since the object distance in aerial photography is always the equivalent of infinity. The excess focal length of the collimator guarantees that its aberrations will be insignificant compared with those of the lens under test. The quality of the illumination of the target is mean noon sunlight since this is characteristic of aerial-photographic operations. In front of the collimator is a sturdy beam that swings about an axis near the end closest to the collimator. The test lens is mounted on the beam over this axis so that its entrance pupil is flooded with light at all angular positions of the beam. The

latter also carries a plate holder that can be placed at varying distances up to 60 inches from the test lens. The test lens carries the filter or filters that will generally be used with it in practical operations. This will include not only colored filters but any filter that is used to equalize the intensity across the field of view of the test lens. The lens can bc adjusted so that the film or plate plane is parallel to the machined surface of its mounting flange. Provision is made for maintaining the film flat either by a register glass or a suction back. The plate holder covers a narrow strip of the field of view from corner to corner of a 9- by 9-inch format. Recordings of the collimator target are made photographically at a number of angular positions in a series of parallel planes in the neighborhood of the visual focus. The emulsion to be used with the lens will be the one to be employed in air operations. The film strips are processed according to the recommended procedure under careful sensitometric control. The exposure time at each position of the field is the same. It is chosen so that the density produced at each position favors the negative material yielding the maximum resolution. If vignetting is severe, positions representing smaller areas of the format are sacrificed. After processing the film strips the resolution is assessed with the aid of a binocular microscope. Thc magnification and illumination are so adjusted that the reading of the maximum possible resolution is favored. From these data through-the-focus curves are plotted for each angular position in the field. From values taken from smooth curves based on these points off-axis graphs are plotted for the several planes studied. The resolution is weighted on an areal basis. The resolution at each position of the field is assumed typical of the annulus in which it occupies the central position. Appropriate allowance is made for the fact that the outer annuli are incomplete because the camera format is rectangular. A final curve can be drawn, which relates average resolving power over the rectangular format for a series of planes parallel to the mounting flange. From this curve the plane of best average photographic focus can be selected.

Visual observations are taken to determine the position of best visual resolution on the axis. The separation of this position from the plane of best average photographic resolution is used to permit routine focusing of lenses of the same type by a simple visual routine.

The Shape and Contrast of the Resolution Target

No reference has so far been made to the shape and contrast of the target best suited to lens resolution studies. Both these factors have a bearing on the actual numerical quantity that will be obtained for the resolution. Therefore it is highly desirable to obtain some measure of general agreement between laboratories on the exact shape and contrast of the target so that values obtained by each may be immediately compared. Unfortunately, there is very wide disagreement on these points. Most targets will place lenses in the same order of merit, but there will be exceptions. These are important ones for performance testing because the proper design of a target will be more severe on inferior lenses. Contrast of target for certain uses of lenses is more important than shape of target. For performance tests in aerial photography, if testing is done at only one contrast, low-contrast targets should certainly be used. This is essential

because of the very great preponderance of low-contrast detail in scenes from the air. However, it is certainly of advantage to do tests also at high contrast, providing the significance attached to these results from a performance point of view as compared with results from low-contrast targets is in proportion to the relative frequency of high- and low-contrast detail in an air photograph. Perfect lenses would be placed in the same order by all shapes of targets at all contrasts and it is again the residual aberrations that complicate the problem of assessment.

The correct basis of choosing a target can be simply stated and there is unlikely to be serious disagreement on the point. The form and contrast of the target should conventionalize the task in the practical application of the lens. It should be as simple as essential requirements allow to make testing an easy routine. In air photography this means that it should attach equal importance to edges in any direction and that it should take account of the mutually destructive effect at the limit of resolution of neighboring edges at any orientation with respect to each other. It should be possible to associate some degree of recognizability with the limit of resolution. It should not make demands for qualities that will serve no useful purpose in the application of the lens or production difficulties may be caused unnecessarily. Important requirements must take precedence over lesser ones.

Some interest has been shown in letters and other targets calling for a greater degree of recognition than the more conventional geometrical forms. It is argued that they have the merit of presenting the observer with an unknown target. The value of this is questioned, since in very extensive observations on resolving-power targets in our laboratories practically no discernible tendency to cheat unconsciously has been noted. The establishment of the limit of resolution of adjoining edges is after all the prime requirement whether the edges constitute a conventional target, a letter, or some natural object in the landscape. It therefore seems unprofitable to complicate the testing for no profitable result. In using letters there is also the further complexity or uncertainty that the recognizability of a given letter will vary with its orientation in the field of view because of the asymmetry of the spot-like image of a point source, which due to aberrations is formed by a photographic objective. It therefore appears that there is no advantage in going to great complexity. A single, simple target conventionalizing to a reasonable degree the practical task will fully suffice.

We think we have met the target requirements outlined in a satisfactory way for routine lens performance tests in the laboratories of the National Research Council of Canada. We have chosen a bright annulus form of target on a darker background. The thickness of the annular ring is equal to the diameter of the central area. In the recorded images of this target resolution is considered to exist as long as the image appears as a line of density enclosing a central light area. If the line of density is broken by astigmatism or if the central area disappears, the limit of resolution is exceeded. This accepts resolution with some distortion as having adequate recognizability. The target has the advantage that it does not place any particular importance on edges in any given direction such as the radial or tangential. This, as stated earlier, is certainly the condition in actual operations where equally important detail will be lying so that its edges are in

any direction at all. The simple single form of this target as compared to lines in the radial and tangential directions, or in more directions, permits an experimental averaging of the various factors. It thus avoids the necessity of making an average on some arbitrary arithmetic basis of the resolving power in two or more specific directions. The mutually destructive effect at the limit of resolution of nearby edges lying in different directions, which obtains in any aerial photograph, is rather adequately represented by the mutually destructive effect of small sections of the inside of the annular ring. Resolution measurements on lines in specific directions take no account of this important factor which is particularly important in the case of poorer lenses. At the limit of resolution the aberrations in all directions at the inside edge of the annulus add up to destroy resolution and cause the disappearance of the central area. It has been definitely shown in the case of lenses of inferior design that the annulus type of target records a relatively more rapid fall-off of resolution towards the corner of the photograph than radial or tangential lines. Under such circumstances it is reasonable to suppose that the resolution picture given by the annulus target is more representative of what will occur in air photography than the performance suggested by the resolution on radial and tangential lines. This revealing property is of obvious value in testing new lenses of unknown performance.

Selwyn and his coworkers first suggested during the last war that because the great majority of aerial detail is of low contrast it would be sensible to adopt a low-contrast target for the assessment of aerial-photographic objectives. Our laboratories in Ottawa have accepted this proposal and our annulus targets have the log contrast of 0.2 suggested by Selwyn. Recent work by Carman and Carruthers has given extensive experimental authority to Selwyn's proposal and has further shown that the typical average log-contrast ratio for a great variety of scenes is considerably lower than 0.2. A log contrast of 0.1 might be considered more typical. It is not thought that any great increase of validity will be obtained by substituting a 0.1 target for a 0.2 target in resolution tests. The difficulty of making the tests would be much increased. Nevertheless, tests are under way in our laboratories to make certain that this conclusion is warranted.

It is sometimes argued that high-contrast targets reveal differences between lenses more readily than low-contrast targets. This contention is undoubtedly true. The fact in itself is not of great importance unless the differences revealed are of real significance in the work to be assigned to the lens. In an aerial scene the amount of high-contrast detail is but a small fraction of the low-contrast detail. It seems unreasonable to evaluate the results obtained with high-contrast targets in any other proportion but this fraction. For aerial photography a lens that has 120 lines per millimeter on a high-contrast target and 15 lines per millimeter on a low-contrast target is not by any manner of means twice as good as a lens that has 60 lines per millimeter on high-contrast target and 15 lines per millimeter on a low-contrast target. Indeed, it is only very slightly better. On the other hand, if the low-contrast figure for the lens performing in the poorer way at high contrast were 25 lines per millimeter instead of 15 lines per millimeter it certainly would be nearly 66⅔ percent better than the other on an overall basis. Such a lens as that postulated is highly desired at the present time.

The Distinction Between Two Types of Testing

The relative merits of performance testing and tests that are designed to study one or more of the various physical factors that contribute to overall performance are quite often vigorously argued. Such arguments are quite pointless since the two kinds of tests serve quite different purposes. The typical designer requires, for the development of his science, physical information specified by him, which will help him to improve his methods and his designs so that his product can meet performance tests. The choice of information must be his. The user of the lens on the other hand is not concerned with these design problems. He wants to know how well a particular lens will cope with the problems with which it will be faced in a particular task in which he is interested and not some other task. Since all optics is generally a compromise one cannot usually meet all working requirements equally well. The two types of experiments may well be quite different in form since they are different in purpose. One is really an investigation and the other is a test. Neither type of work should have its usefulness limited by confusing the two separate requirements and the user must be left as free as is the designer to specify the testing requirement that best suits his problem. In our laboratories we are interested in both types of work. Experiments of the second type are at present going on, but no reference will be made to them here since the concern of this paper is to emphasize the proper general requirements for performance tests and to treat more specifically those required for photogrammetry and air reconnaissance. Agreement on performance tests does not need to await the solution of the long term problems of improving image formation. Performance tests must judge the end result in the most informative way and need not contribute directly to improvement of design.

Although performance tests do not of necessity contribute aid to the designer in his efforts to meet the user requirement, those outlined here have in fact made such a contribution. It is a general one but of the greatest importance. Because it does not point a detailed path it does not seem to have commended itself to many designers.

By contact printing a resolution target on to Aero Super XX it can be shown that Aero Super XX will resolve about 25 lines per millimeter with a line target having a log contrast ratio of 0.2. For the many aerial photographic objectives that have been tested in our laboratories in combination with Aero Super XX, the effective maximum resolution of the combination turns out with monotonous regularity as about 15 lines per millimeter. This is approximately 60 percent of what the emulsion itself can do. The apparent visual resolution may be 150 or more lines per millimeter and will vary greatly from lens to lens and type to type. The high-contrast photographic results may also be quite high and varied. It is suggested by these results that the designer is endeavoring to get some fraction of the energy from a point source into a very compact spot in the image space and letting the rest fall where it will outside this spot. Visually this central spot is sufficiently concentrated, compared with its surroundings, to yield quite high resolutions. It is often of the order of

hundreds of lines per millimeter. The resolution may even be quite impressive photographically for high-contrast targets on Aero Super XX. However, for low-contrast targets on Aero Super XX, the light that falls outside of the central compressed spot is very significant, and in consequence very substantially lower resolutions are obtained. Much better results, and a far more useful lens, would be obtained if the designer ceased from making a concentrated spot from a relatively small fraction of the light for visual interpretation and concentrated his attention on packing nearly all the light to which the emulsion is sensitive into a spot of quite significant size but one that would correspond to a resolution of 25 lines per millimeter on Aero Super XX with a low-contrast target, mean noon sunlight and a minus blue filter. In such a case we would have a lens ideally suited to work with this particular emulsion in air photography. Present lenses have characteristics that fight with those of the emulsion. Since there is little hope of changing the emulsion the lens designer must change his design criteria.

It is of interest to note that some time ago a number of metrogon lenses were tested by the methods outlined earlier and that their performance on low-contrast targets were very uniform. Visually and photographically at high contrast there were considerable differences. Nevertheless, no distinction has ever been made between these particular lenses by those continually using photographs from them for photogrammetric purposes over a period of years. This would seem to be practical subjective evidence to justify concluding that the low-contrast results are judging the important factors of performance in so far as the mappers are concerned. It has also been found in our laboratories that designs said by designers to be sensitive to certain curvatures, separations or surface figures are in point of fact quite insensitive to one or all of these factors with Aero Super XX, low-contrast targets, mean noon sunlight and minus blue filters. Again, this suggests that there is a straining to attain something that gives little profit in the practical photographic problem although the designer's concern about close tolerances is shown to be fully justified on visual performance of the same lenses.

Conclusions

The foregoing discussion leads to some important conclusions. Performance tests of lenses must be based on the condition of use. They must not be confused with experiments that the designer may require to obtain information on ways to improve his product. Performance tests for aerial photographic objectives must take account of the quality of the light and the characteristics of the instrument that will record the image. Resolution tests are generally an adequate means of testing the performance of these lenses if the main purposes is to secure information. A target for resolution tests must conventionalize, both in form and contrast, the task to be assigned to the lens in practice. Results of such testing suggest that we would get much better lenses for air photography if all the energy from a point source could be concentrated in a spot in the image space that would give, at any position in the useful field of view, 25 lines per millimeter on a

target having a log contrast of 0.2. On the basis of tests carried out on most of the available aerial photographic lenses this is obviously not being done at the present time although enough energy is being sufficiently concentrated into a smaller spot to lead the eye to observe a much larger resolution because of its ability, by reason of its response curve, to ignore to a sufficient degree radiation that falls outside the main concentration.

5. Quality Aspects of the Aerial Photographic System

By Duncan E. Macdonald [1]

As the work at the Boston University Optical Research Laboratory is primarily concerned with the field of aerial photography, the considerations of this paper have been directed to that field. However, it is felt that aspects of this discussion may be considered for more general application.

In aerial photography, pictures are taken of ground objects through long air paths by cameras, lenses, and associated equipment mounted on unstable and translating platforms. The resulting photographic negatives, consisting of nonhomogeneous clumps of silver, arranged in numbers and sizes as to approximate the geometric pattern of reflectivities of the ground objects, are then printed. These positive prints are viewed by trained observers known as photointerpreters, generally under low-power stereoscopes. These interpreters then perform the important transduction process—that of converting the message coded in silver clumps into a verbal report. This system is shown in figure 5.1. This paper will discuss the role of several of these components in the system performance.

Aerial photography has as its objective the gathering of information and, thus, the capacity of this photographic system for revealing information is taken to be a meaningful performance criterion. It is perhaps in order to review briefly the development of a common quality criterion that is now in general use, namely that of resolution. In 1850 Dawes [1] [2] reported on the result of his experiment on the conditions for separation of two stellar images, and in 1879 Rayleigh [2] presented a theoretical criterion for the resolution of two point images. Inasmuch as point images are difficult to treat experimentally, it has generally been the custom of the experimentalist since that time to employ test objects consisting of a periodic array of finite geometric details (usually lines) and then to score the resolution of the system in the appropriate coordinate in terms of the fineness of detail that it records (lines per millimeter).

In such fields as astronomical and spectroscopic photography, the original concepts of Rayleigh and Dawes appear to coincide in many instances with the technical objectives. Although at the present time a resolution score is used with some measure of success as a quality criterion for aerial photography, the message presented to the system is of a character quite dissimilar to that of the usual test object. It therefore appeared in order to examine how this and other criteria related to the performance of the system.

In the many approaches to the problem of photographic performance, there is, in general, one point in common—the consideration that the limit of resolution of the system is that point where contrast

[1] Director, Optical Research Laboratory, Boston University, Boston, Mass.

[2] Figures in brackets indicate the literature references on p. 72.

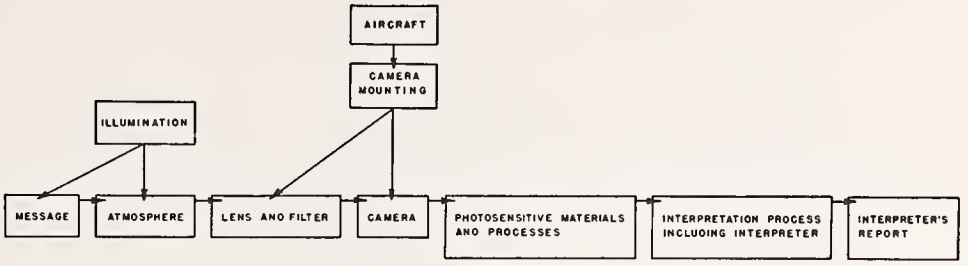


FIGURE 5.1. *The aerial photographic system.*

is degraded to a value below that of the visual contrast threshold. This implies that resolution is a function of the degradation of contrast of a system and suggests that a contrast-reduction function might provide some insight for the judgment of aspects of the system performance.

Consider the resolution limit of a lens-camera-film combination determined in the laboratory as being the limit of a band-pass filter. The nature of a vast majority of man-made objects, and indeed of the majority of natural objects of aerial photographic interest, is such that perfect imaging and recording would reveal square waves of intensity. (This is one characteristic of the message.) However, if square-wave detail is incident on the filter as is observed in the laboratory when employing a parallel-line test object, the shape, amplitude, and mean power of the waves become more and more altered as one approaches the band-pass limit.

The filter concept coincides with the one-dimensional aspect of the resolution target. For example, in the case of scanning pictures in a line-scan, video-type system, the limit of frequency of the system is VR, where V is the velocity of scan in millimeters per second and R the resolution in lines per millimeter. Whether the frequency characteristic is spatial, as in the case of a between-the-lens shutter, or is temporal, as in the line-scan application, the resolution number is directly proportional to the limiting frequency for a given set of parameters.

A photographic system may, then, be described in terms of its response to (a) detail that is of ordered occurrence, as in the case of the resolution target, (b) detail that is isolated, and (c) detail that is random, as, for example, the spatial frequency characteristics of the objects on which the system is to be employed as a detecting device. Because of the mutual influence of areas in proximity due to the spreading of the energy by the system components, particularly by the emulsion, the contrast-reduction functions differ for each of the above classes. (Analogy may be made to network theory, where in linear systems the steady state, transient or random responses uniquely determine the characteristics of the system.) Schade [3] has treated optical elements as low-pass filters and discussed the use of a "transfer characteristic" to describe the performance of a system component in terms of temporal responses for a given set of operational parameters.

Such characteristics in the aerial photographic system have been recognized as contrast-reduction functions. Contrast-reduction characteristics of a 6-inch $f/6.3$ Bausch and Lomb Metrogon, a 12-inch $f/5.0$ Eastman Kodak Aerostigmat, and a 24-inch $f/6.0$ Eastman Kodak

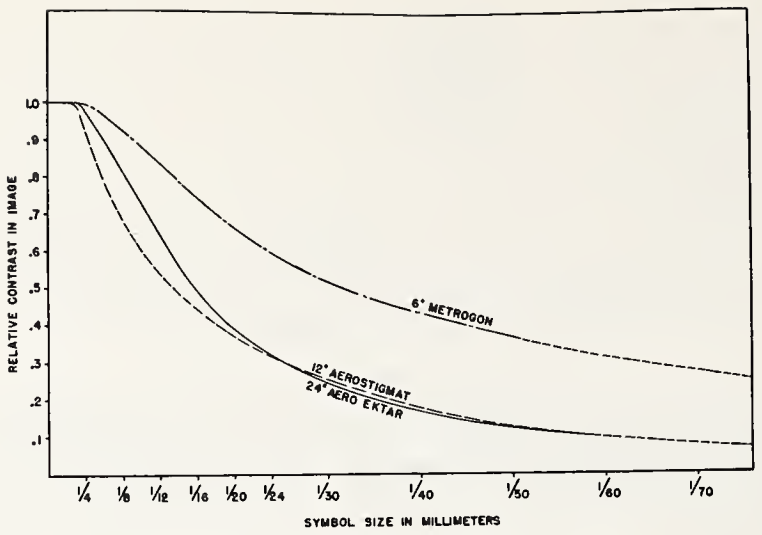


FIGURE 5.2. *Relative contrast in image.*

Aero-Ektar have been studied in the laboratory. Image contrast has been defined as $C=(T_b/T_l)-1$, where T is the transmission as measured on the negative, with subscript b denoting background, l denoting line or object. Data were obtained from photographs of line patterns, employing a photographic plate inclined to the focal plane, as discussed by Zschokke [4] and later modified by Wetthauer [5]. The length of the target lines falls normal to the axis of tilt.

Figure 5.2 shows the contrast-reduction characteristics as measured on Super-XX emulsion for the case of isolated symbols, that is, isolated white lines on a dark field. The contrast is expressed as a function of image size.

Because of the nature of reduction in contrast, it is clear that in order to detect a symbol, more contrast is required in the object as the image becomes smaller. The limiting contrast expressed as a function of image size may be thought of as the contrast-detection threshold of the lens-emulsion system.

The quotient, visual contrast threshold (under the conditions of viewing) divided by the percentage contrast reduction for a given image dimension, is the object contrast threshold for this dimension of image. Assuming conservatively a visual threshold of 0.1 the data from figure 5.2 are taken to show the detection thresholds of these lens-emulsion systems. These are presented as figure 5.3. It is clear that this concept breaks down when the image becomes effectively a point. In the same figure the systems are compared for the case of symbols occurring at regular spatial positioning (resolution charts). It should be noted that this function for periodic symbols approximates that for the case of isolated dark lines on a bright field. In this latter case we have assumed that the visual contrast threshold varies linearly with the reciprocal image size from 0.1 at 2 symbols per millimeter to 0.4 at 22 symbols per millimeter. The 0.4 threshold appears in reasonable agreement with laboratory measures. The validity of the assumptions bears little on the main point. It is clear that there is a marked difference in the lens-emulsion performance for the two types of symbols.

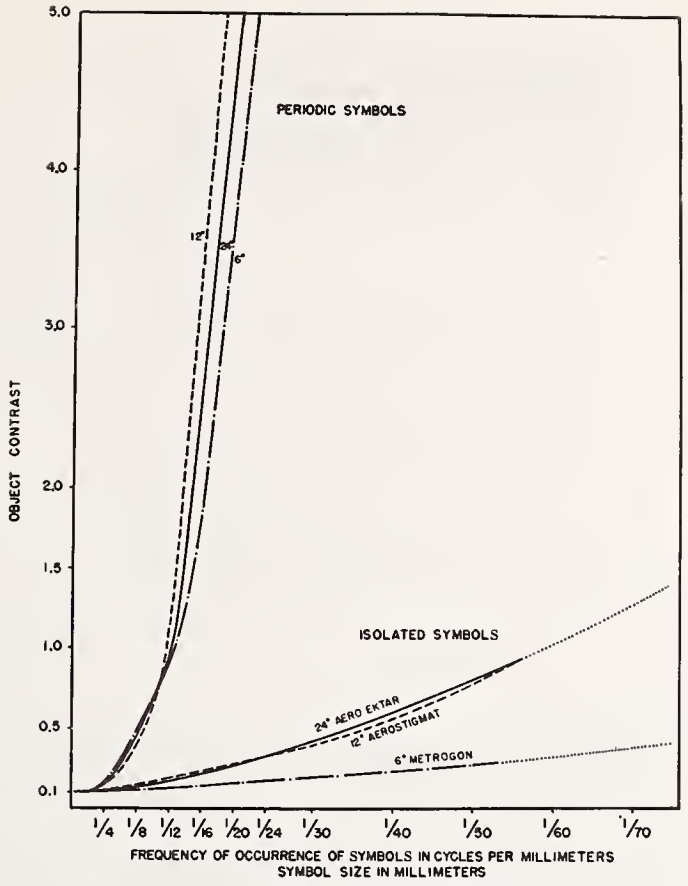


FIGURE 5.3. *Object contrast thresholds.*

The curves shown in figure 5.3 may be taken as representative of the maximum contrast capacity of the lens-emulsion combination for the typical aerial scene, as background and object are located on the toe and straight-line portions of the photographic characteristic, respectively. The test-object brightness ratio is of the order of 30:1.[3]

Let us now assume for purposes of discussion that the object space consists of a universe of isolated symbols. Figure 5.3 describes the response of a lens-emulsion system to this type of detail. We may assume these symbols are distributed such that for any given region $(\Delta x)(\Delta c)$ the same number of symbols occur. Then, in a single exposure the system explores a portion of the detail universe within the area Δ,

$$\Delta = kC_R(f\theta - x_L), \tag{1}$$

where C_R is the contrast range in the scene,[4] f the focal length, θ the angular coverage of the system, x_L the limit of detail size passed by the system, and k a constant that includes a reciprocal scale factor.

[3] The maximum brightness range encountered in an aerial scene, that is, highlight to shadow, is about 150:1. This maximum range is reduced by atmospheric haze to less than 30:1 by only a few hundred feet of air path.

The data employed for this presentation have been taken from film processed to a gamma of 0.98 ±.02. If the sensitometric data are known, it is straightforward to predict the alterations in the form of the curves (fig. 5.3) for any changes in median exposure level for other conditions of processing, or for any specific object if its photometric characteristics are known.

[4] $C_R = C_0\epsilon^{-\beta h}$, where C_0 is the maximum contrast in the object, β the attenuation coefficient, and h the flight altitude.

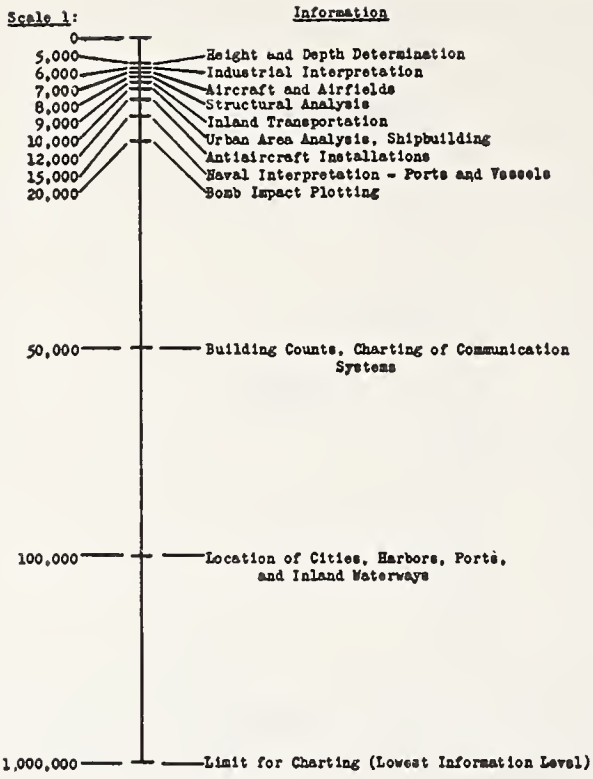


FIGURE 5.4. *Information-level diagram for aerial photographic interpretation.*

It should be noted that the system may be tuned with respect to the detail universe. The tuning is accomplished by varying the altitude of flight, that is, by varying k, which enables the system to explore any preselected range of detail size.

If the contrast threshold of a particular lens-emulsion system (fig. 5.3) is described by $\psi(x)$, then the probability P that any particular unit of detail within the region will occur above the threshold may be written

$$P=\frac{\Delta-\int_{f\theta}^{x_L}\psi(x)dx}{\Delta}. \tag{2}$$

P may then be thought of as a measure of the range of symbols subject to detection through the use of the system.

At the present time, the type of significant detail for a given form of reconnaissance is of such a nature that military photointerpreters state their requirements for photographs solely in terms of scale. Whatever the significant quality criterion may be, the interpreter has achieved, through his experience with many analyses, a concept of a quality criterion that is meaningful to him. In limiting his requirement to a statement of scale, the interpreter tacitly assumes that the other quality aspects will not differ greatly from the present-day average. Thus, it appears that in military aerial photography, as in many other kinds of photography, levels of information are quantized, that is, changes in emission of information from aerial photographs appear to occur at discrete scale levels. This quantized aspect of the aerial photographic problem, as shown in figure 5.4, should not be

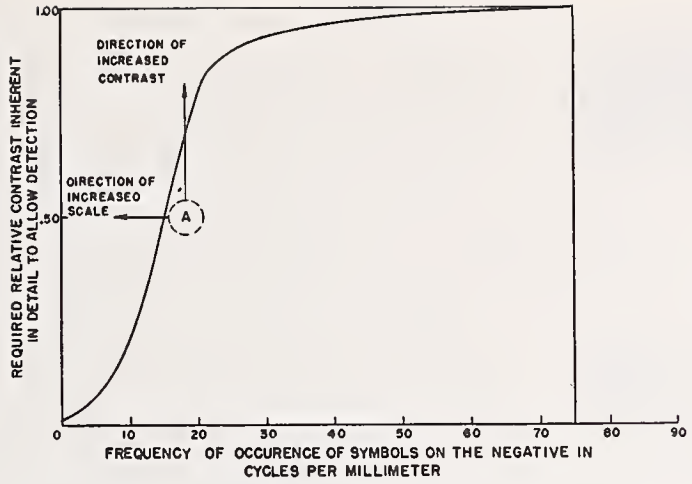


FIGURE 5.5. *Location of significant symbols, "A", in a photograph of not quite adequate quality for a particular purpose.*

taken to be more than an illustration of the gross character of information levels. These levels really define band limits and not discrete lines.

It is more efficient, for all types of intelligence information, to work near the upper limit of the allowable scale, since decreasing the scale reduces the area coverage per exposure. It is not, however, without sound basis that scale has evolved as a quality criterion. Aerial photography has been restricted, at present, almost totally to the use of Super-XX emulsion for work in the visible spectrum. A limiting ratio, detail size/grain size, thus has a constant denominator. An improvement of the detection characteristic for a given system is thus achieved by increasing detail size. This same conclusion holds in comparisons between present systems because of the emulsion dominance of the contrast-reduction characteristic in these systems. It has been suggested that those symbols of significance to a particular photointerpretation task must be so located as to have a high probability of being above the incident contrast thresholds [6]. Thus, one might expect that very similar distributions in size and contrast apply to those symbols that are significant for many types of interpretation. In photographs that are not quite adequate for a particular job these symbols, as illustrated by A in figure 5.5, are submerged below the threshold. The most common method to assure bringing them above the threshold in the repeat photography is to tune the system by increasing the scale. This translates the threshold-characteristic origin to the right with respect to the fixed position of the symbols in the detail universe. The translation can also occur in the Y direction by refocusing (or a change in processing) to drop the threshold characteristic below the significant details.

An effort has been made to study the statistics of aerial photographic targets and, in particular, to attempt to isolate characteristics of the significant symbols for a photointerpretation task. In one approach to this problem, many of the using agencies of aerial photography were contacted. For example, agencies employing aerial photography for forestry purposes have been asked to submit representative prints of aerial photographs taken for their work, which they found to be

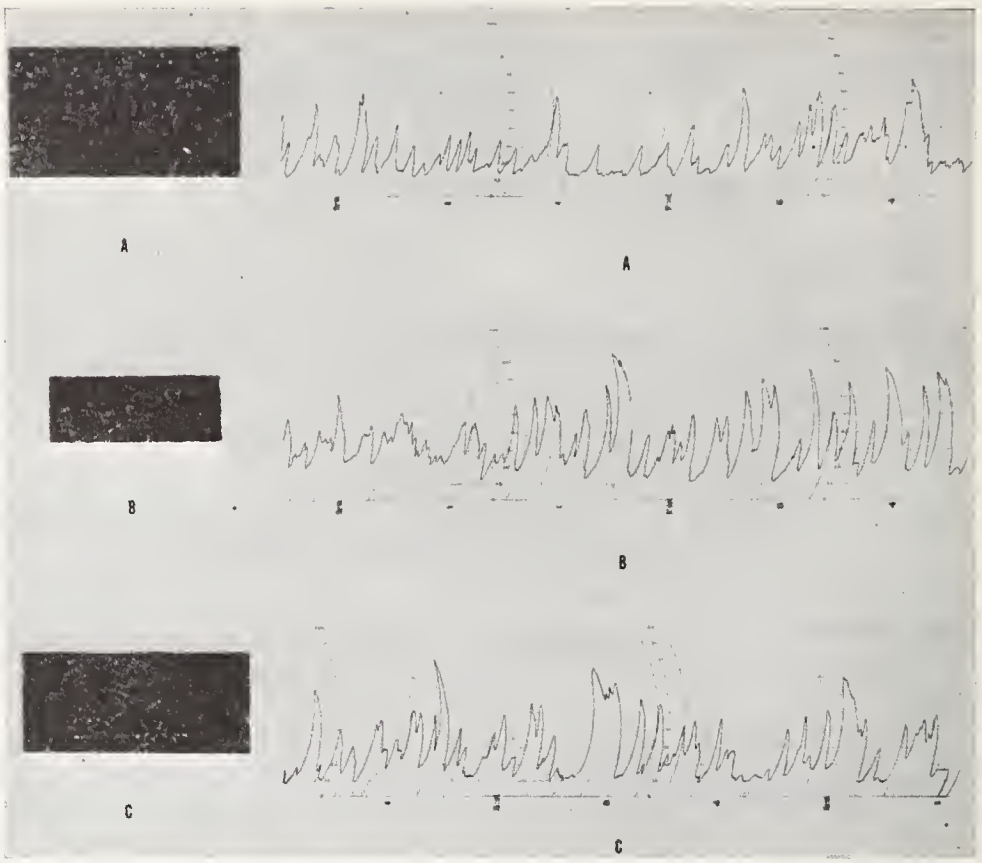


FIGURE 5.6. *Enlarged sections of photographs used to determine classification of forested land as to species, stand height, and stand density.*

Classified by the using agency as: A, Not quite good enough; B, just good enough; C, excellent. (Corresponding microphotometer traces to the right.)

not quite good enough (A), just good enough (B), and excellent (C) for a particular purpose in forestry survey. The purpose of the photography was also requested. By this approach, it was felt not only that the statisties of this type of message eould be found, but that some insight might be gleaned as to the nature of those signifieant symbols that the photointerpreter requires in order to aehieve sueeess in the partieular job. Figure 5.6 shows areas from three prints submitted by the Great Northern Paper Co., Bangor, Maine, and elassified by the user in terms of A, B, and C. Seale of the photography was 1:15,840. The photographs were employed by GNP to determine speeies, stand height, and stand density. Six additional photographs, reported on here, were submitted by the Air Survey's Engineer, Land and Forest Department, Vietoria, B. C., obliques taken with a 3.25-ineh lens from altitudes of 17,400 to 20,000 feet. This photography was used to extend eontrol over large areas of unsurveyed eountry. Analysis was eondueted in the region of a 60-degree depression angle.

Mierophotometer traees were made over three portions of each print. Samples of these traees are shown in figure 5.6, to the right. The scanning aperture was set an order of magnitude below the limit of resolution under normal viewing conditions. Although the seale of the photography differed from set to set, the areas traeed were in all

cases over the image of wooded portions of the terrain. It is assumed by this restriction that this type of photography involved a message with circular symmetry; thus, the resultant plots should be similar for any coordinate of trace.

The analyses were made from the traces. An edge was scored for each case where a change in brightness (reflectivity) equal to or greater than 10 percent occurred (that is, $\Delta B/B=0.1$). A marked change of slope was requisite before the next edge was recorded. The 10-percent threshold value was chosen as representative of a weighted threshold after consideration of the limit of visual contrast threshold on the photographic material at the resolution limit, the range of brightness adaptation levels encountered in the interpretation process, and Blackwell's [7] liminal contrast data. If absolute rather than comparative values were required, it would be necessary to correct the data to allow for the ability of the eye to see edges at lower contrast in the case of larger details.

The close agreement from print to print within any user classification, regardless of scale or purpose of photography, seemed to justify the lumping together of the data from all prints under any one classification.

A plot of the relative frequency of occurrence of edges as a function of distances between them shows no significant difference with picture quality (fig. 5.7, A, not quite good enough; B, just good enough; C, excellent). In this and the following figures each plot deals with nearly 4,000 edges.

The data on reflectivity differences across edges, shown in figure 5.8, indicate a marked difference with picture quality. The better the picture the more skewed is the distribution of reflectivity differences toward the greater differences. From this one infers that the symbols are recorded with greater contrast (that is, with less contrast reduction by the system) in the better pictures.

Figure 5.9 shows the relative frequency distribution of edge gradients, dR/dx, R being reflectivity. Here the excellent photographs stand out, while the difference between A and B categories is slight, the just-usable photographs having relatively more steep-edge gradients than the non-usable pictures.

Additional determinations of the relative frequency of $\Delta R/R$ and dR/Rdx have been made. In each case the excellent photographs stand out from the others as would be expected, but the difference between classes A and B is again less obvious, apparently dealing with some 10 percent of the total information. Coincidentally, this figure is in good agreement with estimates obtained from experienced scientist-photointerpreters who have expressed the feeling that it is the appearance or nonappearance of some 5 or 10 percent of the symbols that differentiates between usable and nonusable photography.

The atmosphere plays two distinct roles in the degradation of the message—haze and turbulence. Atmospheric haze reduces the contrast. Duntley [8] has treated the theory of atmospheric contrast reduction and reports that all attempts to observe the so-called "ground glass plate" effect or the "edge" effect failed. Carman and Carruthers [9] have reported on their experiments treating the brightness characteristics of the aerial photographic message. Because of this fine work there seems to be little need to devote time here to the role of haze in aerial photography. The flight-test data show the expected reduction in contrast with increasing altitude.

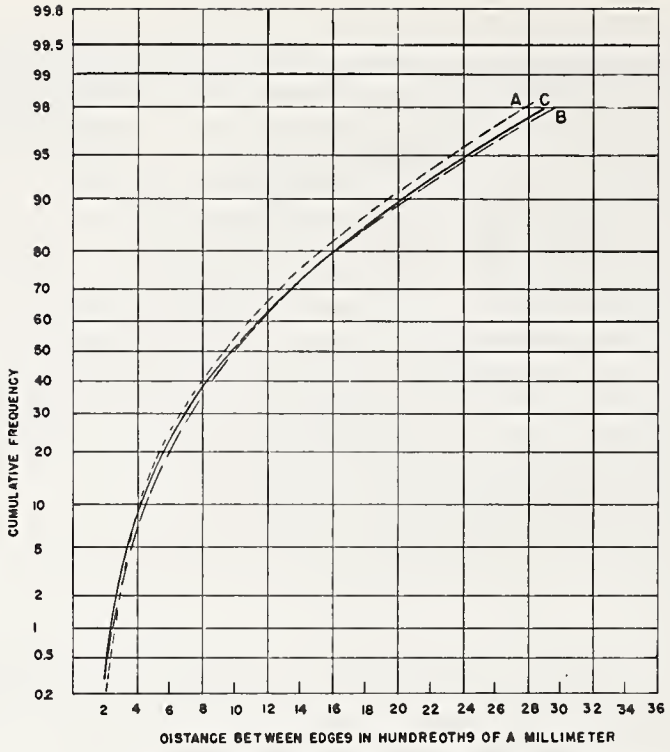


FIGURE 5.7. *Cumulative frequency distribution of distance between edges.*

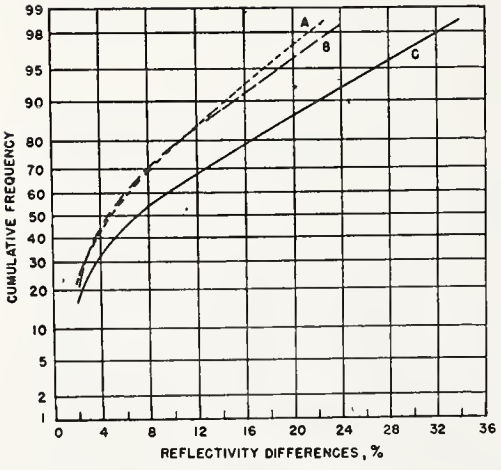


FIGURE 5.8. *Cumulative frequency distribution of reflectivity differences.*

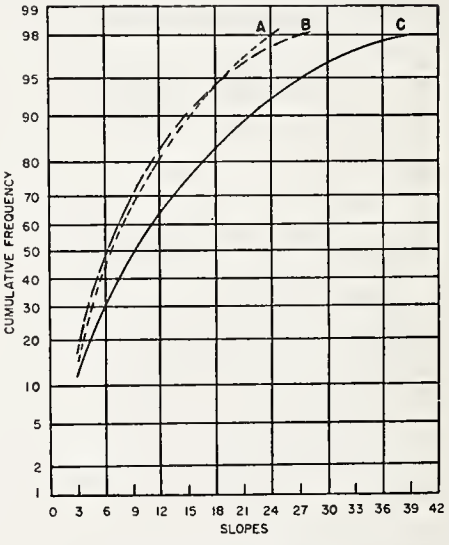


FIGURE 5.9. *Cumulative frequency distribution of slopes.*

On the other hand, air turbulence scatters energy out of the central image, which causes a lower edge gradient and a resultant reduction of image contrast. It appears logical to assume that the limit of definition obtainable in aerial photography will be set by the optical homogeneity of the atmospheric path. The inhomogeneities may be considered to be local motions of small masses of air. These motions are caused by micro-weather conditions and the motion of the aircraft through the atmosphere.

The average condition of the turbulence field introduces effects similar to those caused by inhomogeneities in the lens and window glass, effects which deteriorate the image through increasing flare and diffraction (that is, decreasing the image contrast). This average is not stationary for a number of reasons. One, a nontransient phenomenon, is the migration of vortices and masses resulting in changes in size, number, and density (air density) of the vortices in the turbulence state from moment to moment. These changes are about a mean condition. This phenomenon results in a change in image quality about an average. A second, a transient phenomenon, is the change in the orientation of pressure gradients with respect to the optical axis of the camera. This introduces an image motion during exposure which will, in general, cause a deterioration of the image.

Wind-tunnel tests have shown a deterioration of the image with increasing air velocities in a free tunnel [10]. Other tests which perhaps have more significance here have been discussed in that same report. Aerial photographs have been made over a point source of light, a 1-inch-diameter, krypton-filled, helical-coiled tube of an Edgerton flash lamp, cycled at 60 flashes per second. Photographs were taken with a shutterless camera from an altitude of 6,000 feet in a B–17 at an indicated air speed of 170 miles per hour. Point images selected for study were all located in the same position as referred to the optical axis of the system. Inasmuch as exposure times were in the order of 1/6,000 of a second, aircraft translation and other aircraft motions and transitory turbulence effects can be neglected. Thus, the variations in point dimensions from point to point were assumed to be attributable to the changes in the condition of the turbulence state from instant to instant and to variations in the grain distribution in the photographic emulsion. The point images, when photomicrographed, appeared to consist of a high-density nucleus and a low-density surround. Measures were made of the extent of the nucleus and the surround in two arbitrarily selected coordinates.

It was then necessary to determine which part of this variation of dimensions from point to point was due to the inherent emulsion properties, inasmuch as, at this limit, one is dealing with the effect of individual grains and their distribution. This consideration involved a laboratory test employing the same light source, a moving film magazine, the same lens as employed in the air, and an image scaled down by the same factor as in the aerial test and located at the same position with respect to the optical axis. The resulting distribution of point diameters, both for the nucleus and the surround, is shown in figure 5.10.

The distribution of the laboratory nucleus diameters shown in curve 3 affords some insight to the role of grain distribution in the emulsion. It is clear, as one compares these results with those shown in curve 1 (obtained from an aircraft in flight), that not only is the diameter of the central nucleus increased slightly but the dispersion of the diameters is greater than that obtained in the laboratory, due to the presence of turbulence. Curve 4 treats the laboratory case for the diameters of the low-density surrounds, and curve 2 the same diameters for the in-flight case. From these latter curves it is seen that turbulence causes a spread of the surround energy over a wide range. It is estimated that some 60 to 70 percent of the total energy of the point is in this surround.

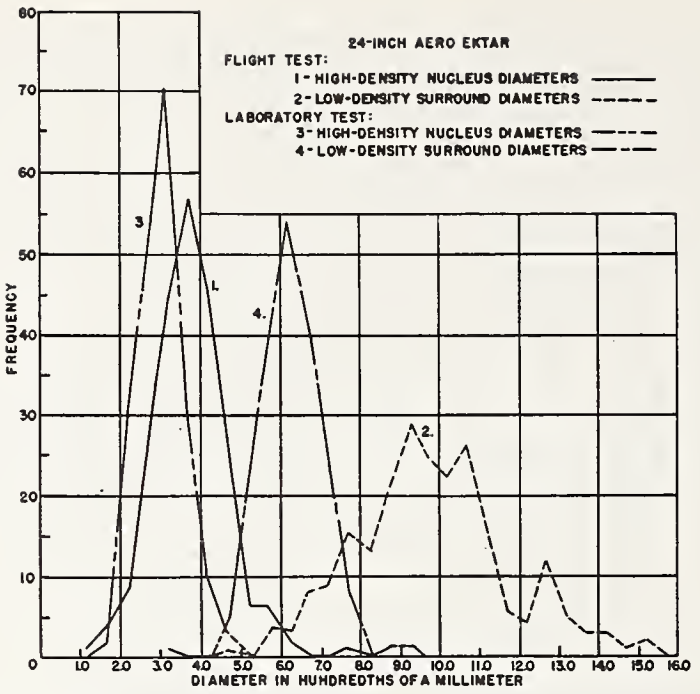


FIGURE 5.10. *Frequency distribution polygons of point-source image diameters on Super-XX emulsions.*

Although lens-emulsion characteristics have been treated previously (figs. 5.2 and 5.3) in introducing the treatment of the message, it is in order to touch briefly on one other aspect of these components. A replot of the data on image contrast, using focal setting as a parameter, indicates a shift in the position of minimum contrast reduction as a function of detail size (fig. 5.11). The conditions of this experiment included full-field illumination. This shift becomes more pronounced as the scattered light is reduced (as may be seen by comparison with previous results obtained with a collimated target [11]) or as one goes to higher-resolution systems.

This is a manifestation of the well-known fact that the focal settings for best resolution and minimum flare (that is, maximum sharpness) do not necessarily coincide. Presented in terms of figure 5.3, this would result in the intersection of the threshold curves for different focal settings of the same lens. It may be stated that focal setting should be selected to provide the minimum threshold, that is, minimum contrast reduction, for the size of symbol under observation. In the case of aerial photography, if the size distribution of the symbols required for operational success be known, then that focal setting which provides the lowest weighted contrast threshold over that size distribution is the best setting. From this it is seen that the system may be tuned in the coordinate of contrast for any given image size by varying the focal setting.

Concerning the photographic emulsion, the concept of microscopic and macroscopic photographic contrast has been discussed previously by Baker [12]. To treat this concept, assume a scattering function

$$i(x)=I_L f(x), \tag{3}$$

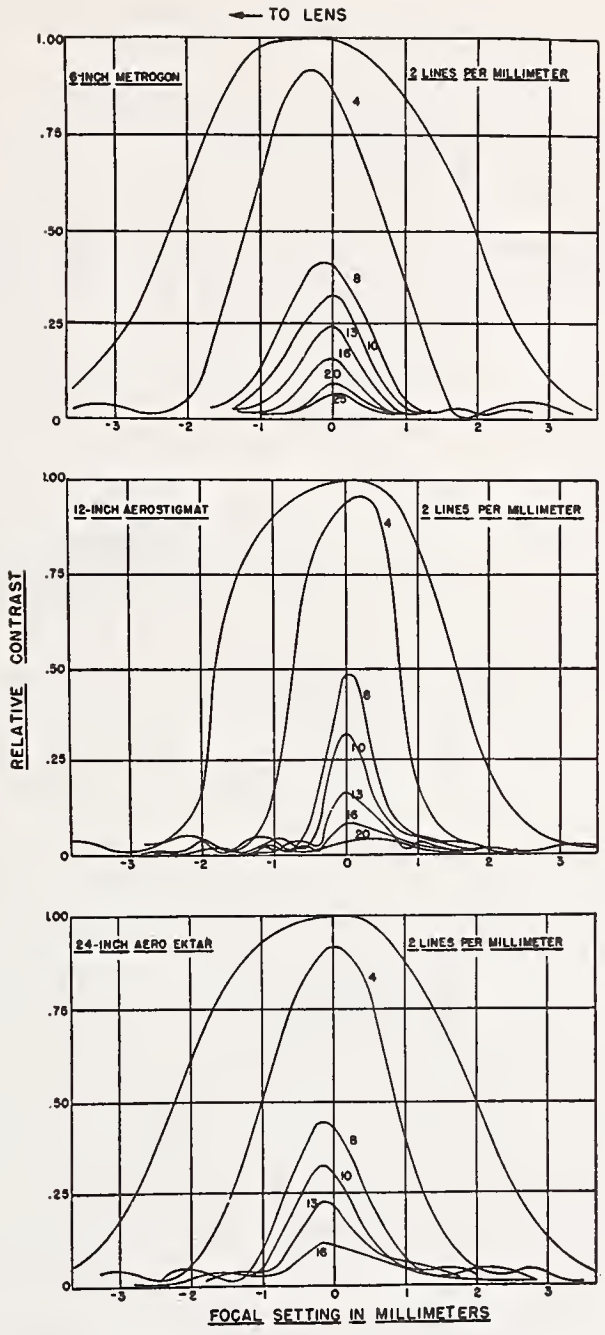


FIGURE 5.11. *Contrast in the image as a function of focal setting, with image size as parameter.*

where $i(x)$ is the intensity at a distance x units from the edge of a line of intensity I_0 as imaged on the emulsion. The intensity that remains in the line image after scattering is I_L. This assumption implies that the intensity over an infinitesimal line of width Δx may be taken as a constant and the fall of intensity regarded as starting from the edge of the line. Figure 5.12a illustrates this condition for a line of width Δx. The dashed line indicates a distribution of intensity as imaged by a perfect optical system. The solid contour indicates the distribution of that energy due to the scattering of light in the emulsion. The

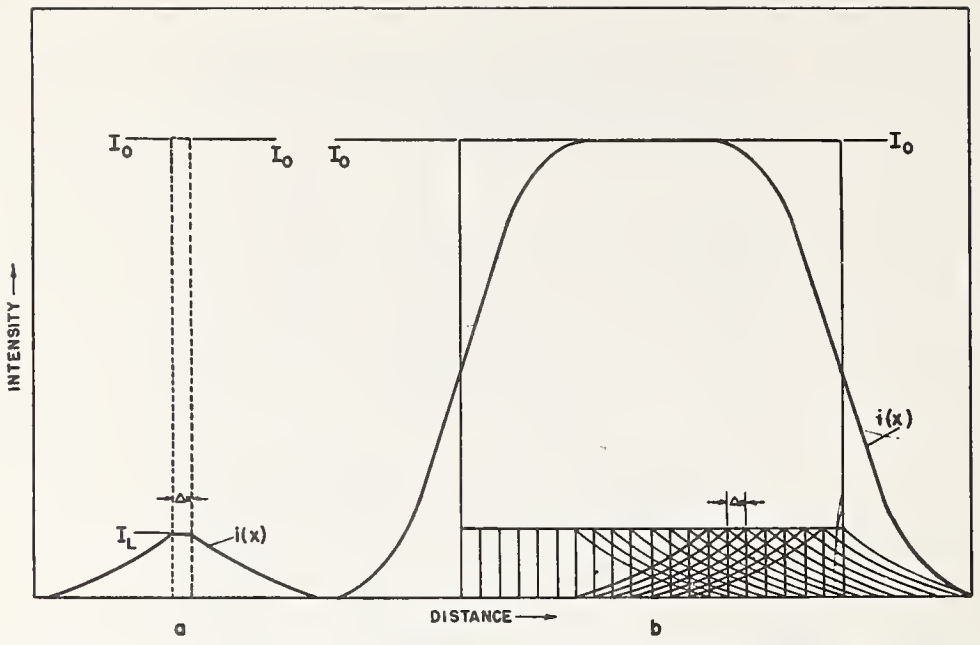


FIGURE 5.12. *Distribution of light intensity from a line assuming exponential scattering.*

intensity remaining in the line, I_L, is considerably less than the incident intensity, I_0.

Figure 5.12b shows the same case for larger detail where the behavior of each line element of width Δx is shown to the same scale as in 5.12a. The ordinates have been added to give the resulting contour of the energy distribution in the emulsion for the gross detail. For a line of this width, $I_L = I_0$ for the central portion. It is seen that there is a critical detail size below which the central portions of the detail image are no longer reinforced by energy scattered from other regions of the detail in sufficient quantity to compensate for the loss of energy from that central portion. Experiments show that for Super-XX emulsion this critical detail size for isolated symbols is about 60 microns. The presence of other symbols in proximity and/or the insertion of any degrading component (for example, the lens) in conjunction with this emulsion increases this critical dimension. The previous statistics, shown by figure 5.7, indicate that some 20 percent of the symbols in the aerial photography examined were recorded below the critical size for the emulsion alone. These symbols are, therefore, no longer in the same sensitometric reference framework, "photographic density: object brightness", as is observed for the macroscopic detail of the photograph.

Redefining I_L as the intensity of the midpoint of the line of width x_0, $\overline{I_L}$ as the average intensity over the line, and taking the origin of coordinates at the point $I=0$ and the mid-point of the line, then

$$I_0 x_0 = \overline{I_L} x_0 + 2 \int_{\frac{x_0}{2}}^{\infty} I_L f(x) dx. \tag{4}$$

Having made the approximation $(\overline{I_L}/I_L)=1$, the intensity I_L may then

be compared to the incident intensity I_0:

$$\frac{\overline{I_L}}{I_0}(x_0)=\frac{1}{1+\frac{2}{x_0}\int_{\frac{x_0}{2}}^{\infty} f(x_0)dx}. \tag{5}$$

The reduction in contrast for symbols occurring at regular spatial frequencies may be treated by an expression of the form

$$\frac{I_L}{I_s}(x_0)=[2f(x_0)]^{-1} \tag{6}$$

and for isolated symbols

$$\frac{I_d}{I_0}(x_0)=\frac{1}{1+\frac{1}{x_0^2}\int_{\frac{x_0}{2}}^{\infty} x_0^2\phi(x_0)dx}, \tag{7}$$

where the subscripts s and d imply space and detail, or symbol, respectively, and where, in going to the case of two-dimensional scattering, the original function $f(x)$ is modified as $\phi(x)$.

The shutter [13], aircraft translation and camera vibration [14] have been previously discussed. They reduce contrast in a manner that may be treated by simple geometric considerations and need not be reviewed here.

To assess the problem relating resolution, scale, and contrast criteria to the detection and recognition of photographic detail, a laboratory test setup has been constructed to provide simulated aerial photography. A report has been made on the method and the preliminary results discussed [15]. It is noted that this experiment concerns itself solely with the information on the photograph as presented to the interpreter. The equipment employs a 20-millimeter $f/1.5$ Biotar lens stopped down to $f/5.6$, used in conjunction with a Fairchild Oscillo-Record Camera. As this camera has provision for moving the film during exposure, it allows opportunity to introduce image motion effects common to most aerial photographs. The camera is aimed in a horizontal direction into a 30-inch-diameter, plane, front-silvered 45-degree mirror located directly over a target array. To vary resolution conditions, a turret head containing 13 ophthalmic lenses and a clear aperture is located in front of the Biotar objective. By this means, 14 different resolution values, from about 4 to 70 lines per millimeter, in approximately equal steps of angular resolution, are obtainable on Super-XX emulsion. It is assumed throughout that this experiment represents photography employing perfect optics used at different focal settings. The validity of the assumption is substantiated by viewing microphotometer traces across the edge of two adjoining photometric areas for each resolution condition (fig. 5.13).

The target array is shown in figure 5.14. The test objects consist of squares and circles. The squares are ¼ of an inch on a side; the circles are of the same area as the squares. Thus, according to Duntley's work [16], the objects are of equal visibility. The objects are of

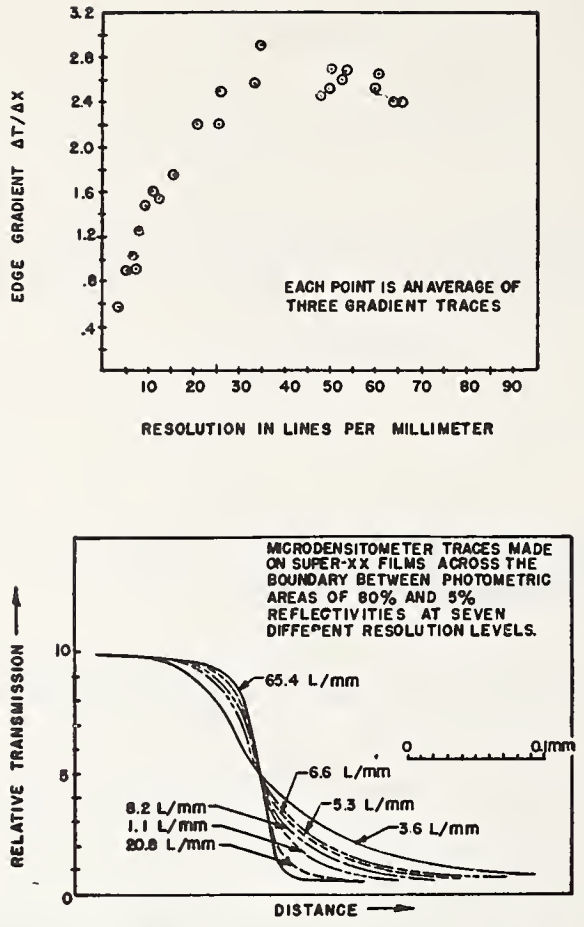


FIGURE 5.13. *Edge traces and edge gradient as a function of resolution.*

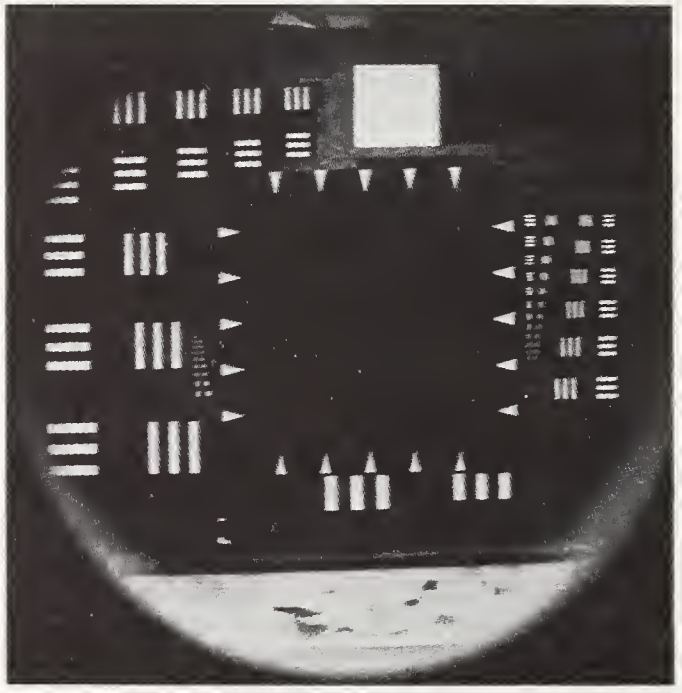


FIGURE 5.14. *The target array.*

three different diffuse neutral-gray reflectivities, 5, 13, and 40 percent, and, in each case, the background for the objects is also one of these same three reflectivities. Thus, six different combinations of object-background reflectivities are available if the zero contrast conditions are excluded.

For scoring purposes the test objects are arranged on a background that is divided into 25 sections, a 5- by 5-unit grid, numbered down the sides and lettered across the top. The observers record the location of the test object by the proper designation, for example, 2C, 4C, 1A, etc. Forced discrimination is employed in all cases. For detection, a single object is located in each of 10 randomly selected positions of the 25 possible. In the analysis the photo reader is then required to indicate the 10 positions that most probably contain objects. For differentiation, or recognition, 12 circles and 13 squares are located on the target, one in each of the 25 positions. The analysis now calls for indicating the 13 positions that contain squares. This work was performed under 7X viewing with binocular microscopes. Throughout the work, the dimensions are normalized in such a way that the test object is taken as a unit square. Photographic scales are then recorded in terms of object dimension, and resolution is expressed in terms of lines per object. No generality is lost by this scaling of test objects. It is only necessary to observe in any transformation that the ratio of grain size to the detail image size be held constant. It is thus possible to translate these results to any desired dimension only by applying the same factor to both the object dimension and the photographic scale.

The form of a typical object differentiation plot as a function of resolution is shown for the scale of 1:3700 in figure 5.15. Incident brightness ratio is the parameter. As all films have been processed to a gamma of 0.98 $\pm$.02, the image contrast, as presented to the observer, may be tabulated for the case of macroscopic detail. The loss of contrast due to decreasing detail size and/or reduced resolution may then be calculated from system functions or measured on the film. The data of this experiment are tabulated below for the macroscopic case.

TABLE 5.1. *Contrast readings made at a resolution of 35 lines per millimeter on the negative and at a scale of 1:3700*

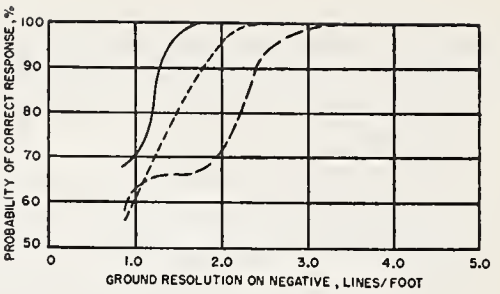
Reflectivities		Contrast
Object	Background	
%	%	
40	5	5.8
40	13	2.4
13	5	1.0

It is apparent from the experimental evidence at hand that count or detection is insensitive to resolution and, in fact, may be considered by treating only the energy distribution in the image and noting if sufficient contrast occurs to bring the object above the visual contrast threshold. Expressed in terms of angular resolution α, the number of objects per resolved spacing, the detection threshold on Super-XX emulsion over the range explored appears to follow the form

$$C-a\alpha-b=0, \tag{8}$$

FIGURE 5.15. *Object differentiation as a function of resolution.*

Object reflectivity	Background reflectivity
——— 40%	——— 5%
- - - - - - - - 40%	- - - - - - - - 13%
— — — — 13%	— — — — 5%



where C is the image contrast and a and b are constants. a is about 0.09 and b is the contrast threshold that provides the same probability of detection at peak resolution of the photographic material (that is, same graininess conditions) for the same size symbol viewed under the same conditions.

Preliminary data from the recognition experiment are shown in table 5.2.

TABLE 5.2. *Resolution and contrast on the negative for various probabilities of recognition of isolated unit cubes*

Scale	Emulsion					
	Super-XX			Pan-X		
	Resolution		Contrast	Resolution		Contrast
	Lines/object	Objects/line		Lines/object	Objects/line	
50% probability						
1:2500	2.22	0.45	0.6	1.69	0.59	0.5
	1.67	.60	2.7	1.35	.74	1.5
	1.53	.65	4.5	1.27	.79	3.0
1:3700	2.56	.39	2.6	----	----	----
	1.81	.55	3.0	----	----	----
1:5000	2.00	.50	1.2	5.55	.18	0.6
	1.72	.58	2.0	2.85	.35	1.8
	----	----	----	2.27	.44	2.3
80% probability						
1:2500	3.12	0.32	0.74	2.22	0.45	0.5
	1.78	.56	2.8	1.81	.55	1.6
	1.56	.64	5.0	1.58	.63	3.1
1:3700	2.22	.45	3.6	----	----	----
1:5000	2.85	.35	1.6	3.70	.27	1.8
	2.50	.40	2.5	3.33	.30	2.7
95% probability						
1:2500	5.26	0.19	0.75	3.33	0.30	0.6
	2.38	.42	2.8	2.38	--	1.6
	2.00	.50	4.8	2.22	.45	3.1
1:3700	3.33	.30	4.8	----	----	----
100% probability						
1:2500	7.14	0.14	0.8	5.00	0.20	0.75
	5.00	.20	3.2	2.85	.35	1.6
	4.00	.25	6.0	2.50	.40	3.1

A portion of these data is plotted in figure 5.16. These show image contrast as a function of resolution. Probability of recognition is a parameter. It is clear that the threshold for this form of recognition is sensitive to resolution but, broadly speaking, insensitive to contrast.

It is perhaps of interest to note in conjunction with these results the mechanism of the visual process. Peripheral vision provides for the detection of objects due to contrast (if we limit ourselves to a static gray scale, as in the case of photography). If the stimulus is sufficient, a reflex feed-back mechanism brings the object onto the fovea. Thus, the presence of the object is detected at low resolution and recognized (or studied with the intent of recognition) in the high-resolution portion of the eye.

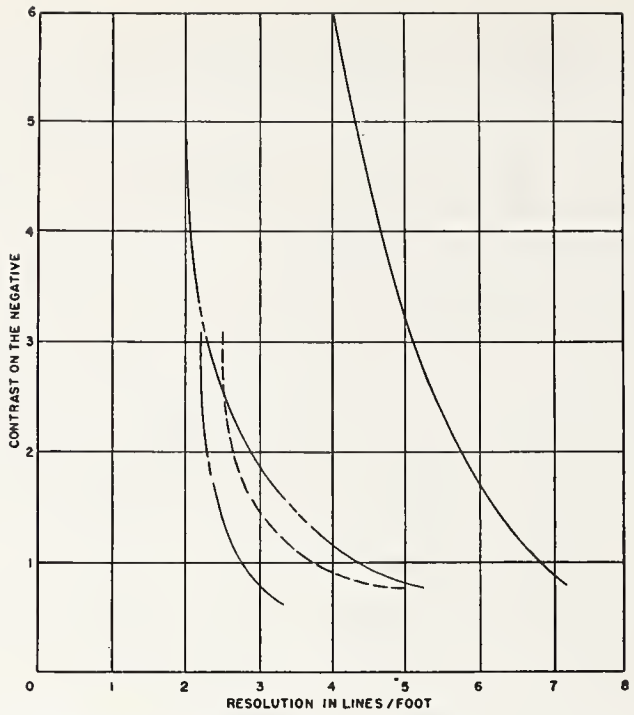
It is possible to compare, by means of figure 5.16, two different emulsions that present the same type of data under the same conditions of resolution or contrast, the only difference being grain. In that class of photography where the purpose of the picture is to reveal information, it appears that consideration should be given to a graininess factor defined in terms of the role of the grain obscuring or obliterating information. In figure 5.16 the effect of graininess is manifest by the shifting of the curves in the direction of better resolution or increased contrast, for any given probability of recognition, as one goes from Pan-X to Super-XX emulsion.

Neglecting the psychological components (experience, recall, etc.) we may regard recognition as a higher-order detection. It is by means of detection of the presence of substructure within a symbol, or by detection of characteristic symbol groupings, or by detection of markings that reveal texture that recognition is achieved.

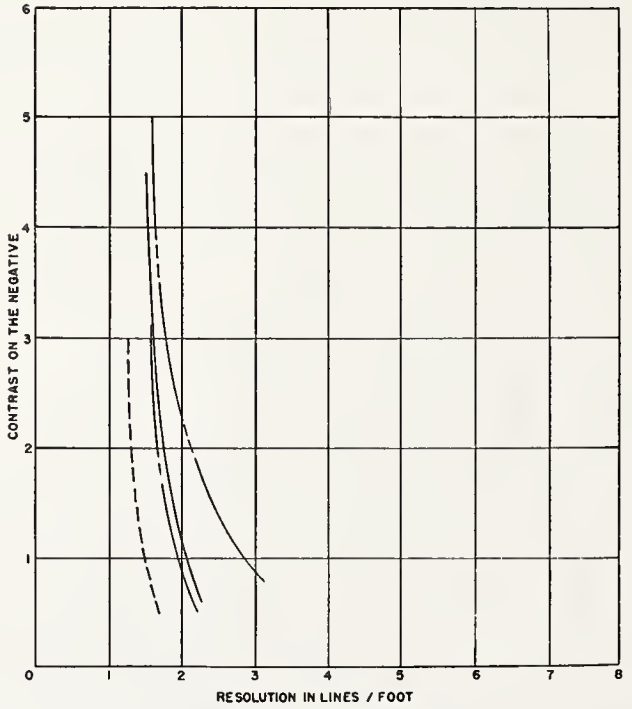
In the experiment described, the detection criteria measure the ability to detect the presence of, that is, to count, symbols. The result is not an absolute threshold for, as in actual interpretation, a search factor is involved. In this same experiment the recognition criteria correspond to detection of substructure. This is related to the decision as to whether a continuity or discontinuity in edge contour occurs. To render this decision the interpreter requires sufficient definition to observe that a finite portion of the object is bounded either by an arc or a discontinuity and/or a straight edge. Most probably the decision is based upon judgments of curvature.

From the particular results shown in figure 5.16, it is seen that when the unit square is imaged at 0.12 millimeter on a side on Super-XX emulsion the significant dimension over which a density difference must be detected appears to be about 15 microns at low contrast and 8 microns at higher contrast. Thus, at these higher contrasts the decisions are rendered by detection of brightness fluctuations occurring over spatial dimensions of the same order of magnitude as those of the individual grain. We may infer that the manner in which the resolution must improve as the image contrast decreases depends upon the symbol shape, and, more important, in the general case, upon the ratio of this significant symbol size to the grain size. From the experiment it is clear that in passing to better resolutions image contrast alone determines whether or not the sub-symbol is detected and, therefore, whether or not the gross recognition is achieved.

For simplicity, let us assume that the photographic interpreter makes his decisions or gains his information through scanning the photograph from blur point to blur point in a manner analogous to an



Scale 1:2500. 100-percent-probability: Super-XX, ———; Pan-X, — — — —. 95-percent probability; Super-XX, — - - —; Pan-X, — - —.



Scale 1:2500. 50-percent probability: Super-XX, ———; Pan-X, — — — —. 80-percent probability; Super-XX, — - - —; Pan-X, — - —.

FIGURE 5.16. *Contrast on the negative as a function of resolution to provide various probabilities of recognition.*

electron beam sweeping the mosaic of an orthicon tube. In this procedure the interpreter faces a basic decision as he proceeds from one blur point to the adjacent blur point: either "yes, the density does differ" or "no, the density does not differ." If there is no detectable change, the interpreter then concludes that there is no change from point to point in the object space as represented by the two image points. On the other hand, if a change does occur, then the interpreter must conclude that there is a change, a boundary, an edge, or a point of a brightness extreme within the corresponding area in the object space. According to the accepted view of the visual process the magnitude of this change is not important, nor are the levels at either side of the boundary. To quote Zoethout [17]:

> When a certain area of the retina is illuminated we are very little concerned with the absolute intensity of the light falling upon this area or the absolute intensity of the resultant sensation; but the ability to discern the difference between the light in this area and that in the neighboring portions of the retina is of prime importance. Upon *visual discrimination acuity* depends practically all our seeing.

Although the magnitude of the change in brightness over an edge is not important in the basic judgments that enter into photointerpretation, the brightness range over which a system can differentiate between small brightness differences enters into any measure of the information capacity of the system. Thus, although figure 5.3 (threshold contrast as a function of image size) shows the relative area of a one-dimensional detail universe that occurs above the system threshold, the one-coordinate case, to be completely represented, must be measured by a volume, the other coordinate being the median exposure level or its equivalent.

The role of the edge in photography is well recognized. Howlett [18] has long emphasized the importance of edges in the photographic problem and has designed his resolution targets with this in mind. More recently, Higgins and Jones [19] have treated subjective picture sharpness and in their work have correlated sharpness judgments with a gradient function.

The conditions of maximum sharpness and minimum flare (maximum symbol contrast) are apparently coincident. However, it is seen in figure 5.11 that, for those lenses examined, no one focal setting provides maximum image contrast for all symbol sizes. This implies that to achieve maximum information the focal setting must be chosen with consideration as to the size of the significant detail in the image. As a consequence, it turns out that the focal setting that provides maximum information in photographs that are studied under unit magnification does not necessarily provide maximum information under any other power.

Therefore, in such cases as aerial photography, when the picture is subject to study under several magnifications in an effort to extract the maximum information, it becomes necessary to consider the distribution of symbol sizes to be recorded. From this distribution it should be possible to determine the focal setting that provides the optimum (weighted) contrast threshold over the format.

Summary

A photographic system is called upon to record symbols of different shapes and sizes and with different spacings and arrangements. These symbols also occur at different contrasts.

This paper considers only a one-dimensional aspect, that is, size not shape. By viewing isolated symbols as well as symbols occurring at regular spatial frequencies two extremes of spacing or arrangement are considered. Because the present photographic emulsion dominates the contrast-reduction characteristic, the threshold curves will be of the same general shape for all present aerial photographic systems.

The statistics of certain photographic messages were analyzed; photographs were classified by the using agencies as not quite good enough, just good enough, or excellent, for a particular purpose. These photographs were taken of natural objects to lend validity to the assumption of similar distributions of distances between edges and inherent contrasts in the object space. The prints were analyzed and similar normalized distributions in the x-coordinate were observed on all pictures. On the other hand, the distributions of brightness differences across edges were different and related to usability of the photography. The results indicate that the greater the contrast reduction by the system the poorer was the picture as classified by the using agency. The presence of symbols on aerial prints, which are finer than the air-borne resolution generally recorded even with high-contrast, line-pattern targets, indicates that the nature of the detail universe is not comparable to that of our target structure. From these analyses there is an indication that the statistics of the recorded messages are quite similar, although the target and/or purpose may differ. This consideration also involves cursory examination of photographs for urban analysis.

The contrast-reduction function of the lens-emulsion system has been interpreted in terms of the object contrast that is required to render a detectable symbol on the emulsion for the two spatial arrangements considered. It is held that the shape of this threshold curve is important in determining the ability of the system to record symbols. Until a statistical weighting can be applied to the size, shape, and spatial distribution in the object space, probably no more meaningful expression of the system capabilities can be given than through an interpretation of eq 2, which considers a weighted mean performance over the format

$$P=\frac{\Delta-\int_{f_\theta}^{x_L}\psi(x)dx}{\Delta}.$$

This expression may be considered as giving the probability that any symbol in a random detail universe will be located above the system threshold. It becomes then a measure of the capacity of the system.

When symbols are in close proximity in the object space (approximated by the data for symbols occurring at regular spatial frequencies) it is improbable that these symbols possess sufficient inherent contrast to be recorded at the high-contrast resolution limit of this system. This fact, coupled with the observation that peak-contrast rendition in the image does not occur at the same focal setting for all image

sizes, implies that judgment of relative performance of aerial photographic systems on the basis of high-contrast resolution scores in the laboratory is not infallible. It further emphasizes that the best operational focus cannot be determined in the laboratory from a maximum resolution setting on a high-contrast target.

The experimental work on the psycho-physical detection-recognition criteria on the emulsion, as a function of resolution, scale, and contrast indicates that the fundamental problem is providing sufficient contrast in the image to permit its detection. For a given shape and size of symbol a certain minimum level of sharpness is requisite. At any point above this level of sharpness, however, the criterion is again one of contrast.

In figure 5.16 it is seen that at a given image contrast the same level of recognition may be achieved on Pan-X emulsion at a lower resolution than on Super-XX. This fact is introduced to point out the significant role that graininess plays in obscuring photographic information and the resulting need for directing more attention to this important factor. With the emphasis directed on achieving a large resolution number we may have overlooked this factor and others of equal significance that have major bearing on the information that the picture can reveal.

---

[1] W. R. Dawes, Mem. Roy. Astron. Soc. London **35,** 158 (1865–66).
[2] Lord Rayleigh, Phil. Mag. **8,** 266 (1879).
[3] O. H. Schade, Electro-optical characteristics of television systems (in four parts), RCA Rev. **9,** 5–37, 245–286, 490–530, 653–686 (1948).
[4] W. Zschokke, Phot. Korr. **36,** 131 (1899).
[5] A. Wetthauer, Z. Instrumentenk **41,** 148 (1921).
[6] D. E. Macdonald, Calibration of survey cameras and lens testing, Photogrammetric Eng. **17,** 383–389 (1951).
[7] H. R. Blackwell, Contrast thresholds of the human eye, J. Opt. Soc. Am. **36,** 624 (1946).
[8] S. Q. Duntley, The reduction of apparent contrast by the atmosphere, J. Opt. Soc. Am. **38,** 179 (1948).
[9] P. D. Carman and R. A. F. Carruthers, Brightnesses of fine detail in air photography, J. Opt. Soc. Am. **41,** 305 (1951).
[10] D. E. Macdonald, A preliminary consideration of air turbulence effects on definition in aerial photography (Boston Univ. Opt. Research Lab. Tech. Note 54, June 21, 1949).
[11] D. E. Macdonald, Calibration of survey cameras and lens testing (Boston Univ. Opt. Research Lab. Tech. Note 54, June 21, 1949).
[12] J. G. Baker and J. S. Chandler, Equlpment for aerial photography, chap. 1, Summary Tech. Rep. NDRC, Div. 16.1 (Optical Instruments) 1946.
[13] A. H. Katz, Camera shutters, J. Opt. Soc. Am. **39,** 1–21 (1949).
[14] D. E. Macdonald, Progress report on aerial camera motions (OSRD Rept. 5178, June 1945). Contractor: Massachusetts Institute of Technology (OEMsr–203).
[15] D. E. Macdonald, Criteria for detection and recognition of photographic detail, Part I—Resolution, scale and contrast conditions for isolated detail, Part II—System performance (Boston Univ. Opt. Research Lab. Tech. Notes 69 and 72, July 10, 1950 and October 15, 1950). Also, paper presented at the Winter Meeting of the Optical Society of America, March 9–11, 1950, Criteria for detection and recognition of photographic detail (Abstract) in J. Opt. Soc. Am. **40,** 258 (1950).
[16] S. Q. Duntley, The visibility of distant objects, J. Opt. Soc. Am. **38,** 237 (1948).
[17] W. D. Zoethout, Physiological optics, 4th ed., p. 167 (The Professional Press, Inc., Chicago, 1947).
[18] L. E. Howlett, Photographic resolving power, Can. J. Research **24,** 15 (1946).
[19] G. C. Higgins and L. A. Jones, The nature and evaluation of the sharpness of photographic images, J. Soc. Motion Picture Engrs. **58,** 277–290 (1952).

6. A Mathematical Model of an Optical System [1]

By Max Herzberger [2]

The new methods of rapid computation (IBM equipment, electronic devices) give the possibility of investigating systematically the validity and practicability of optical methods. The author has started an ambitious enterprise. He is tracing a set of more than a thousand rays through each of three optical systems: (1) An aerial lens with an aperture of $f/5.6$ and a half-field angle of 22 degrees; (2) a wide-angle lens with an aperture of $f/6.3$ and a half-field angle of 37½ degrees; (3) a lens with an aperture of $f/7$ and a half-field angle of 9 degrees.

The first problem is to see whether it is possible to find a mathematical model for the lens in question, i. e., to find a mathematical formula with not too many constants with which object and image rays can be coordinated analytically with an accuracy of a few units of the fifth decimal (for focal length equal to one). The second problem is to find out how many of the thousand rays are necessary and sufficient to derive a function with this accuracy. Figure 6.1 shows that the desired accuracy was obtained.

This mathematical model is used to analyze the optical images. The intersection points of a set of rays equally distributed over the entrance (or exit) pupil, (a) with a set of image planes, (b) with the plane through object point and system axis (the meridian plane), are plotted (fig. 6.2, 6.3, and 6.4).

The plot of these last points, which are called the *diapoints* of the object, lends itself particularly well to a graphical analysis of the system and permits one to investigate the aberrations of each single ray.

In this paper, a preliminary study is made of an aerial lens having an aperture of $f/5.6$. The agreement between the model, the exact calculation, and the photomicrograph of the images is shown, and an analysis of the diapoint diagram is given.

Once a mathematical model of an optical system is given, it can be investigated. In a future paper an attempt will be made to follow up recent ideas in order to compute the influence of diffraction, to analyze the image errors as function of the aberrations at the single surfaces, and to study different approaches to determine the quality of the optical image for the three lenses mentioned in the introduction.

The analytical approach taken in the present paper rests on the mixed characteristic function of Hamilton and Bruns.

Let the object and image origins be put at the entrance and exit pupils of the optical system (in this case, the nodal points). Let the optical axis coincide with the $z(z')$ axis of our coordinate system.

[1] Communication No. 1463 from the Kodak Research Laboratories.
[2] Research Laboratories, Eastman Kodak Company, Rochester, N. Y.

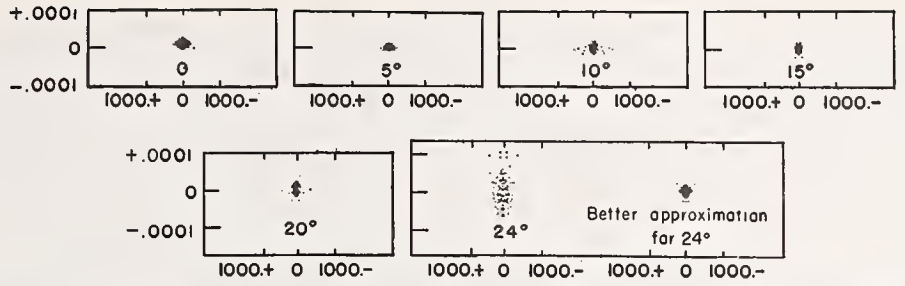


FIGURE 6.1. *Deviation of the data obtained by ray tracing from the mathematical model of the optical system.*

Let object and image rays be given in each case by the intersection point of the ray with the plane $z,(z')=0$, i. e., by $x, y, (x', y')$, respectively. The optical direction cosines may then be designated by ξ', η', ζ' and they will be normed in such a way, that

$$\xi^2+\eta^2+\zeta^2=n^2, \qquad \xi'^2+\eta'^2+\zeta'^2=n'^2, \tag{1}$$

n, n' being the refractive indices of object (image) space.

Since the optical system has symmetry with respect to the system axis, the characteristic function, V, is a function of the symmetric functions of the coordinates, i. e., of

$$u=\tfrac{1}{2}(\xi^2+\eta^2), \qquad v=\xi x'+\eta y', \qquad w=\tfrac{1}{2}(x'^2+y'^2), \tag{2}$$

and for a system of parallel rays with an inclination of 0, 5, 10, 15, 20, 24 degrees from the axis, respectively. V is a function of v and w, and the direction cosines of the image ray are

$$\xi'=\frac{\partial V}{\partial v}\,\xi+\frac{\partial V}{\partial w}\,x'_N, \qquad \eta'=\frac{\partial V}{\partial v}\,\eta+\frac{\partial V}{\partial w}\,y'_N. \tag{3}$$

From (3) the coordinates of the diapoint can be computed, i. e., the intersection point of the rays with the meridian plane, as

$$x'_D=(V_2/V_3)\xi, \qquad y'_D=(V_2/V_3)\eta, \qquad z'_D=-1/V_3, \tag{4}$$

using the abbreviation,

$$\frac{\partial V}{\partial v}=V_2, \qquad \frac{\partial V}{\partial w}=V_3.$$

Let V be given by a fifth-order development, i. e.,

$$V=a_0+a_2v+a_3w+\tfrac{1}{2}(a_{22}v^2+2a_{23}vw+a_{33}w^2) \\ +\tfrac{1}{6}(a_{222}v^3+3a_{223}v^2w+3a_{233}vw^2+a_{333}w^3), \tag{5}$$

and therefore,

$$V_2=a_2+(a_{22}v+a_{23}w)+\tfrac{1}{2}(a_{222}v^2+2a_{223}vw+a_{233}w^2), \\ V_3=a_3+(a_{23}v+a_{33}w)+\tfrac{1}{2}(a_{223}v^2+2a_{233}vw+a_{333}w^2).$$

Inserting these into eq 3, the coefficients a_i, a_{ik}, $a_{ik\lambda}$ can be calculated from a small number of rays traced through the system by the method of least squares. (At least five meridian rays and five skew rays should be traced.)

Figure 6.1 gives the fit of the mathematical model. The system had a focal length of 3, and the deviation for ξ', η', ζ' calculated by (5) from the ray-trace data for the rays up to a 20° field angle is, in general, smaller than 1×10^{-5}. Only for the field angle of 24° did the deviation amount to a few units in the fifth decimal. Introducing a few more coefficients reduced the errors to insignificant size.

The coefficients of the functions are given by table 6.1.

TABLE 6. 1.

	0°	5°	10°	15°	20°	24°
a_2	----------	+1.000005	+0.999999	+1.000008	+1.000216	+1.000776
a_3	+0.326322	+0.325219	+.321972	+.316264	+.307558	+.297747
a_{22}	----------	−.351799	−.346312	−.335021	−.313910	−.278191
a_{23}	----------	−.087948	−.088330	−.091106	−.094766	−.094814
a_{33}	+0.089396	+.091207	+.087397	+.067329	+.050488	+.030552
a_{222}	----------	−.06390	−.07172	+.02139	−.00011	−.03586
a_{223}	----------	−.00680	−.03740	−.06315	−.06494	−.28434
a_{233}	----------	−.39280	−.45960	−.49228	−.58776	−1.09514
a_{333}	−4.15539	−4.26160	−4.44175	−3.88328	−4.41015	−4.72242

Inspection of table 6.2 shows that the coefficients are slowly changing functions of $u=1/2\ \sin^2\sigma$. They can be replaced by

$$a_i=a_i^{(0)}+a_i^{(1)}u+1/2a_i^{(2)}u^2+1/6a_i^{(3)}u^3, \tag{6}$$

thus obtaining V as a function of u, v, w.

When V is obtained, the desired intersection points are calculated for a large system of rays, evenly distributed over the exit pupil (the points were ordered in equilateral triangles evenly distributed over the vignetted exit pupil). Thus, the intersection can be plotted (a) with a set of planes perpendicular to the axis (in this case, three planes were chosen, one through the Gaussian focal plane, and two 0.75 and 1.5 mm, respectively, in front of it); and (b) with the meridian plane (calculation of the diapoints).

Figure 6.2 shows these spot diagrams. On the right will be seen the size of the vignetted apertures, this giving a measure of the light going through the system for different points of the field. The lines y_N=constant are drawn on these apertures, and they contain the points used for the spot diagrams. The symmetry of the problem with respect to the meridian plane makes it only necessary to scan the rays of one-half of the aperture.

TABLE 6.2.

	24°		24°		24°
a_2	+1.000731	a_{222}	+.17443	a_{2223}	−11.1725
a_3	+0.0297783	a_{223}	−.11766	a_{2233}	−4.4855
a_{22}	−.283235	a_{233}	−.58800	a_{2333}	−24.0728
a_{23}	−.096540	a_{333}	+2.62365	a_{3333}	−501.5666
a_{33}	−.009835	a_{2222}	+4.7495		

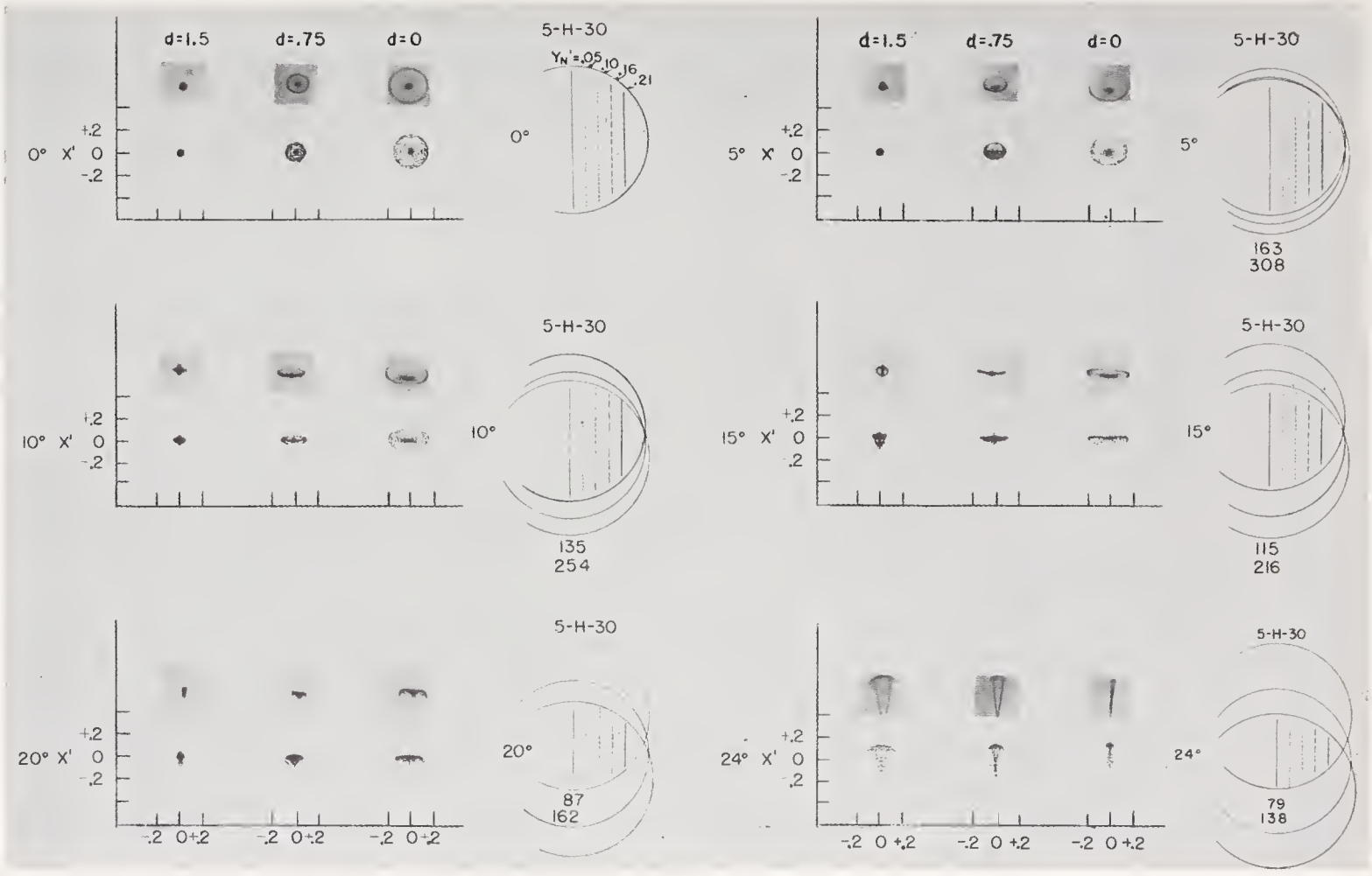


FIGURE 6.2. *Comparison of computed images with microphotographs.*

Figure 6.2 shows the quality of the optical image in detail. Obviously, the best image plane is not the plane through the Gaussian focus, but the plane about 1.5 mm in front of it. The field is very slightly curved forward and comes back at about 20 degrees and then curves rapidly backward. The image at 20 degrees looks slightly more compact than at 15 degrees, and at 24 degrees there is no longer any image to speak of.

Also, Figure 6.2 contains reproductions of photomicrographs taken by L. A. Jones and R. N. Wolfe[3] of these laboratories printed side by side with the spot diagram, showing the agreement between computation and photographic image.

The difficulty of analyzing the spot diagrams lies in the fact that near the plane of best focus there is a heaping of singularities of the wave surface. This is illustrated in Figure 6.3. The image is analyzed by computing the intersection points of the lines, $y''_N = \text{constant}$, going through a set of parallel lines in the exit pupil.

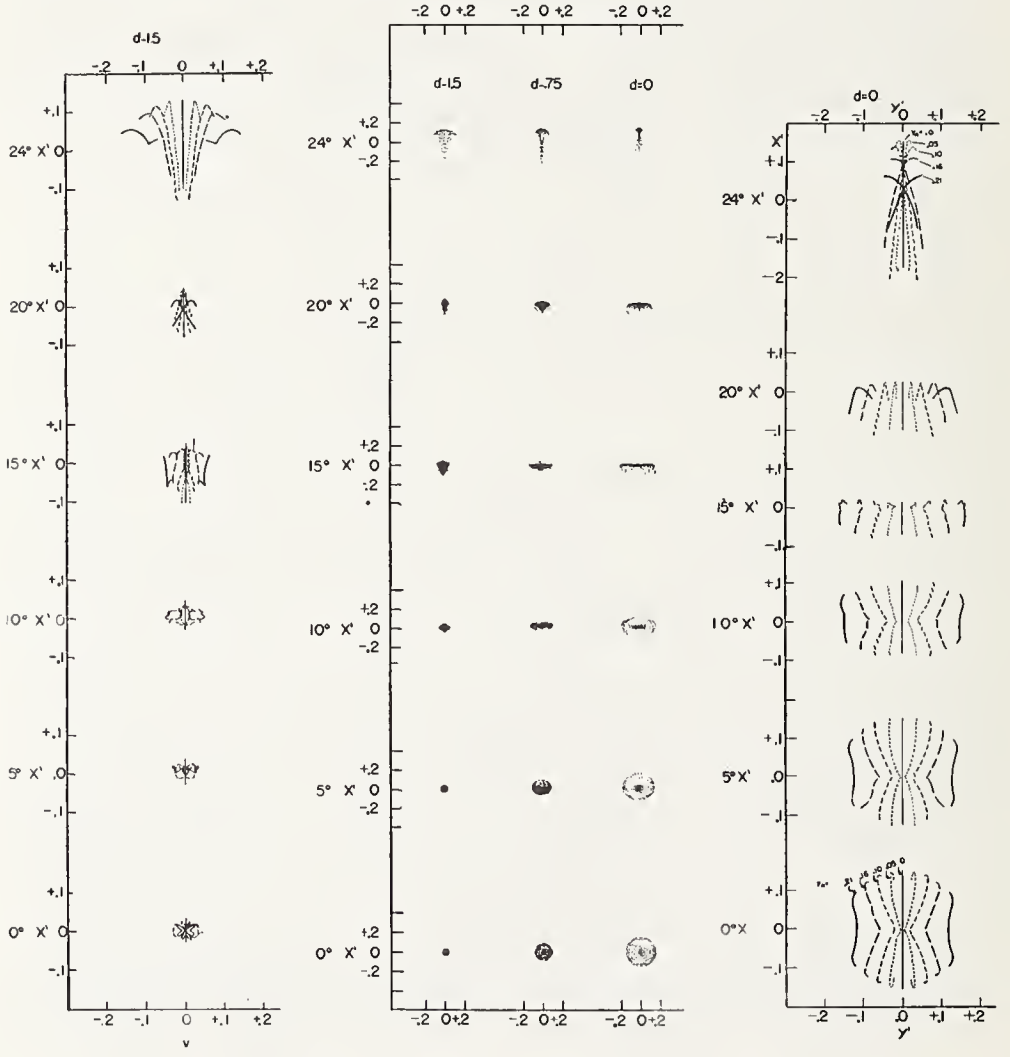


FIGURE 6.3. *Analysis of the optical images obtained by scanning the vignetted aperture.*

[3] L. A. Jones and R. N. Wolfe, J. Opt. Soc. Am. 35, 559 (1945).

The figures are drawn for the Gaussian focal plane and the best image plane 1.5 mm in front of it, and they are four times magnified with respect to the spot diagrams, thus showing where the condensation of light occurs. Inspection of these figures shows that these curves are highly irregular and multivalent, which makes it very obvious why it is so difficult to compute the intersection points in the image plane as simple functions of the direction of the ray and the intersection height in the aperture of the plane.

The author has, for a long time, drawn to the attention of the optical designer the importance of computing the diapoints, i. e., the intersection of the rays (from a given object point) with the meridian plane, i. e., the plane through the object point and the axis (for a meridional ray, the diapoint coincides with the sagittal focus). It can be proven that the knowledge of the diapoints gives complete information about the optical image and, moreover, that the diapoints allow a simple analysis of the image qualities. If the diapoints form a straight line, the image rays form a symmetric image with the line in question as the axis symmetry. The deviation of the diapoints from the best straight line gives a measure of asymmetry.

Figure 6.4 is a plot of the spot diagrams and, on a large scale to

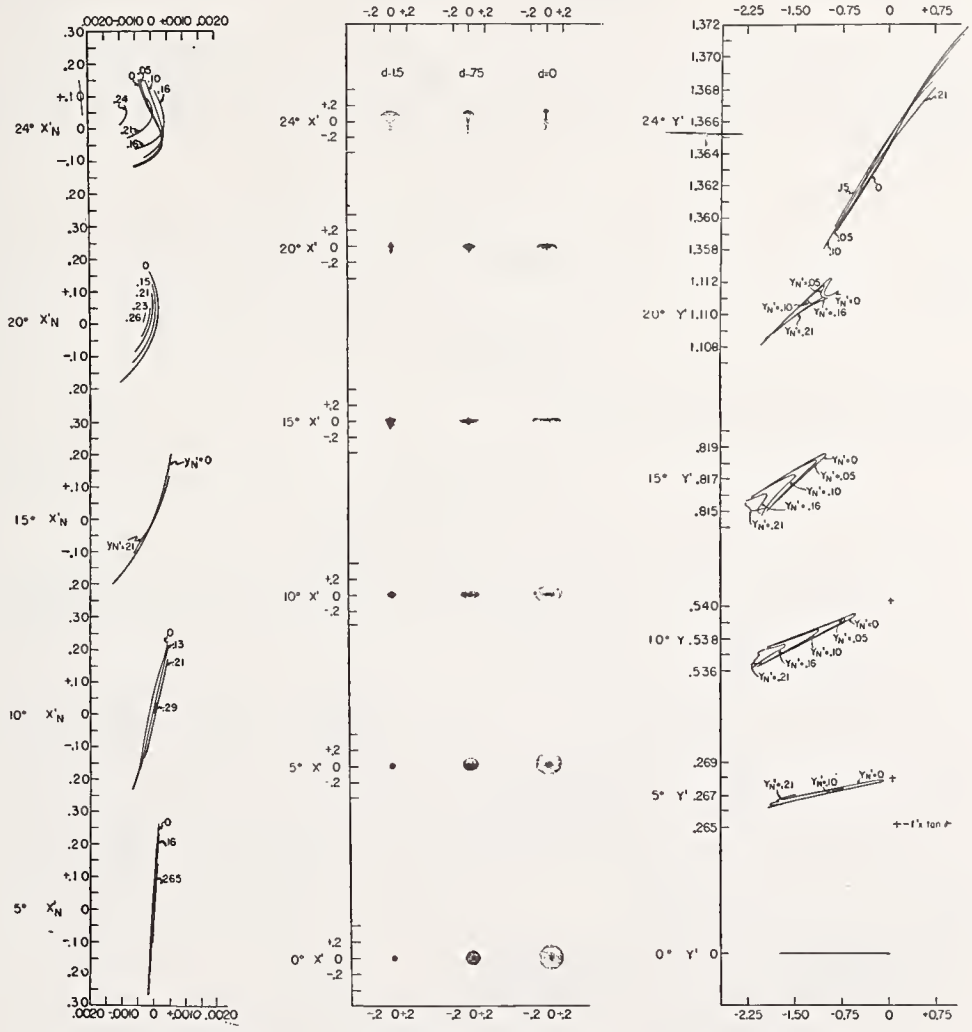


FIGURE 6.4. *Diapoint analysis of an optical system.*

the right, the diapoints for each angle, the lines hereby corresponding to lines $y'_N = \text{constant}$. On the left is plotted for each of these lines the deviation from symmetry. The attention of the reader is called to the fact that these lines are simple curves that can easily be analyzed mathematically.

Even a superficial inspection of these curves shows that the spherical aberration remains practically constant up to 20 degrees, that a slight forward curvature exists which is corrected at about 20 degrees, and not much of an image remains at 24 degrees. These curves show the effect of vignetting, and also that the image at 20 degrees is slightly better, though slightly less symmetric, than at 15 degrees off-axis.

The mathematics of the problem in question is best attacked when the angle characteristic, W, is employed and used as a function of the variables,

$$a=\zeta, \qquad b=\xi\xi'+\eta\eta', \qquad c=\zeta'. \tag{7}$$

The diapoint coordinates then become, for an infinite object point ($a=\text{constant}$),

$$x'_D=\frac{\partial W}{\partial b}, \qquad z'_D=\frac{\partial W}{\partial c}, \tag{8}$$

thus determining W completely. Moreover, distance K of the points of the caustic from the diapoints can be computed by the expressions,

$$n'^2K^2-[\{n'^2(n^2-a^2)-b^2\}W_{22}-2bcW_{23}+(n'^2-c^2)W_{33}]$$
$$+[(n^2-a^2)(n'^2-c^2)-b^2](W_{22}W_{33}-W_{23}^2)=0 \tag{9}$$

with

$$W_{22}=\frac{\partial^2 W}{\partial b^2}, \quad \cdots,$$

and the values of the function

$$W_{22}W_{33}-W_{23}^2 \tag{10}$$

give a measure of the asymmetry of the image.

It will be shown in a future paper how these diapoint aberrations can be split up into the contribution of the single surfaces, according to the methods already described.[4,5]

Diffraction effects also can easily be computed from a knowledge of the characteristic function, V, since V gives, for each point of the exit pupil, the light path for a plane-entering wave. By using Fresnel's integrals, the amplitude and phase of the resulting light vector can then be computed at any point in image space.

Furthermore, the geometrical optical data will be coordinated with the recent attempts to find an analytic measure for the sharpness of an optical image.

---

[4] M. Herzberger, J. Opt. Soc. Am. **38**, 324–328 (1948).
[5] M. Herzberger, J. Opt. Soc. Am. **37**, 485–493 (1947).

The author thanks H. Jenkins, Miss N. McClure, and Miss S. Hall for their help in carrying out these calculations. Special thanks are due to D. L. MacAdam and C. Price for advice and help in setting up the computations for the IBM machines, and to F. Malley, of the Eastman Kodak Co., for putting his IBM equipment at the author's disposal.

7. Methods and Apparatus for Measuring Performance and Quality of Optical Instruments

By A Arnulf [1]

Introduction

From the user's point of view, and for a given point in the field, the value of an optical instrument is completely described by two figures. The first is its performance (Leistung), which, after Löhle, is the number showing how much the instrument multiplies visual sharpness. In France this number is called the "amplification" of the instrument. The second figure is the quality, which, in its simplest form, is the ratio of the performance of the instrument under test to the performance of a perfect instrument with similar characteristics.

Visual Instruments

Theory

Let us first consider the case of visual instruments in which the aerial image is observed with an eye-piece. The work on resolving power relative to perfect instruments, which I reported a few years ago, allows us, first, to find an absolute criterion of the quality of the instrument, and secondly, to find a simple connection between its quality and its performance.[2]

We define the perfect instrument as an instrument in which the distribution of light in the image is determined by diffraction alone. For such an instrument, the resolving power is given by the equations

$$S \cdot \Omega = 2n \sin U \cdot T = s_\omega \cdot \omega, \tag{1}$$

where S and T are the angular and linear resolving powers of the instrument, Ω and $n \sin U$ are the linear aperture and the numerical aperture of the objective, and s_ω is the resolving power of the eye for the diameter $_\omega$ of the pupil.

These equations may be written

$$S \cdot G = T \cdot P = s_\omega, \tag{2}$$

where G is the angular magnification, and P the power of the instrument.

---

[1] Institut d'Optique, Paris, France.

[2] A. Arnulf, Compt. rend, ac. sc. **200,** 52 and 306 (1937); La Vision dans les Instruments, Edit. Rev. opt. (1937).

Instrumental Efficiency

Let us consider an imperfect instrument with the same exit pupil and the same magnification as the perfect instrument. S',G',T',P' are corresponding terms for the defective instrument. We know that $S'>S$. Therefore,

$$S'\cdot G>S\cdot G \quad \text{or} \quad s'_\omega>s_\omega. \tag{3}$$

Furthermore,

$$\frac{S}{S'}=\frac{s_\omega}{s'_\omega}. \tag{4}$$

Therefore, if we compare a defective instrument with a perfect one with identical optical characteristics, the ratio of the resolving powers, or of the performances, is given by the ratio of the resolving powers of the eye in the image field. The quality of the image is then defined by

$$E_i=\frac{s_\omega}{s'_\omega}, \tag{5}$$

which is called *instrumental efficiency* because it depends only upon the effect caused by the defects of the image.

Let us compare the quality of two different instruments, for example, the big telescope in Yerkes Observatory and a low-magnification microscope. The above method will give for each one its instrumental efficiency and the ratio of these two efficiencies will give the relative instrumental efficiency of the two instruments.

We may sum up the principle of this method as follows. The resolving power of the eye in the image field of a perfect instrument is the same as the resolving power of the eye alone with the same diameter of the pupil. The defects of the instrumental image degrade this resolving power, and this degradation determines the quality.

Total Efficiency

The criterion defined above gives the loss of quality compared with a perfect instrument of the same type, and it is practically the most important fact. However, where the user is to be considered, there is a loss of the effective quality caused by the stopping of the pupil, which increases the resolving power of the eye. Therefore, it is necessary to consider this stopping effect in order to compare the quality and the performance. Similar considerations to those previously used show that when we include both the stopping of the eye and the defects of the image, the quality is given by

$$E_t=\frac{s_n}{s'_\omega}=\frac{s_n}{s_\omega}\cdot\frac{s\omega}{s'_\omega}=E_p\cdot E_i, \tag{6}$$

S_n being the lowest resolving power obtainable with the eye, corresponding to a pupil diameter of about 2.5 to 3 mm. This is, for all practical purposes, the resolving power of the eye in daylight. $E_p=s_n/s\omega$ is the efficiency of the stopping process. This can be obtained by measuring visual sharpness.

Radius of pupil (mm)	0.2	0.4	0.75	1.0	2.0	4.0
Pupil efficiency	0.07	0.23	0.50	0.64	0.88	1.0

According to what has been said, one conclusion that may be reached is that an optically perfect instrument might be a visually moderate instrument. This is the case especially with instruments operating at the magnification giving the best resolving power.

Relation Between Efficiency and Performance

In the case of visual instruments, there is a precise relation between the performance and the efficiency. The performance is given by the amplification A, where $A \equiv s_n/S$. This means that

$$A = G \cdot E_t. \tag{7}$$

In this equation G is the angular magnification for telescopes, or the conventional magnification for microscopes or such instruments. For a perfect instrument, $E_i = 1$ and $A = G \cdot E_p$.

Experimental Procedure

Measuring the efficiency requires: (1) The determination of the resolving power of the tested instrument from which is deduced the resolving power of the eye in the image field by using the formula

$$s'_\omega = S\frac{\Omega}{\omega} = S \cdot G,$$

or

$$s'_\omega = T \cdot \frac{2n \sin U}{\omega} = T \cdot P;$$

(2) the resolving power of the eye without the instrument, for one or several apertures of the pupil, which gives the relation to the perfect instrument.

The quality of an instrument is a function of the experimental conditions. Therefore, it is necessary to choose conditions such that the result might be valid in as many cases as possible, and also to duplicate as nearly as possible the conditions under which the instrument is to be used. The conditions of testing once defined will be taken as a constant standard.

1. The test object is seen against a background of uniform brightness, which covers the entire field of the instrument. This is necessary if one is to introduce all the stray light that the instrument might give under use. The effect eventually produced by the sun or moon will be realized by sources situated in precise points in the field.

2. No optical system will be interposed between the instrument and the test object in order to avoid any effect caused by its defects and stray light it might introduce.

3. No telescope is placed behind the eye piece to increase the magnification for the same reasons given in item 2, and also because the increased efficiency makes the test unnecessarily strict and does not duplicate the conditions of use. It is not reasonable to expect the instrument to have a better quality than that given by a perfect instrument with the same magnification.

4. The standard resolution test object is the Foucault test object, or a test object consisting of two parallel lines of infinite length and of varying contrasts. For simplicity one can use a contrast of 1 and a very low contrast, 0.01 or 0.02. Tests with the higher contrast are more sensitive to astigmatism or focusing defect, whereas with the lower contrast they are more sensitive to stray light, spherical aberration, and chromatism. It is desirable to use several brightnesses representing daylight, twilight and night. Night instruments tested under daylight conditions might give results completely different from the desired results.

Description of the Apparatus

Figures 7.1 and 7.2 show the arrangement of the apparatus. The background with a uniform brightness is obtained by placing the nose of the instrument I in a lighted sphere S, and by using several plane screens E_1, E_2 which illuminate the field. The black test object is seen against the last screen.

The variation of the contrast is obtained by using a rotating white disk D, of the same brightness as the backround, and consisting of two similar sectors between which the aperture can be varied while the disk is rotating. The contrast is given by $C=\alpha/360$, (α in degrees), α being the total aperture of the disk.

Furthermore, the angular size of the test object T can be varied continuously by projecting its image on a fixed plane, the objective O and the test object T being moved simultaneously with suitable cams. The apparatus is simple to use with any form of test object (fig. 7.3 and 7.4).

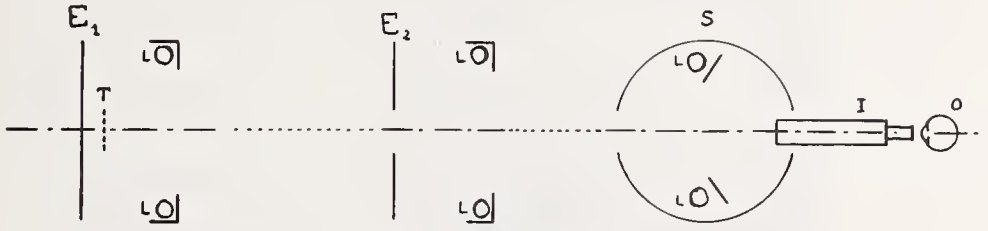


FIGURE 7.1. *Testing apparatus for visual instruments.*
(General arrangement.)

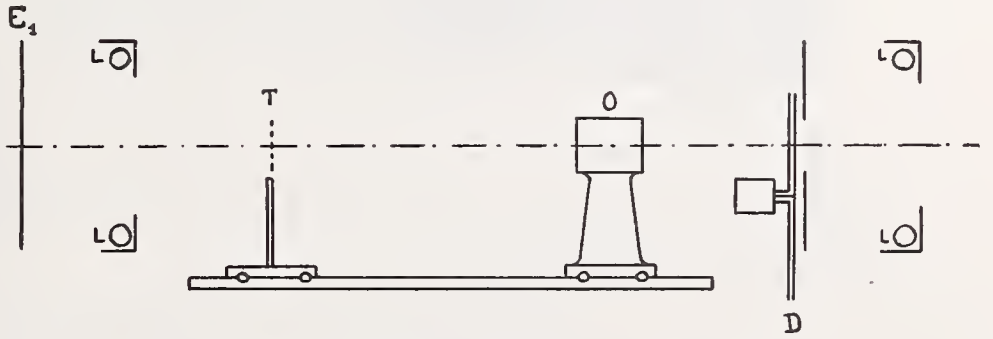


FIGURE 7.2. *Testing apparatus for visual instruments.*
Control of the contrast and the angular size of the test object.

518-743 O - 73 - 21

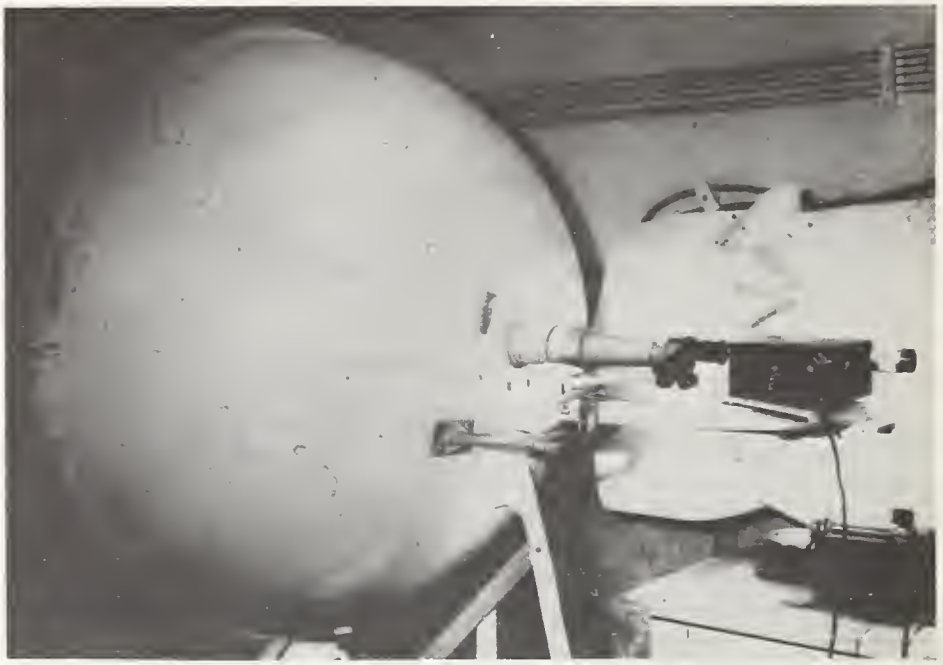


FIGURE 7.3. *Testing apparatus for visual instruments.*
Instrument to be tested and sphere S. For efficiency measurements, the photometer behind the eye piece is removed.

FIGURE 7.4. *Testing apparatus for visual instruments.*
Rotating disk D; test object A; optical bench R; round bright screen for the background.

Standard characteristics for the examination of visual instruments are as follows:

Brightness of the background.

$5 \cdot 10^{-3}$, 10^{-4}, $2 \cdot 10^{-6}$, $5 \cdot 10^{-7}$ 10^{-8}, $5 \cdot 10^{-10}$ Stilb.

The values most used are in italic.

Nature of the test object. Foucault resolution test object, or equivalent test object with two parallel lines.

Contrasts.

1.0	0.5	0.25	0.10	0.06	0.03	0.015	0.01

Generally we use two contrasts, contrast 1.0 and the contrast nearest the threshold for the brightness used.

Points tested in the field. The center, a circle equal to two-thirds of the radius of the field, a circle 2° from the edge of the field.

The instrument is always tested in the same conditions under which it is used, without any accessory optical aids.

Examples of the Effect of Isolated Defects

Decrease in brightness. Figure 7.5 shows the variation of the efficiency with the brightness for an optically perfect instrument with a transmission factor 0.35, using a Foucault test object of unit contrast. The efficiency is related to the slope of the curve of visual sharpness and brightness. It is equal to 1 for daylight and very nearly 1 for night vision (the slope is zero or very small). It is least in the brightness region where visual sharpness varies most rapidly with the brightness, that is, in twilight, and again near the threshold.

Stray light. Stray light produces a decrease in perception for low contrasts, which can be computed in every case if we have previously measured the stray light. Figure 7.6, curve 1, is the curve for an instrument the correction of which is nearly perfect. Table 7.1 gives the values of efficiency versus contrast for two binoculars in the center of the field. Lines A refers to an instrument without blooming, lines B to the same instrument with bad blooming which increases the scattering.

Monochromatic spherical aberration. This effect has been reported by M. Françon for both theoretical and experimental methods.[3] The

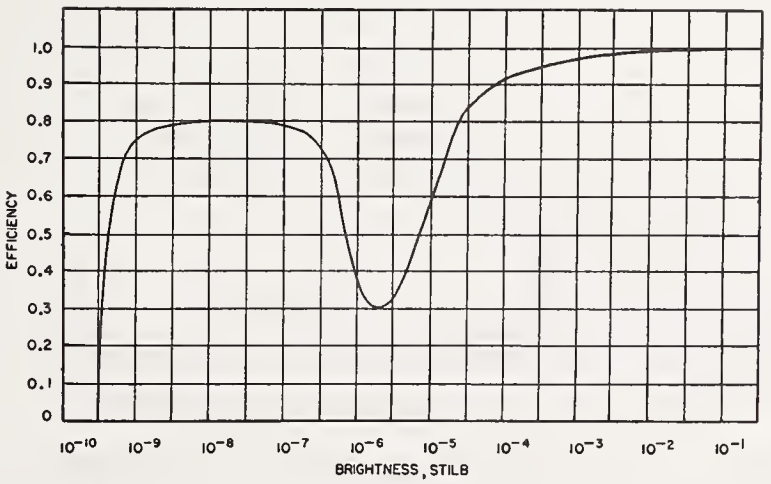


FIGURE 7.5. *Efficiency versus brightness of a perfect instrument having a transmission factor of 0.35.*

Test of resolution with full contrast.

[3] M. Françon, Rev. opt. [10] **26**, 354 (1947); [3] **27**, 157 (1948).

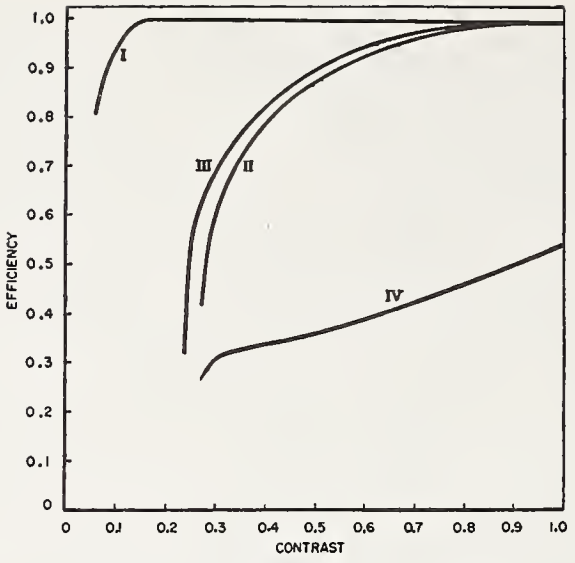


FIGURE 7.6. *Efficiency versus contrast.*

TABLE 7.1. *Efficiency versus contrast for two binoculars, before and after a diffusing blooming, showing the effect of stray light*

(Center of the field) Lines A: Before blooming. Lines B: After blooming.

Contrast	1	0.6	0.3	0.1	0.08	0.05
1{A	0.73	0.91	0.97	0.98	0.97	0.90
1{B	.73	.90	.91	.91	.15	0
2{A	.85	.85	.95	.95	.85	0
2{B	.85	.85	.94	.60	.25	0

curve (fig. 7.7) gives, as a function of the third-order spherical aberration given in phase differences, the values of efficiency for various contrasts in the test object, when the exit pupil of the instrument corresponds to the best resolving power (magnification giving $\omega=0.6$ mm). We can see that for small contrasts and the Rayleigh tolerance ($\lambda/4$) the fall in efficiency is considerable. Figure 7.8 shows experimental results obtained with a contrast of 0.03, varying the diameter of the pupil of the eye. An aberration corresponding to the Rayleigh tolerance has no effect on pupils larger than 3 mm. For smaller pupils the efficiency decreases quickly. Figure 7.6 III and IV shows the results for a very large spherical abberration.

Astigmatism. The results in table 7.1 are those of the very common case of an instrument with a little astigmatism. The efficiency is better for medium contrasts than for high contrasts. Figures 7.9 and 7.10 show how the resolving power varies with the diameter of the pupil when astigmatism is present, and this explains the above results.

Chromatism. This question is being studied both theoretically and experimentally. Curve II figure 7.6, shows the variation of efficiency with contrast for a very large chromatism (aperture f/12, no chromatic correction).

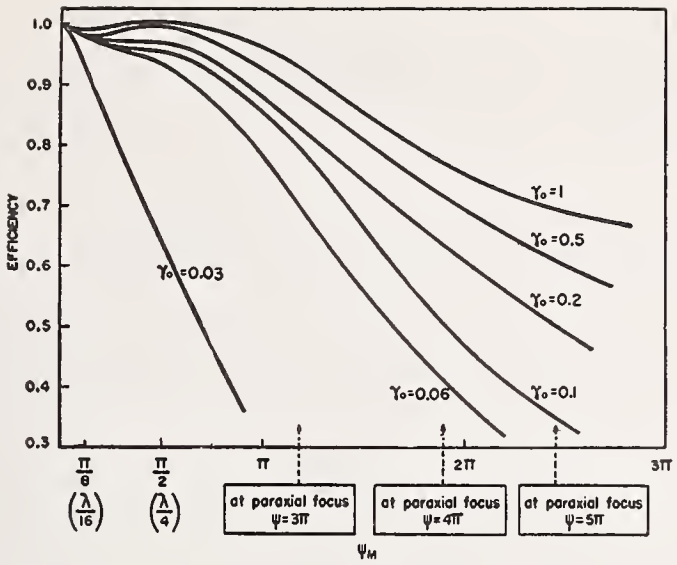


FIGURE 7.7. *Efficiency versus spherical aberration expressed in phase differences for various contrasts.*

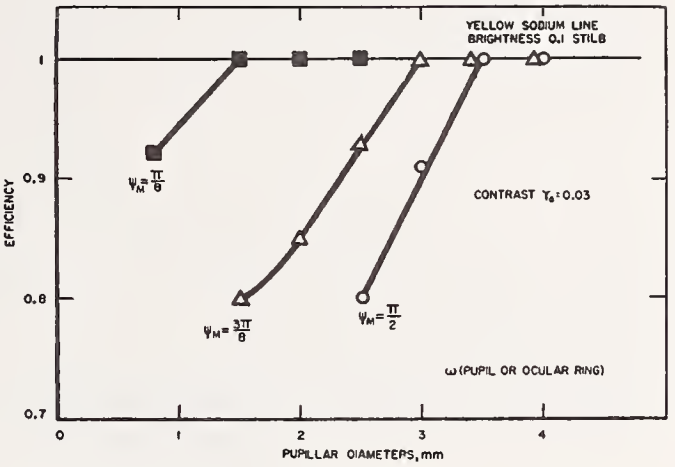


FIGURE 7.8. *Efficiency versus pupillar diameters for various amounts of spherical aberration expressed in phase differences.*

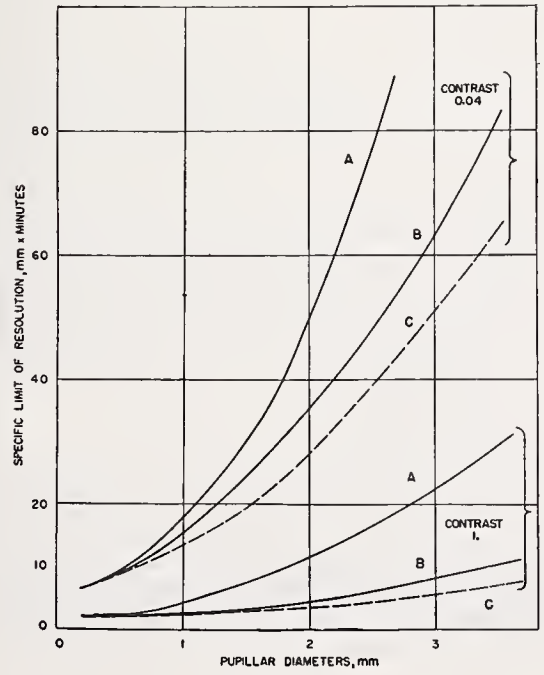


FIGURE 7.9. *Specific limit of resolution of the eye versus the pupillar diameter.*

A, without astigmatism; B, astigmatism (4 diopters, parallel to the lines of the test object); C, astigmatism (4 diopters, perpendicular to the lines of the test object).

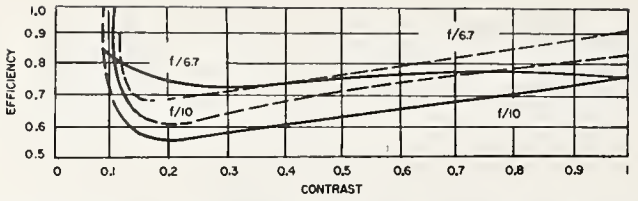


FIGURE 7.10. *Efficiency versus contrast of photographic objectives*
Solid lines, lens with third,order spherical aberration; dotted lines, well corrected lens.

Quality in the Whole Field of an Instrument

The question arises as to whether it is possible to evaluate with one number the quality of an instrument over its entire field. Two very different cases occur.

1. *Instruments for pure observation.* These are used to examine minutely, details in objects seen in the field of the instrument. In this case we are interested in the resolution of details in the center of the field. It would not be reasonable to include in the measurement the edge of the field when it does not enter into the examination of details. This instrument will be characterized by the area within which the efficiency is greater than 0.9 times the efficiency in the center.

2. *Instruments to discover objects.* In this case the periphery of the field has the greatest importance, as is the case for the periphery of the retina, the efficiency in foveal vision having practically no importance. The efficiency to be introduced is the efficiency that corresponds to a point on the retina situated at a distance from the fovea equal to the field of the instrument, and the whole efficiency of the field will be equal to the mean peripheral efficiency for the whole field.

Instruments With Diffusing Screen (Photographic Emulsion)

Theory

The method that is described here has been completed by Mme. Marquet.[4] Similar methods have been studied by various other workers.[5]

The principle of the method is the same as for the visual case, but instead of the angular resolving power of the eye, we have here the linear resolving power of the eye-emulsion combination, the optical image of the object being focused on the emulsion with a perfect objective of variable aperture. If S and T are the resolving powers of the instrument, Ω and $n \sin U$ the linear or numerical apertures of the objective in the object space, $n \sin u$ the numerical aperture in the image space for a perfect instrument, we have

$$S\Omega = 2n \sin U \cdot T = 2n \sin u \cdot t.$$

The total efficiency E, which is similar to the total efficiency in

[4] M. Marquet, Sciences et Industries Photographiques, [2]*18*, 129–142 (May 1947).
[5] A. Couder, Cahiers phys., No. 14, pp. 35–48 (May 1943); Sciences et Industries Photographiques,[2] **14**, 170–174 (August 1943). L. E. Howlett, Can. J. Research (July 1946). E. W. Selwyn and J. L. Tearle, [B] (Sept. 1946).

visual instruments, and is used in the calculation of the performance, is given by $E_t = t_\infty / t$, where t_∞ is the resolving power with a perfect objective, the numerical aperture of which is very large. t_∞, being independent of the aperture, is called the resolving power of the emulsion.

The absolute efficiency introduces two kinds of losses in the resolution

1. The increase in resolving power of the perfect instrument as it is stopped down, because the diffraction pattern becomes more and more important in comparison with the area of granularity-diffusion in the emulsion. We evaluate it by the aperture efficiency $E_p = t_\infty / t_0$, where t_0 is the resolving power of the emulsion for a perfect instrument with the same aperature as that of the tested instrument.

Let us note that for ordinary emulsions, $t_0 \sim t_\infty$ for small apertures, 2 sin $u = 0.1$, for example. In the case of photographic lenses, the aperture efficiency is rarely important. However, it is always important in the photographic microscope.

2. The decrease in resolving power caused by all the defects in the image produced by the instrument is given by the instrumental efficiency $E_i = t_0 / t$. This is the ratio between the resolving power t_0 of the emulsion with a perfect instrument, the aperture of which is the same as the aperture of the tested instrument, and the resolving power t of the emulsion with the tested instrument.

3. We have then, $E_t = E_p \cdot E_i$.

Experimental Procedure

The test objects are illuminated as for visual instruments. We generally use the higher brightnesses only. In order to measure t in the case of the perfect objective, it is necessary to use an instrument sufficiently good to be considered practically perfect. We therefore use astronomic lenses with very short focal lengths, between 20 and 50 mm, or apochromatic microscope objectives, set in a small camera with a high-precision movement for focusing.

A very important point is to know what magnification is to be used when observing the photographic images, this magnification depending upon the conditions of use.

When the photographs are to be seen without magnification, the observation is done with the naked eye at a standard distance of 250 mm.

Frequently we want to determine the best resolving power. It is obtained, for all practical purposes, when the area of granularity-diffusion in the emulsion is seen from the eye within an angle of 4 to 10 minutes. This is obtained for a resolution test when the angular resolution for the eye is between 3.5 and 8 minutes.

Example of an Efficiency Determination

Figure 7.10 gives the variation of efficiency with contrast in the center of the field for an excellent photographic lens operating at apertures of f/6.7 and f/12, and for a lens with third-order spherical and chromatic aberrations.

We might notice two interesting results. A photographic lens, considered to be an excellent one, never reaches an efficiency of 1 for

any contrast, even for $C=1$. This has already been noted by many observers.

The loss in the efficiency is largest for a contrast between 0.1 and 0.2, after which the efficiency increases rapidly, until the perceivable image disappears. This is a different result from that obtained in the case of a visual instrument. This occurs because the emulsion does not reproduce small contrasts very well, and the defects of the instrument have less and less effect on the resolving power.

Conclusion

The method given above, used now for 20 years for visual instruments, and nearly 10 years for photographic instruments, has satisfactorily solved, for us, the problem of determining the quality of an instrument in any point in the field. It avoids the discrepancies that occured very often between the results of tests in the laboratory and the practical use of the instrument. However, it has the serious disadvantage of not separating the eye from the optical image. If it were possible to separate them, then by combining the quality of the optical image, without any receiver, and the special properties of any receiver, it would be possible to predict the quality of the instrument for any condition. This is not yet possible because we do not yet know enough either of the variations of the distribution of illumination with various aberrations taken together, or of the properties of receivers. However, the experimental method maintains its interest because it is rapid, and because it takes into account the totality of the effects produced by instrument defects, some of which cannot be forecast by computation.

Discussion

Dr. H. Osterberg, American Optical Co., Stamford, Conn.: A question for Dr. Macdonald. If your answer to this question is "Yes," I would like to ask one more. You mentioned the discrimination of squares and circles.

Dr. D. E. Macdonald, Optical Research Laboratory, Boston, Mass.: Yes.

Dr. Osterberg: In connection with one of your charts?

Dr. Macdonald: I have answered "Yes."

Dr. Osterberg: Have you studied the proposition as to when an individual would judge a square that has rounded corners to be a square as a function of the distance he stands from this figure?

Dr. Macdonald: I could answer the question "Yes" but I can't give you the information as to what the functional relationship is at this time. I would like to point out that I mentioned in qualifying this—.

Dr. Osterberg: May I interrupt? I am not going to ask for such a complicated functional relation as you think. What I am interested in is this. What is the ratio of the radius at the corner of the square to the diagonal when this person reports this square with rounded corners to be a square?

Dr. Macdonald: We don't have the information in that form at all. As I tried to point out at the start, this was preliminary data presentation on the basic discrimination or recognition only of this

one type, the simplest type of discrimination we could think of, that is, between a square and a circle. Under different conditions of resolution—this involves the rounding of the corners, but we did not purposely introduce a form factor of this type—we have presented the same information to the analysts at different observing distances to bring in the effect of variable magnification.

As I say, I am not prepared to give any answers on any aspect except the basic square-circle differentiation at this time.

Dr. Osterberg: Thank you.

Chairman: Are there other questions that anybody would like to ask?

Dr. H. R. J. Grosch, General Electric Co., Lockland, Ohio: I would like to point out that these problems of interpretation of photographs, et cetera, remind me very much of the work done in educational and psychological testing because we have essentially a subjective phenomenon and we have experienced photointerpreters who would be willing to tell us this was or was not a good photograph but who would be unable to describe exactly why they thought so.

On the other hand, we have a series of measurements or tests that can be performed under controlled conditions but that do not measure the parameters of the problem partly because we don't know which are important. For instance, we assign arbitrary names such as "contrast" and "resolution" to certain things, but it is by no means certain that these are pure factors just as the terms "intelligence," "quickness," "retentiveness," don't mean too much.

Now, there is a mathematical technique that ranges over into the field of fantasy at times, known as "factor analysis." There are many proponents of this in Washington, hiding in cubby-holes of the Pentagon and similar sin spots, and I would like to suggest both to Howlett and Macdonald that it might be possible to run a standard factor analysis relating—or perhaps I should say correlating—the subjective judgments of their trained observers with measurements that they have made of the objectives, such as motion and stability of the camera.

Dr. Macdonald: I think that is exactly what we are trying to do with this statistical analysis of photographs. We are trying to bring out these subjective factors that are so enmeshed in the personality structure of the photointerpreter. The only way to drag them out seems to be through some statistical form.

We thought perhaps the incentive in a commercial organization for fuller interpretation might result in a more uniform background of training, one that we could trace back more easily than we could in the case of military photointerpreters. So analysis in this type has so far been restricted to the commercial applications of photography.

In terms of the basic units—and this ties in with the definition that we have to make if we are going to define the information content of a picture—we need, perhaps, a basic point where there is a binary decision that the photointerpreter makes. There are several ways of deciding what type of decision he makes. For example, if he is scanning along a line, his decision would be either "Yes, there is an edge," or "No, there is not an edge," as he goes from blur point to blur point.

I am talking about a discreet one-dimensional case and I have mentioned previously that the problem is one of moving to a two-

dimensional source and a continuous case. If the existence of an edge is a logical unit of information, then its evaluation depends on (a) the absolute magnitude of the change, and (b) the gradient.

CHAIRMAN: I believe we had one other question.

MR. A. H. KATZ, Photo Reconnaissance Laboratory, Wright Air Development Center, Dayton, Ohio: This is just an additional comment on the remarks made by Dr. Grosch, to the effect that it is only lately, after 10 years of talking to photointerpreters, that they have even been willing to realize that there is something in the picture other than the scale. This is a waste of good standard deviations.

DR. R. A. WOODSON, Armour Research Foundation of Illinois, Institute of Technology, Chicago, Ill.: I would like to direct a question to two gentlemen, Dr. Howlett and Prof. Arnulf. They both deal with the same question. Should we test a lens in the way it is intended to be used, or should we magnify the image for its inspection? Now, let's consider this first on the basis of the decision of whether a basic lens design is satisfactory, and, second, on whether an instrument made to such a design satisfies the performance requirements.

If you consider that in production manufacture of an optical instrument the operators will have variable visual accuity, you will have disagreements as to the rejection and acceptance of an instrument, whereas if the image is viewed under magnification you eliminate that human factor. I would like comments, please.

CHAIRMAN: Mr. Arnulf, do you wish to comment? Is Dr. Howlett here?

DR. L. E. HOWLETT, National Research Council, Ottawa, Canada: Mr. Chairman, I want to say that it would be acceptable under the circumstances outlined to use whatever degree of magnification was required. Whatever visual criterion is used in the test could be one well correlated with the circumstances of use. Such a procedure would make the test easier, too.

DR. W. WALLIN, U. S. Naval Ordnance Test Station, Inyokern, China Lake, Calif.: Dr. Howlett, it strikes me as somewhat anomalous and bothers me that the philosophy you express seems quite adequate as a functional test, suitable as inspection technique; but let us suppose that we have established to our satisfaction that such a test is inadequate. Then we want to arrive at something more analytic, divide and conquer. There are factors present in the lens system, itself, factors present in the photographic emulsion, let us say, and factors imposed by external circumstances.

If we are to understand the problem we have to analyze them and separate them, and yet once we do that we depart from your basic philosophy, don't we? I would like your comments; or haven't I made my question clear?

DR. HOWLETT: Well, perhaps you have. I will try and answer it. I think there are two distinct things involved there. Perhaps I did not make them sufficiently clear but I did refer to them. Certainly, any philosophy of evaluating quality of an image from the point of view of the user must, in the last analysis be a routine test, but that still has no bearing on any other test. You can do all the analyses that you like but keep it separate until it is in a form where it can be incorporated as an explanation or where it will lead to improvements in the evaluation. You can still evaluate the quality of an image on the photographic emulsion even if you don't know all about the for-

mation of the image, the developers, the various lens elements, et cetera, all of which are extremely important.

Having evaluated in an orderly manner the quality of the image, you can still go on.

Dr. Wallin: You are not saying then that the information we get is not going to help us?

Dr. Howlett: I did not understand that.

Dr. Wallin: The question was, should we really rely on visual analysis in evaluating a lens that is to be used photographically?

Dr. Howlett: I think it is quite reasonable to do so if you can show that by so doing your results correlate precisely with the photographic procedures that are going to be used. We have found that there are particular visual points that are always within a reasonable tolerance of where the photographic plate should be. I think that is perfectly acceptable and saves an enormous amount of time.

8. Image Quality as Used by the Government Inspector of Visual Telescopic Instruments

By H. S. Coleman [1]

Introduction

There are a considerable number of points of view from which image-quality evaluations can be considered. It is the purpose of this paper to discuss image-quality evaluation from the point of view of a government inspector whose responsibility is to determine whether or not a given optical system submitted to the Government conforms with its specifications. Because of the widely different procedures that might be followed in inspecting optical systems, it is the purpose of this paper to describe what are perhaps the most popular three inspection processes used in inspection and to show the correlation among the image-quality evaluations that are obtained among these three methods.

For purposes of presenting a description of the above mentioned test equipments, this paper is divided into three parts, corresponding to the three test procedures under consideration. The first of these test procedures involves the use of a device referred to as "the Kinetic Definition Chart Apparatus" [2] (because of its motorized parts), the second involves the use of a "Twyman-type Interferometer," [3] and the third involves the use of a device referred to as a "Dioptometer." [4]

The Use of the Kinetic Definition Chart Apparatus in the Inspection of Visual Telescopic Systems

The following section of this paper describes the Kinetic Definition Chart Apparatus (hereafter abbreviated to K. D. C. apparatus) and presents examples of the type of optical data that can be obtained with this equipment. The basic principles upon which this device depends were first described by Fabry [5] in 1935.

The K. D. C. apparatus can be regarded as an apparatus for measuring the "resolving power" of optical systems in which the spacing between the elements of the test object used and the contrast of the elements of the test object used can be varied continuously.

The Apparatus

The K. D. C. apparatus is essentially a continuously variable resolving-power apparatus in which the target is made to appear (by optical collimation) at any desired range and in which the contrast of

[1] Director of the Scientific Bureau, Bausch & Lomb Optical Co., Rochester, N. Y.
[2] H. S. Coleman and S. W. Harding, J. Opt. Soc. Am. **37**, 263 (1947).
[3] H. S. Coleman, D. G. Clark, and M. F. Coleman, J. Opt. Soc. Am. **37**, 671 (1947).
[4] H. S. Coleman, M F. Coleman, and D. L. Fridge, J. Opt. Soc. Am. **41**, 94 (1951).
[5] C. Fabry, Proc. Phys. Soc. London **48**, 747 (1935).

the target may be varied thus simulating conditions existing in the field. In addition, the test object may be surrounded with an illuminated region designed to simulate natural backgrounds against which targets may be viewed. The apparatus consists of a test object, an optical-reduction unit, a collimating unit, a standard telescope, an auxiliary telescope, an artificial sky, an off-axis fixture, and a mechanism whereby the angular subtense of the elements of the test object may be varied continuously. Each of these parts is discussed separately.

Test object. The basic type of test object used was first introduced by Foucalt [6] in 1858. This object consists of alternate white and gray bands of equal width of the form shown in figure 8.1 In general, it has been found desirable to use test objects in which the markings are oriented in four directions. The relatively coarse numbers shown in figure 8.1 are for purposes of identifying the test object and for focusing. The K. D. C. test objects can be made having any desired spacing and contrast between the white and gray bands. These are calibrated photoelectrically for the number of bands per unit length and for contrast. The test object is illuminated by means of a sphere as shown in figure 8.2.

Optical reduction unit. The apparent size of the test object is reduced optically by means of any one of a set of lenses mounted in a turret shown in figures 8.2 and 8.3. These lenses are referred to as the optical-reduction unit. The apparatus in its present form utilizes microscope objectives having focal lengths of 4, 8, 16, and 32 mm., and Plössel-type eyepieces having focal lengths of 20 and 50 mm., respectively. The optical-reduction unit is required in order to obtain practical limiting distances of resolution. As a result of this feature, the entire length of the K. D. C. apparatus is approximately 6 feet. Early models were as long as 120 feet.

Collimating unit. The collimating unit is used to make the test object appear (optically) to be at any desired range. It consists of a high-grade telescopic objective located such that its focal plane may be placed at any desired distance from the image of the test object formed by the optical-reduction unit.

Standard telescope. The standard telescope consists of highly corrected optical parts and is provided with efficient stray light stops. The objective has a clear aperture of 2.5 inches and an effective focal length of 15.75 inches. Provision is made for varying the aperture of the standard telescope by means of a set of stops so that its entrance pupil can be made equal to that of the optical device under test.

The standard telescope is used to establish the minimum angle of resolution for a "perfect" optical system of a given aperture, to acquaint the observer with the appearance of an image formed by a high-grade optical system, and to adjust the collimating unit for the desired target distance. This adjustment is made by focusing the standard telescope on an outdoor target placed at the desired range and then by adjusting the distance between the collimating lens and the image of the test object formed by the optical-reducing unit so that the test object is at the same focal setting as the outdoor target.

Auxiliary telescope. The auxiliary telescope is an astronomical-type telescope having an entrance pupil of 1.250 inches, which is provided with a set of eyepieces making it possible to vary its magnification from 1 to 30x. The auxiliary telescope is used in series with the instrument under test as shown in figure 8.2.

[6] L. Foucault, Ann. Observ. Paris, 5, 197 (1859).

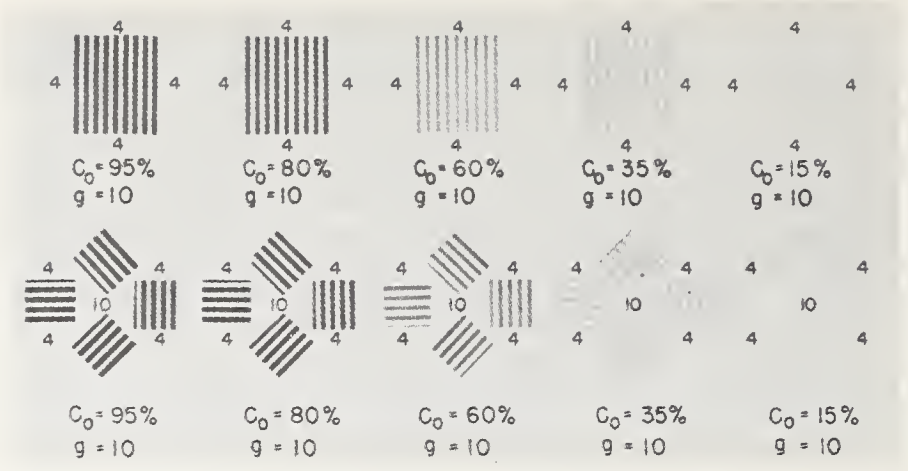


FIGURE 8.1. *Modified Foucault resolution targets.*

$$C_0=\frac{B-b}{B}=\frac{\text{Reflectivity of white band}-\text{Reflectivity of dark band}}{\text{Reflectivity of white band}};$$

C_0=inherent target contrast; g=number of white bands per inch

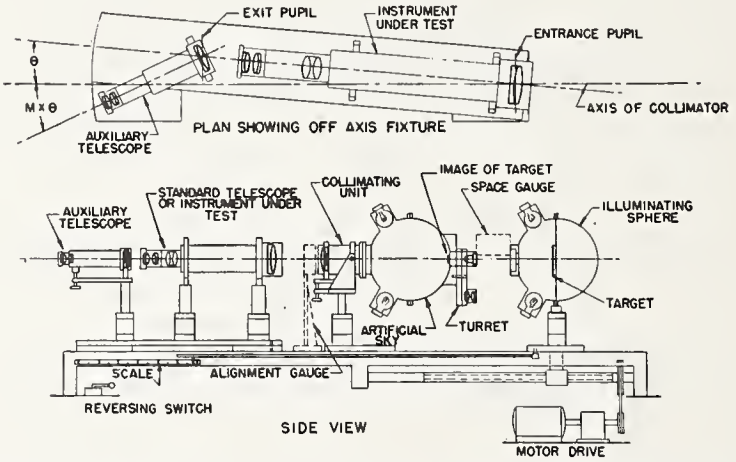


FIGURE 8.2. *Schematic diagram of Kinetic Definition Chart apparatus.*

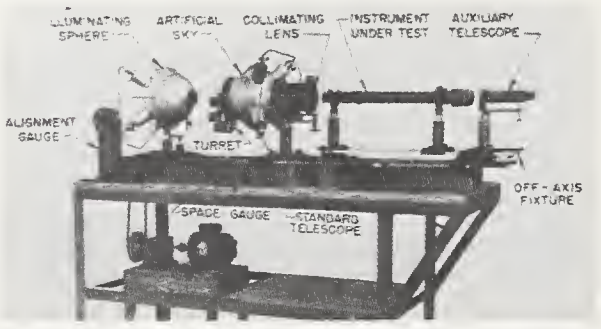


FIGURE 8.3. *The Kinetic Definition Chart apparatus.*

In the quality-control processes, the auxiliary telescope has three functions. The first of these is to make it possible to test the system under consideration at full aperture. This is done by reducing the exit pupil of the combination of the system under test and the auxiliary telescope to less than that of the natural pupil of the eye. Frequently the aberrations in optical systems are greatest at the margin of the aperture and may not be noticeable if the pupil of the eye becomes the effective stop of the system. In addition the pupil size varies from inspector to inspector. The second function is to reduce the portion of the pupil of the eye being used in the measurement to such an extent (1 mm or less) that the eye performs as a "perfect" optical device. This practically eliminates differences in measurements made by various observers. The third function of the auxiliary telescope is to produce a sufficiently large image on the retina of the observer's eye to make an observation "comfortably." In this connection, it should be noted that the auxiliary magnification does not provide a means of exceeding the limit of resolution of the eye because of the inverse relation between the size of the exit pupil at the eye and magnification, and as a result of the limitations imposed by diffraction theory.

Artificial sky. In order to simulate outdoor conditions, the image of the test object formed by the optical-reduction unit is surrounded by an illuminated region. This region is referred to as "the Artificial Sky" and subtends an angle of approximately 30° measured from the center of the collimating objective. The artificial sky is constructed so that it forms an adjustable field stop for the optical-reduction unit. This adjustment is accomplished by means of a series of spherical segments having various apertures and mounted in a turret as shown in figure 8.2.

The artificial sky has two functions. The first is to produce a known brightness level to which the observer's eye is to be adapted. The second function provides a means of taking into account the stray light normally introduced into an optical system by natural backgrounds when the optical instrument under consideration is performing its intended task. Stray light in optical devices reduces the contrast of the image on the retina and, consequently, the range at which objects are visible.

Variable spacing unit. The mechanism for controlling the angular subtense of the target markings consists of a motorized unit by means of which it is possible for the observer to vary continuously the distance between the test object and the optical-reduction unit. This variation changes the size of the image formed by the optical-reduction unit and, consequently, the angular subtense of the target markings. The distance between the optical-reduction unit and the test object is indicated on a scale graduated in tenths of an inch. The optical parts are arranged so that the scale reading is directly proportional to the minification of the image of the test object formed by the optical-reduction unit. In this connection it might be noted that for the range over which the target is moved, the longitudinal displacement of its image formed by the optical-reduction unit is much less than the focal range (depth of focus) of the collimating objective and hence the image viewed is always in focus. This means that neither the instrument under test nor the auxiliary telescope have to be changed in focus when making a K. D. C. measurement.

By virtue of the continuous feature of varying the distance between the test object and the optical-reduction unit, precise resolution measurements can be made since it eliminates the use of the discreet angular intervals of conventional resolution objects.

Gages. In order to facilitate the adjustment of the K. D. C. apparatus two special gages have been devised. These are shown in figure 8.2 and 8.3.

The first of these gages, referred to as the space gage, is used to set the distance between the target and the optical-reduction unit so that the scale reads linearly with minification. This gage is simply a metallic block 5 inches long. Appropriate scale readings have been determined for each of the optical-reduction units when the gage is placed between the face plate of the test-object illuminator and the cell of the optical-reduction unit.

The second gage, referred to as the alinement gage, provides a means of alining the optical axes of the test object, the optical-reduction unit, and the collimating unit. This gage is essentially a combination two-dimensional square and height gage and is used as indicated schematically in figure 8.2.

Off-axis fixture. The K. D. C. apparatus is provided with a mechanical assembly for mounting the optical system under test for making resolution measurements at various field angles. This fixture, referred to as "the off-axis fixture," is designed so that the optical device under test can be alined with the optical axis of the collimating unit and can be rotated about its own entrance pupil. The off-axis fixture also provides a means of rotating the auxiliary telescope about the exit pupil of an optical device under test.

The Procedure

The end product of the K. D. C. test is a quantity referred to as the K. D. C. efficiency. This efficiency is a numerical measure of the quality of an optical system. The axial K. D. C. efficiencies of the better present-day optical systems approximate 100 percent. The K.D.C. efficiency may be used in either of two different forms depending upon whether the measurements are to be used for predicting the performance or for determining the quality of an optical device under consideration.

In predicting the performance of telescopic systems, it is desirable to know how much farther a given object can be seen with the instrument under consideration than with the unaided eye. This point of view gave rise to the definition of the K. D. C. efficiency given by eq 1.

$$\text{K. D. C. efficiency} = \frac{X_i'}{M_i X_e} \times 100, \tag{1}$$

where

X_i' = K. D. C. scale reading at limit of resolution for the eye plus instrument under test,
X_e = scale reading at limit of resolution for the unaided eye,
M_i = the magnification of the instrument under test.

In determining maximum resolution of an optical system, auxiliary magnification is generally used for the reasons indicated above. In

this case, K. D. C. efficiency is defined by eq 2, which is equivalent to the definition presented in eq 1, when less than 1-mm diameter of the eye is used.

$$\text{K. D. C. efficiency} = \frac{X_i'}{S} \times 100, \tag{2}$$

where

X_i' = K. D. C. scale reading at the limit of resolution for the instrument under test using the auxiliary telescope,
S = K. D. C. scale reading at the limit of resolution for the standard telescope having the same entrance as the instrument under test and using sufficient auxiliary magnification to eliminate the imperfections that may be introduced by the observer's eye.

It will be noted that when one uses eq 2, he is simply comparing an instrument under test with one known to be of high optical quality and to be practically free from stray light. It should be noted, however, that since the standard telescope has a constant specific coefficient of resolution (defined to be the product of the aperture in inches and the minimum resolvable angle in seconds) of 4.51 inch seconds, the K. D. C. efficiency may be converted into angular units. In making K. D. C. measurements, the limit of resolution is taken to be the scale reading at which the direction of the markings at any one of the four orientations cannot be specified.

The procedure for making K. D. C. measurements for telescopic systems may be divided into two parts. The first of these consists in axial measurements using test objects of various contrasts. The second is the measurement of resolution at various field angles with a test object of fixed contrast. The procedure consists simply in noting the K. D. C. scale readings at the limit of resolution for each target contrast or field angle. The results of such measurements are plotted in the form shown in figures 8.4 and 8.6. As indicated, measurements may be made without the Artificial Sky in operation and at different brightness levels. In order to illustrate the importance of surface defects and cleanliness a thumbprint was placed on the center of the objective of the standard telescope to provide a source of stray light as this instrument is practically free of such defects. In the measurements presented, a uniform surround was used and the target was at the same apparent brightness as the surround and an auxiliary telescope was used. Figure 8.5 is presented to illustrate the relation of K. D. C. efficiency to the scale readings presented in figure 8.4. It is to be noted that the K. D. C. efficiency for any field angle is 100 percent by definition for the standard telescope with clean optics. Figure 8.7 shows an example of the K. D. C. efficiency obtained for a 5x telescope at various field angles without the use of auxiliary magnification.

Possible errors. There are two principal sources of error in making measurements using the K. D. C. apparatus. The first of these is in focusing on the image and the second is the failure to expose the image of the object to the eye long enough. The error in focusing has been reduced by focusing on numerical figures rather than on the resolution markings of the test object. The exposure time required may be as

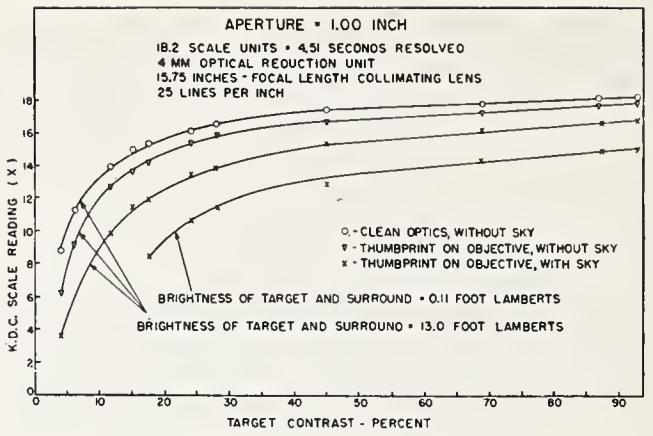


FIGURE 8.4. *Kinetic Definition Chart scale readings (x) versus target contrast for a standard telescope with and without artificial sky.*

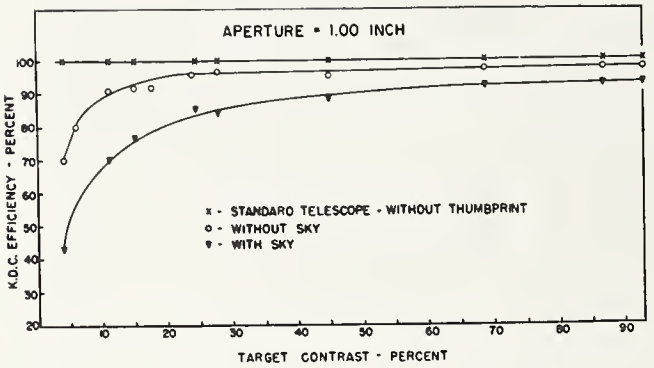


FIGURE 8.5. *Kinetic Definition Chart efficiency versus target contrast for standard telescope with thumbprint on the front of the objective.*

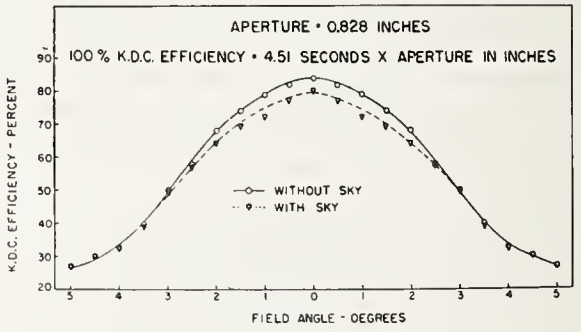


FIGURE 8.6. *Kinetic Definition Chart efficiency versus field angle for 3x telescope with and without artificial sky.*

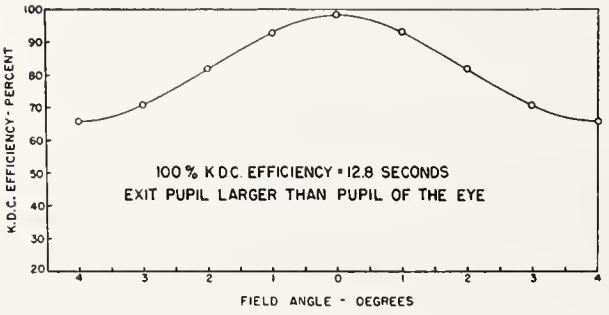


FIGURE 8.7. *Kinetic Definition Chart efficiency versus field angle for 5x telescope without auxiliary magnification.*

much as 10 seconds for low-contrast objects or for poor-quality images. This error appears to be related to unsteadiness of the observer's head as the required exposure time can be reduced by using a chin rest. If the above errors are minimized, the probable error in the resolution measurements made using the K. D. C. apparatus does not exceed 2 percent in the case of optical devices of high quality.

The Dioptometer

The dioptometer used to test the instruments mentioned in this paper is of a fairly classical design and therefore is described briefly. It is shown schematically and photographically together with its accessories in figures 8.8 and 8.9. The dioptometer is a simple telescopic system consisting of an objective, an eyepiece, and a reticle, and is calibrated to measure the distance an object appears to be from its objective in units referred to as "the diopter". The diopter is the reciprocal of the apparent target distance in meters. This device is used to focus on the image of a test object, in the shape of a cross, formed by an optical instrument under test. If both the vertical and horizontal parts of the image of the cross can be brought into sharp

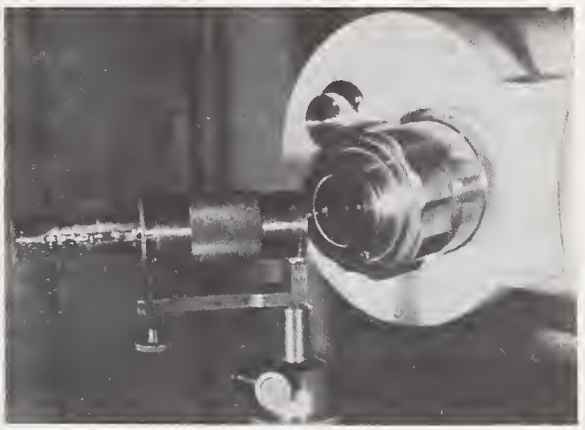


FIGURE 8.8. *D–2 dioptometer used to measure the astigmatism of telescopic systems.*

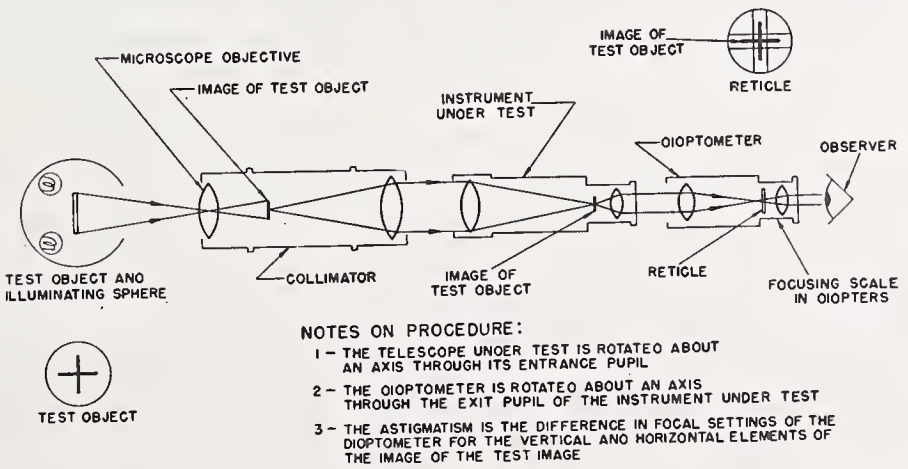


FIGURE 8.9. *Schematic diagram of the apparatus used to measure the astigmatism of telescopic systems by means of the dioptometer.*

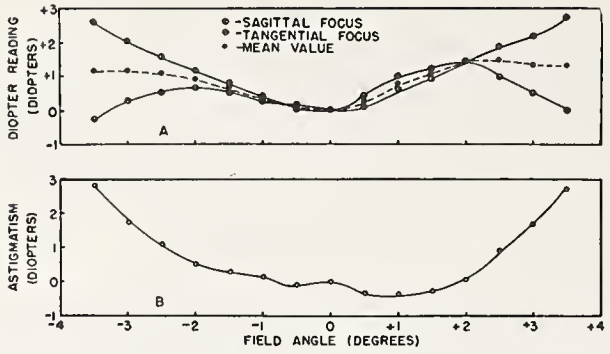


FIGURE 8.10. *Dioptometer readings versus angle of view.*

A. Analysis of a 7- by 50- by 7-degree prism-erecting telescopic system; B. Difference in dioptometer reading between the two foci.

focus when viewed with the dioptometer, the instrument is said to be free from astigmatism. If the two parts of the image of the cross cannot be brought into sharp focus for the same scale reading of the dioptometer, the instrument under consideration is said to have astigmatism. The magnitude of astigmatism is determined by focusing the dioptometer first on the vertical and then on the horizontal parts of the image of the cross and noting the dioptometer scale readings thus obtained. The difference in focus between the vertical and horizontal lines thus obtained is referred to here as astigmatism. Measurements of astigmatism are usually made for various field angles.

Typical dioptometer data are shown in figure 8.10.

The Interferometer

The interferometer used to test the instruments mentioned in this report is shown schematically and photographically in figures 8.11 and 8.12. It resembles the well-known Twyman instrument except in certain mechanical details. The procedure used to evaluate the quality of an optical system, however, is new and is therefore described. Most attempts to use the interferometer for purposes of evaluating the quality of optical systems have been confined to lenses and prisms and depended upon the counting of the number of interference fringes obtained in some specified manner. Such task is almost hopeless in the case of the great variety of complex interference patterns obtained in testing telescopic systems. Because of the difficulty in analyzing the interferometer patterns produced by complex mixture of aberrations, consideration was given to the possibility of evaluating the quality of an optical system on a basis that did not require the specification of the magnitudes of the specific aberrations. Since the main use of the interferometer, from the point of view of this paper, was to evaluate the optical quality of visual telescopic systems, an investigation was made of the type of interferometer patterns that were produced by telescopic systems focused visually. In this investigation a number of instruments were focused on numerals subtending angles well above the limit of resolution and placed optically at infinity. The investigation indicated that the best visual focus was found to give rise to interferometer patterns having one part covered by a single interference area. This region is the area over which the optical-path-difference does not exceed one-quarter of a wavelength (taking

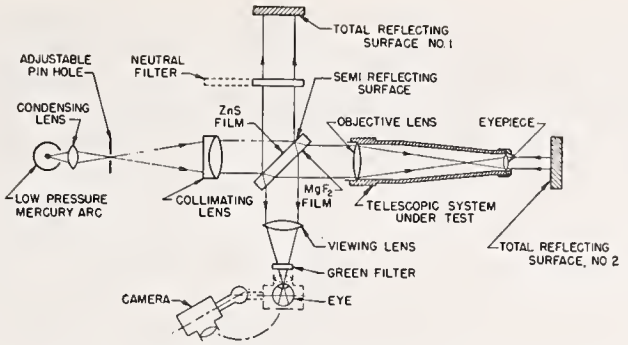


FIGURE 8.11. *Schematic diagram of interferometer used to test telescopic systems.*

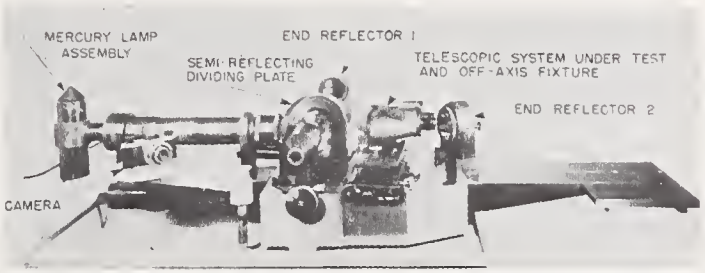


FIGURE 8.12. *Interferometer used to test telescopic systems.*

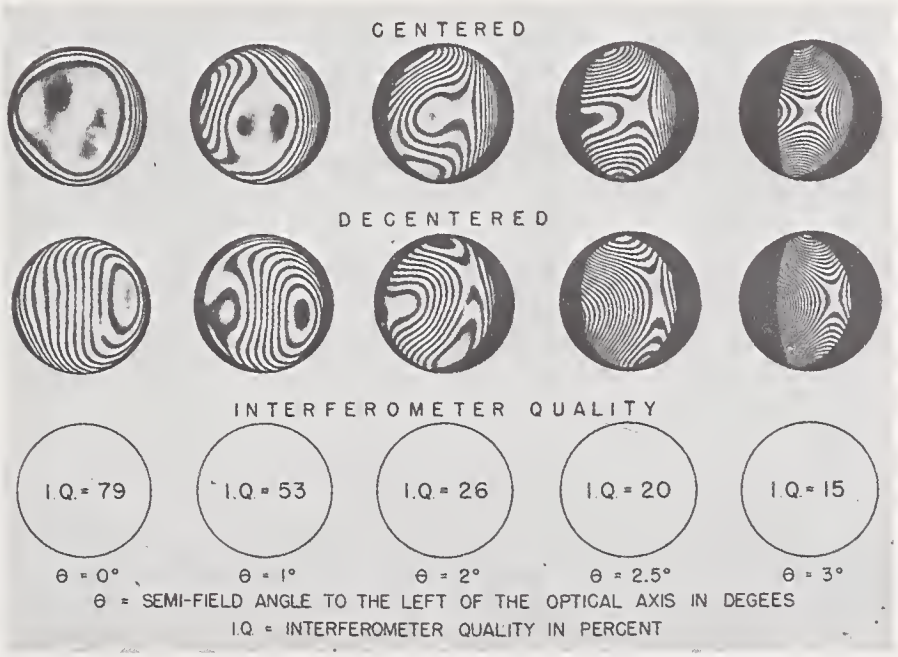


FIGURE 8.13. *The interferometer quality at various field angles of the right barrel of the 7- by 50- by 7.1-degree binocular, Mark 28 Model O serial no.* 197549.

Adjusted for best focus.

into account the double passage of the light through the telescopic system under test). The diameter of this area was found to decrease by any slight change from best visual focus. The diameter of the largest circle that can be inscribed in this area is taken to be a measure of the optical quality. When this diameter is expressed as the percent of the axial entrance pupil of the system under test, the percentage is referred to as the interferometer quality (hereafter interferometer quality may be abbreviated to I. Q.). The diameter of the inscribed circle is referred to as the I. Q. circle, and is defined in terms of the area over which the optical-path-difference, introduced by a system under test, does not exceed one-quarter of a wavelength. For practical purposes, the interferometer quality might be thought of as the percentage of the entrance pupil of the system under test that is free from aberrations.

A sample of the data obtained showing the interferometer quality of an instrument is presented in figure 8.13. Because of the symmetry found for the aberrations across the field of the optical systems studied, data are presented only for the semifield angles of 0, 1, 2, 2.5, and 3 degrees to the left of the optical axis.

Conclusions

Certain general conclusions have been drawn concerning the various experimental procedures mentioned in this paper. These conclusions are stated briefly for each type of apparatus.

The K. D. C. Apparatus. The conclusions concerning the use of the K. D. C. apparatus are summarized as follows: (1) The K. D. C. apparatus is a versatile device that may be used to measure resolving power of a variety of optical devices including telescopes, photographic objectives, microscope objectives, eyepieces, and the human eye; (2) the K. D. C. apparatus provides a means of taking into account the brightness level to which the human eye is adapted, the ill effects introduced by stray light, and the target contrast; (3) the K. D. C. apparatus is sufficiently rapid and impersonal to be used for quality-control purposes for visual telescopic systems being mass produced; (4) the K. D. C. apparatus provides a means of objectively evaluating the design of optical systems; (5) the K. D. C. apparatus may be used to obtain data required for predicting the performance of optical systems.

The Dioptometer. The dioptometer described in this report provides a simple means of evaluating the quality of optical systems in which the predominant aberration is astigmatism.

The Interferometer. The conclusions reached concerning the use of the interferometer mentioned in this report are as follows: (1) The interferometer may be used to evaluate the quality of optical systems of moderate dimensions being mass produced; (2) the procedure described for using the interferometer in this paper is simple and rapid and does not require a technically trained operator; (3) the interferometer is the most precise test fixture for evaluating the quality of optical systems available at the present time; (4) It is evident from figures 8.14, 8.15, 8.16, and 8.17 that a high degree of correlation exists among the data obtained using the K. D. C. apparatus, the interferometer, and the dipotometer for the instruments mentioned in this paper; (5) The high degree of correlation obtained

between the Dioptometer values of astigmatism values and the K. D. C. efficiency, indicates that astigmatism is the principal aberration responsible for the rapid decrease in optical quality across the field for the present design. This may be seen by comparing figure 8.14 with figure 8.17.

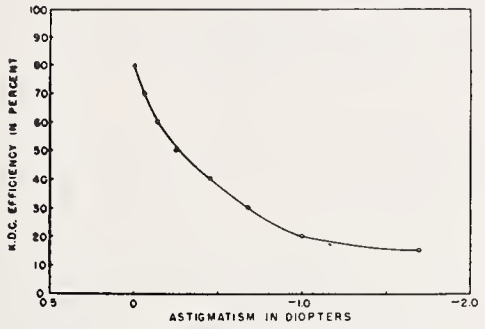


FIGURE 8.14. *Average values of the Kinetic Definition Chart efficiency and the astigmatism in diopters for all the instruments tested.*

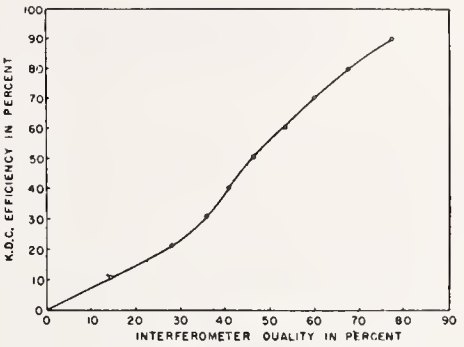


FIGURE 8.15. *Average values of the Kinetic Definition Chart efficiency and the interferometer quality for all the instruments tested.*

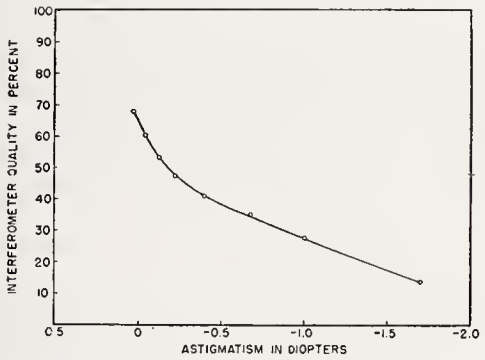


FIGURE 8.16. *Average values of the astigmatism and the interferometer quality for all the instruments tested.*

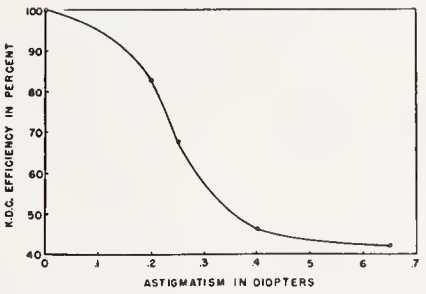


FIGURE 8.17. *Relation between Kinetic Definition Chart efficiency and astigmatism for pure primary astigmatism.*

9. Application of Fresnel Diffraction to Measurements of High Precision

By A. C. S. van Heel [1]

Introduction

On a previous occasion it has been emphasized that a direction can be determined with surprising precision by means of a diffraction pattern of the Fresnel class and especially when settings are made on the color transitions, occurring in the usually very diversely colored patterns obtained when a white light source is used.[2,3] As one such application, a spherometer for large radii and a flatness tester were described. As this work has been continued and the results have been found to be of desirable accuracy, although the simple method has not been modified essentially, a description of these developments, together with some comments and corrections of the previous publication, appears justified. It will be shown that the scope of the method is exceedingly general and its application to the measurement of even very small aberrations is quite possible. The method is essentially a means of studying the structure of wave fronts, or, when the conception of light rays as normals to these is preferred, the structure of light pencils.

Study of a Flat Surface

The determination of the form of a nearly flat surface may serve as a relatively simple example and will be described in some detail.

In figure 9.1, S is a horizontal slit, some tenths of a millimeter wide, receiving light from a glow lamp or, better, from a carbon arc. At a distance a of about 2 m, a metallic grating G is placed, at TT′ the surface to be tested is placed. The light reflected from this surface is observed at MM′ by means of a traveling microscope of low magnification (10), reading to approximately 1 micron. The distances b and c are approximately 1 m.

In order to remove the observing microscope from the path of the incident light the surface is tilted about the line TT′ by an angle α, so that MM′ is not in the plane of the figure but behind it. When calculating the results this tilt has to be taken into account, a simple procedure for flat surfaces, but requiring a rather complicated correction formula when the surface is convex or concave. Fortunately, the formula can be simplified to

$$\delta = g\bar{y} + hR\bar{y} + \ldots, \tag{1}$$

[1] Laboratory of Technical Physics, Technical University, Delft, Holland.
[2] A. C. S. van Heel, J. Opt. Soc. Am. **40**, 809 (1950).
[3] A. C. S. van Heel, J. Opt. Soc. Am. **41**, 277 (1951) correction.

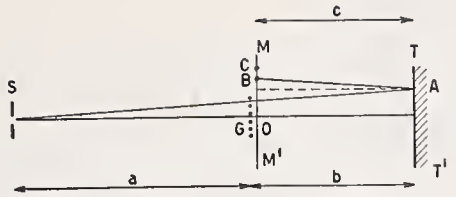


FIGURE 9.1. *Surface tester.*

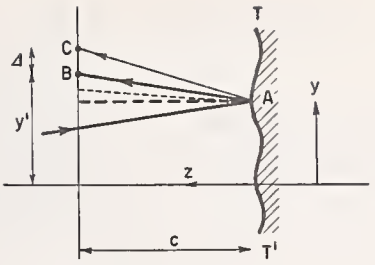


FIGURE 9.2. *Slope derived from deviation.*

where δ is the correction to the observed distance OB, $\bar{y}$ the distance from SO at which the incident ray intersects the plane of the grating, and R the curvature of the surface; g and h are constants, depending on α, a, b, and c. The value of $\alpha=2.09°\pm0.05°$ allows the following terms of (1) to be neglected, as they amount to much less than the tolerances given by the precision of pointing, as long as the radius of the surface is greater than about 1 m.

In the following this correction is always assumed to have been applied, in order to simplify the discussion.

Derivation of the Surface Profile

The diffraction pattern displays its multiple and significant colors MM′ exactly as if this plane were at a distance $b+c$ beyond G and the reflecting surface were not present, provided that it is perfectly plane. Deviations from planeness give rise to changes in direction of the normal. Thus in the case of figure 9.1, if TT′ is not plane the normal at A might have a different direction, and the "correct" position of the intersection with MM′ of the normal to the reflected wave front or reflected "light ray" would be displaced from B to C.

When the deviations from planeness are small the variations of the slope are given with ample accuracy by 1/2 BC/c. Referring to figure 9.2 we can write

$$dz/dy=-\tfrac{1}{2}\Delta/c, \tag{2}$$

where $\Delta=BC$ is the observed deviation from the theoretical point of intersection. The last is known from

$$y_0'=\bar{y}(a+b+c)/a, \tag{3}$$

where $\bar{y}$ has the meaning given above.

Integrating the curve for dz/dy with respect to y we find the form of the surface z (y).

Once the shape of the reflecting surface along a line lying in the plane of figure 9.1 has been established, the coordinates may be transformed (translation and rotation) when this is necessary to compare different series of measurements on the same surface. Again, when the surface has been tilted about an axis normal to the plane of figure 9.1, as may happen when the object has been removed, cleaned, and replaced, a rotation usually must be taken into account together with an accompanying division of c by the cosine of the tilting angle.

Straightness of the "Light Rays" and Basis of Measurement

Before proceeding to give detailed results, we must say some words on the "light rays" and the way measurements are performed. As has been stated in a previous publication (see footnotes 2 and 3), the maxima of light in the diffraction pattern correspond more or less to the slits in the grating and the minima to the wires, provided the observing distance $(b+c)$ is small enough to obviate any "forkings" as indicated in figure 4 of the cited article. Still the assumption that the loci of the colored maxima (and of the otherwise colored minima) are straight is even then only approximately true and needs testing. This can be and is done by removing the surface to be tested and measuring the pattern at different distances from G.

Since 1949 the traveling microscope has been improved; the consistency and accuracy of the readings have been considerably bettered, while magnification has been increased from 5 to 10. The small autocar lamp has been replaced by an arc lamp, which permits the color transitions in the pattern to be more clearly and precisely determined. Thus the measurements are more trustworthy.

It appears from careful observations that the observed curvature of these loci [4] is considerably smaller than was then thought and, for a large part, is to be ascribed to observation errors. Instead of the not very consistent curvatures of these loci, we can now state, with more confidence, that the loci are practically straight (with a grating period of about 0.6 mm and at distances from G up to 2 m). Deviations from straightness are never more than 3 microns, even up to the eleventh grating slit above or below the middle slit O.

This fact has emboldened us to speak of "light rays" as a short expression for these loci and facilitates in no mean measure their use. The way in which the curvature begins to play a part when the dimensions are altered and especially how the loci behave when the light is reflected has to be studied separately. For our present purpose we will let this point rest and make use of the fact that the loci are straight within less than $3 \cdot 10^{-3}/2 \cdot 10^{3}$, or 0.3 second of arc.

Not only the maxima and minima are colored, they are also accompanied by fringes of divers hues and intensities. Each period repeats itself with the same sequence of colors, symmetrical about the maximum as well as about the minimum. Thus there is more than one characteristic to determine their position.

Even with one characteristic for each period we found a mean error of 8.2 microns in five series of determinations on 19 maxima with four pointings on each maximum; the five series gave respectively 6.4, 8.3, 8.1, 9.6, and 8.3 microns.

The distances a, b and c were measured several times in a period of 6 months and could be trusted to about 0.2 mm.

At 19°.0 C their values were: $a=2342.1$ mm, $b=1204.2$ mm, $c=1179.4$ mm. A discussion of errors shows that the uncertainty of these distances has no appreciable influence on the final results. The supports being beams of steel, the values can be corrected for temperature with a dilatation coëfficient of 12.10^{-6}. The object whose surface was to be tested was applied by a double spring to a contact ring. The distances b and c were measured to and from this plane. When

[4] A. C. S. van Heel, J. Opt. Soc. Am. **40**, 812 (1950).

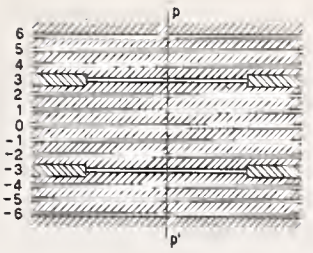


FIGURE 9.3. *Grating with partly blocked slits.*

FIGURE 9.4. *General view.*

convex or concave surfaces are studied a correction for the sagitta has to be applied.

During the measurements it is desirable to know with which slit of the grating one is concerned. This was effected by blocking parts of the slits as indicated in figure 9.3. The slit 0 is at O in figure 9.1, the slits 3 and −3 have been blocked except in the central parts. The coloration of the diffraction fringes corresponding to these is distinctly affected, thus they can be identified, but only at their marginal parts. Although the influence of the blocking extends farther towards the center than corresponds to the geometrical shadow of the stops, the neighborhood of PP′, where the measurements are made is unaffected. This was ascertained by a separate series of measurements.

Practical Realization

Figure 9.4 gives a general view of the instrument. It has been set up in a basement, whose temperature is constant for long periods of time within 1° C. Brackets in the wall support two strong steel pipes of 4-m length and 3.2-cm outside diameter. Slit S, grating G, microscope M, and object A are mounted on robust carriages clamped on the pipes. The light from an arc lamp (not shown in the figure) is thrown on to the slit by means of a mirror, so that the arc lamp is at a sufficiently large distance from the apparatus to avoid warming it by radiation or convection.

The microscope has a doublet object glass. The light pencils reflected by the tested surface, which are approximately horizontal, are received in the plane of a cross wire on which the microscope is focused. By means of a plane mirror the light is turned into the vertical direction to facilitate observation.

It may be noted in passing that optical parts are not inserted in the light path between the source and the cross wires (with exception, of course, of the test surface itself), but that beyond the cross hairs no error is introduced by any reasonably good optics. The warning may

be repeated here not to make use of a grating on glass, as every slight inhomogeneity of the glass and every deviation from flatness of the surfaces of it are deleterious to the perfect formation of the diffraction pattern. Thus only metal gratings with air gaps can be used.

Probably the gratings are best made by stretching wire of constant thickness over the grooves of two parallel precision screws. We have applied ourselves to the construction of such a harp, plugging the ends of each wire into holes in a robust frame, on which also the screws were mounted, but as yet either the wires were not sufficiently straight or their thickness, about 0.3 mm, changed by the too strong stress. We also tried metal sieves, whose holes act as slits, the light source being a slit. This entailed the necessity of calibrating each row of holes, this calibration being by means of a precision traveling microscope, provided with a supplementary cylindrical lens. Thus the position of each "slit" was known to within a few microns and had to be taken into account in the calculations. Although precision is not impaired seriously in this way the evaluations are somewhat tedious, and we are still trying to produce a grating with sufficient precision to avoid correction terms.

Flat Surface

A standard flat of natural quartz, cut perpendicular to the axis, diameter 65 mm, thickness at the center 20 mm, was one of the first objects to be studied. The cylindrical slab is slightly prismatic, having a thickness of 18 mm at one side and of 22 mm at the other side. The form of the standard surface was investigated in two perpendicular directions. In position I the virtual line of intersection of front and back surface is vertical; this gives the profile along the diameter a–a in figure 9.5. In position II the intersection line is horizontal, the thickest part of the slab being at the bottom; this gives the profile along the diameter b–b in figure 9.5. The back surface had been dulled with paraffin wax. The front surface has been studied as it came from the polisher.

The diameter of the contact ring, on which the slab was pressed with a force of the order of 1 to 2 kg, was 52 mm.

In figure 9.6 we give the results of the measurements in position II, as these were repeated under different conditions. Position I gives comparable results.

The first series gave the results indicated by open circles. The slab was then removed, cleaned again as it had been before the first series with a piece of cotton wool and acetone, replaced and the second

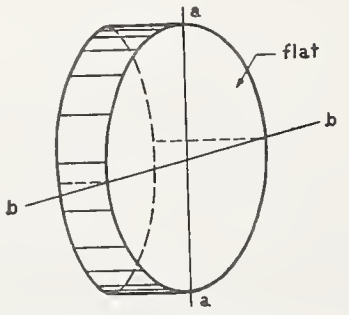


FIGURE 9.5. *Tested slab.*

FIGURE 9.6. *Profile of flat; position II.*

series was made, giving the dots in figure 9.6. Although the correspondence was very good, we were not satisfied with the fidelity to the true form of the profile unless it was tested again with a transverse translation of the surface. Only thus could we make certain that the ups and downs of the curve are to be ascribed to the surface and not to some faulty calibration of the grating or other error. The slab, therefore, was removed again, cleaned in the same way and replaced in a slightly higher position, shifted 6.5 mm compared to the former positions. The third series of observations then gave the values indicated by crosses in figure 9.6.

To each of the curves yielded by these series small vertical translations (in the vertical direction of the figure corresponding to the z-direction in figure 9.2) and small rotations have been applied, in order to obtain a satisfying coincidence. Even a small particle of dust on the contact ring gives rise to such variations of the position of the flat.

The curve corresponding to the third series, moreover, has been shifted in a horizontal direction along a distance of 6.5 mm.

We were gratified to find a very good coincidence of the three independent curves. The mean deviation from the mean curve amounts to only $1.2 \cdot 10^{-6}$ mm or 12 A or about $\lambda/475$.

The integration of the dz/dy-curves gives, as is well known in such cases, reasonably good results.

Further, a parabola (thin line in fig. 9.6) has been calculated to account for a mean curvature of this part of the surface (and along the line b-b of fig. 9.5). The radius of curvature is 2.8 km (concave).

It seems to us that the following can be inferred. (1) Irregularities of a reflecting surface can be studied in this way to about 12 A; it is to be understood that only the relative positions are found of *means* over areas of, as yet, unknown extension, because we have not yet ascertained how far the diffracted light from a given slit has an appreciable influence; a preliminary investigation seems to indicate 5 mm^2 as the order of magnitude of these areas. (2) The mean radius of a reflecting surface can be found, *again without touching the surface and without a reference surface*, over a range of 1 to 2 cm with a precision corresponding to an error in the sagitta of 12 A.

The last conclusion can also be put as follows: In the manner described, deviation from flatness of a reflecting surface is found in an "absolute" way when the surface is 1 to 2 cm in diameter and the radius is smaller than 34 km.

It is not possible without serious variation of the method, to find the thickness of parallel layers. Thus the surfaces a and b in figure 9.7 are not to be distinguished in the manner described.

As has been mentioned, the observations of the same flat in position I give analogous results (See fig. 9.8). Here two series were observed.

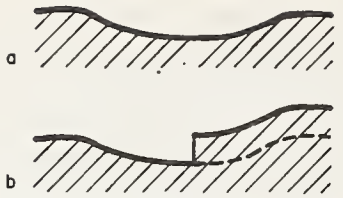


FIGURE 9.7. *Parallel layers not indicated.*

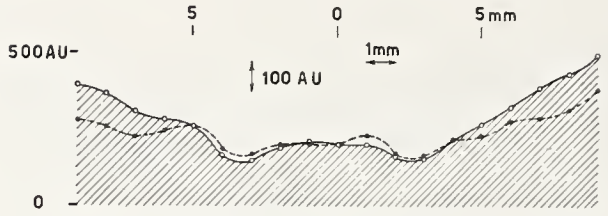


FIGURE 9.8. *Results of two series of measurements on a flat surface.*

For the central portion (11 mm) the mean error of the mean values is 11 Å or $\lambda/520$. The whole range of 18 mm of the measured profile, however, has a larger mean error (32 Å or $\lambda/180$), but inspection of the curves for the first series (open circles) and the second series (dots) of figure 9.8 shows that the general form of the surface is changed between the two series. The mean radius of curvature is 1,4 km for the first series and 2,4 km for the second series. Presumably, the force applied by the springs to press the slab on to the contact ring has been distributed in a somewhat different way in the two cases. This conclusion was strenghtened by an inspection of the construction of the springs, which are not very apt to act evenly when the wedge-like slab has its edge vertical.

A rather crude estimation of the stress made this explanation plausible. In position II the force of the two symmetrically placed springs will be less liable to be disturbed unevenly, even when the slab is shifted upwards.

Curved Surfaces

The study of curved surfaces proceeds along the same lines. Again the observations must be corrected for the tilt α and then can be compared to the theoretical positions, calculated with a preliminary value of the surface curvature.

In figure 9.9 we give the results obtained with the convex and concave parts of a gage of about 4 m radius. Again two positions, I and II, are investigated as in the case of the flat surface.

From the dz/dy-curves the profiles follow by integration and the curves thus obtained are reduced to a mean orientation of the surface. It is to be remembered that here the effects of the *mean* curvature has been left out. The radius for the convex surface is 3987.5 mm, for the concave surface 3984 mm.

On applying the two parts to one another and illuminating with sodium light, figure 9.10 was obtained; for this the surfaces were silvered with a rather thick transparent layer to get sharp interference fringes. The diameter of the gages is 62 mm, but only the central part with diameter 18 mm, where the fringes are nearly straight, has

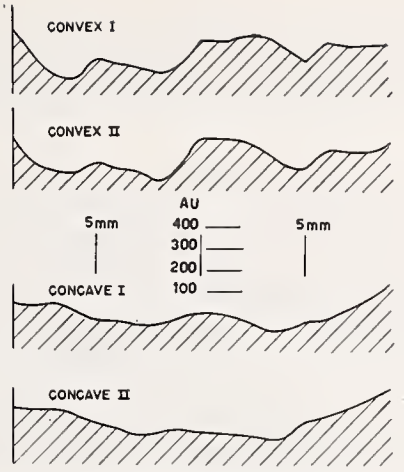


FIGURE 9.9. *Profiles of convex and concave surfaces.*

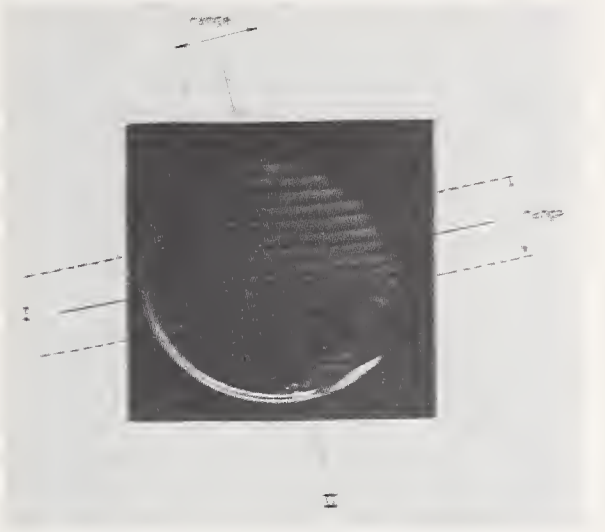


FIGURE 9.10. *Interference test.*

been tested by the diffraction method. The directions I and II are indicated in the figure.

The general trend of the profiles can be recognized in figure 9.10.

As an essential difference between the two methods of observation, the described diffraction method and the interference method, we want to emphasize that with the first the mutual influence of the two glasses is obviated. Indeed, the possibility now arises of investigating the eventual influence of the presence of surface forces on the form of the surfaces.

Further, it is no mean advantage that each surface is studied separately without the aid of an auxiliary surface. In that sense the diffraction method might be called an absolute one.

The spherometry of surfaces with large radii offers no difficulty with the apparatus as described. The precision tends to be too great for practical and workshop uses. For these purposes, however, there is no need to spend too much time on the observation of many lines. If only the two marked or blocked lines are observed a pre-

cision is easily and speedily attained, which is amply sufficient for current use. In this connection it is valuable that no silvering or other treatment of the surfaces is needed to insure sufficient illumination of the image.

Focussing of Lenses or Mirror Systems and Determination of Aberrations

The preceding account of the study of reflecting surfaces applies largely to the investigation of the performance of lenses. Indeed, the reflected wave front has been the object of study. Only by dividing by 2 c were inferences made about the reflecting surface. Essentially the described method is a tool to look into the structure of plane or moderately curved wave fronts, that is, of parallel or moderately divergent or convergent pencils.

The determination of the position of the focus of a well corrected positive system can be done with extreme precision by placing the slit approximately at that point and studying the convergence or divergence of the emergent pencil by means of the grating, observations being made at say 2 m distance from it.

From the mean curvature of the emergent wave front, the distance of the slit from the focus can be calculated with any desired accuracy if the focal length is approximately known.

Assuming a mean error of the sagitta of the wave front of 25 A, the error in the position of the "mean" focus due to these observational errors amounts to:

0.6 micron for a focal length of 100 mm,
16 microns for a focal length of 500 mm,
60 microns for a focal length of 1000 mm.

If the aperture of the system is large enough and its correction sufficiently good this accuracy can even be bettered by making use of gratings of larger breadth.

Of more importance perhaps is the determination of aberrations. Path differences can be measured with a precision of 25 A. Conversion into other measures of aberration is only a question of elementary theory.

As was mentioned before,[5] the (axial) aberrations of a very good objective lens, f=60 cm, aperture 1/10, were determined with the given precision by means of the diffraction method.

A decided advantage over the diffraction methods working with a grating in the neighborhood of the focus is the great simplicity with which the observations are reduced to data of path differences, while on the other hand the apparatus is much simpler in construction than the lens interferometer. It is not known to the author what accuracy can be attained in practice with the interferometer. Only tests of given well corrected systems by the described method, might give some ides of how the accuracies of the two methods compare.

The same applies to the elegant and simple method of Gardner and Bennett.[6]

When aberrations of optical systems are the chief object, the ap-

[5] A. C. S. van Heel, J. Opt. Soc. Am. **40**, 815 (1950).
[6] I. C. Gardner and A. H. Bennett, J. Opt. Soc. Am. **11**, 441 (1925).

paratus, up to now only used incidentally for this purpose, should be reconstructed in order to facilitate adjustment, especially when the observation is to be made photographically. A vertical mounting then seems to be indicated.

Photographic Method

It was felt during the somewhat lengthy observations that a more rapid recording of the diffraction patterns would be an improvement, as for instance, the influence of temperature variations should be obviated. Therefore, exposures on Kodak Ektachrome and Kodachrome films have been made. It appeared that the times of exposure were reasonably short, well below 1 minute.

Measurement of the films showed that practically the same precision is attained (the grain not being observable at the low magnification used), even when the rendering of color is not perfect. The mean error amounts to 9 microns, which compares, not unfavorably, with the values given on page 111 for direct observations.

The regular and irregular variations of the film during the processing can, with careful handling, be kept within this limit. We should make most of our observations photographically, as measurement can then be done at leisure, but development of the films still forms a drawback. The processing is lengthy and cumbersome and sending to a processing factory of the firm still entails too long a delay. We expect that these drawbacks will be solved in future.

Concluding Remarks

We hope to have it made clear that it is advantageous to use Fresnel diffraction for the study of the form of wave fronts in the field of technical optics. Adaptation to special cases can facilitate the production of the desired data with great accuracy. No very special parts are needed, although a good wire grating helps much to simplify the evaluation of the results.

Precision might still be bettered by making use of more than one color transition at each period of the diffraction pattern. An error in the determination of light path differences of 15A seems to be attainable.

The author thanks the members of the optical group for their help, especially H. C. Voorrips, who has made most of the observations.

518-743 O - 73 - 22

10. Image Structure and Test Data

By James G. Baker [1]

It has been the good fortune of the author over a period of 16 years to have been closely concerned with the design, construction, testing, and application of over 50 uniquely different imageforming optical systems and in fact, with the projects requiring such equipment. Many of these systems have been of substantial size and complexity, in both the visual and photographic domains. Several have employed as many as 35 surfaces in an effort to comply with requirements on field angle, aperture, resolution, and basic task.

The instruments in which these systems are mounted list among them large astronomical telescopes, visual telescopes, periscopes, theodolites, spectrographs, navigational sights, and to a very considerable extent, photographic objectives. These latter objectives have had as many as 11 elements and for astronomical purposes have focal lengths up to 320 inches. Apertures have been as large as 33 inches, and formats as large as 27 by 27 inches.

Because of this close association between calculation and the building of equipment, supplemented by observing experience in the use of large astronomical instruments, the author has been able to accumulate a fund of visual testing experience difficult to transcribe on paper. In the case of large aerial lenses the writer has been on numerous high-altitude flights for testing purposes, and has undertaken cold- and pressure-chamber tests of the same lenses. The purpose of all of this procedure was to leave as small a gap as possible between various phases of development of the optical equipment, from conception to final application.

Because of the necessary expenditure of many thousands of hours of computations in the design of these systems, it has been of the utmost importance to the author to form a reasonably consistent picture in his own mind of the many factors governing the performance of the systems for their intended purposes. The follow-through to the final application of any given system, especially where photography is involved, had to be pursued to furnish a safe basis for further work. In this way aerial flight testing as early as 1942 provided the author with first-hand experience on the conditions of aerial photography in the field. The accumulated results over a period of 5 years during World War II have led to what the author hopes are improved compromises in the instruments of the post-war period in which he has been involved.

Unsatisfactory testing conditions encountered during the manufacture of several of the large astronomical Schmidt telescopes in 1938–40 caused the author to set up testing equipment at the Harvard

[1] Research Associate of Harvard College Observatory, and Consultant to the Perkin-Elmer Corp., Norwalk, Conn. Acknowledgment is also made to the Flight Research Laboratory at Wright-Patterson Air Force Base for sponsoring a portion of the work described in this paper, and to the Photographic Laboratory at WPAFB without whose continual cooperation the long term work described here would not have been possible.

College Observatory as soon as circumstances permitted. In 1941 a "testing tunnel" 22 feet long and 4 feet square was constructed. An existing 24-inch aperture f/5 astronomical mirror was mounted on a concrete pier for a collimator. This testing tunnel with its collimator was used for visual examination of a number of instruments built during the war at Harvard under the NDRC program. Photographic testing and photomicrographic enlargements were employed at infrequent intervals to supplement the extensive visual work. Photographs of star trails and planets were occasionally made.

No systematic laboratory testing by photographic means was employed at Harvard but the writer maintained close contact with the copious testing accomplished under OSRD contracts at the Massachusetts Institute of Technology, Mount Wilson Observatory, the National Bureau of Standards, the Eastman Kodak Company, and at Wright Field. Some of the Harvard lenses were tested independently at several of these places, and abundant data are available.

If one adds the extensive testing carried on in England and Canada during the war, the volume of data becomes very great indeed. The material is far from being obsolete. Any one interested in the detailed performance of a variety of optical systems would do well to have these data at hand. It would be very difficult even to reproduce the data. The author has drawn liberally on this information in the postwar years.

During the fall of 1945 a number of reports were prepared by the staff of the NDRC project at Harvard on the optical developments carried on there during the period from 1940 to 1945. In the first 3 months of 1946 a further opportunity was presented to the writer in reviewing the testing work and optical developments throughout the nation, and summarizing these as a coauthor in the NDRC volume on instruments. This period of review led to the presentation of a conception of the problems of image quality, as touching on resolution and contrast, for purposes of optical design.

It was pointed out that the lens design could be adjusted in the final balancing of aberrations to go with the level of resolution to be expected as an average in the air, and that by putting as much light energy as possible within a circle of confusion determined by the adopted level of resolution, the designer could improve the *microscopic contrast* at that level. In fact, these considerations were made use of in 1941 in the design of a 40-inch f/5 telephoto. It is very probable that the Zeiss designers were doing more or less the same thing before the turn of the century, inasmuch as lens bench practice almost inevitably requires it.

Users of lens systems often tend to look at the problem from another side, where a different focus may be found to exist for maximum resolution and for maximum contrast. However, it should be observed that the user of a finished system has far fewer variables at his disposal than the designer, and in fact is faced with simply the question of what focus to use. It is incumbent on the designer from the start of his work to control the image in matters of contrast and resolution. Consequently, it is vital that the designer understand as much as possible the details of image structure and test data for the guidance of his efforts and the success of his output. Where the design results can be supplemented by final laboratory adjustments of the finished lens, so much the better, and the old-time optical manufacturers made many test models for the purpose.

Visual Testing

It is well known that in the presence of image aberrations there may be a very distinct difference between the visual and photographic performance. For convenience, however, it is often necessary for the observer to omit the photographic process and to judge from the visual image insofar as he can what the probable photographic performance will be. In systems having very little aberration the visual evaluation can be applied with confidence. The observer looks for a sharply defined image of a source pinhole, even though the pinhole image may be embedded slightly in unfocused light. Many observers prefer to use a brilliant star point, but the writer has found that it is difficult to judge the distribution of light in the image, or the relation between the intense central point surrounded by diffraction rings and the probable photographic performance.

In a few cases where absolute perfection is required in the image within the possibilities of wave optics, a detailed analysis of the structure of the diffraction rings may prove necessary. Photographic systems do not require this kind of analysis directly. One is more interested in whether a bright but not dazzling pinhole, whose image in the focal plane measures perhaps 0.020 mm, presents a "clean" edge even in the presence of color aberrations or monochromatic aberrations. If the color correction by design differs appreciably from the visual range, one is obliged to make use of a suitable filter in which the response of eye and photographic emulsion receive proper weighting. Obviously, this process becomes more and more unreliable in outlying regions of the spectrum. If visual testing is to be employed at all under these circumstances, one must check the performance in the nearest accessible color region for comparison with the calculated design performance in the same region. Obviously, the designer of the system ordinarily prefers to do this kind of checking himself, owing to his familiarity with the aberrations and their color gradients. The success of the venture will depend greatly on the fund of experience available to the designer at the lens bench.

In 1943 a demonstration instrument was made at Harvard to show the difference between the performance of a two-element achromatic doublet, designed according to purely geometrical optics for smallest axial circle of confusion, and of a similar two-element achromatic doublet designed according to strictly physical optics. Two f/4 objectives were constructed from the same melts of optical glass and to the same test plates. The curves and thicknesses were all within critical tolerances. The only difference between the pair of doublets lay in the rear surface radius. The one differed by a calculated number of fringes from the other to span the gap between geometrical and physical correction. On the test bench the performance of the "physical optics" instrument proved to be in close agreement with the usual diffraction image structure and the image was sharply defined. The "geometrical optics" lens showed quite marked spherical aberration in the image.

Again in 1943 a "spherical-aberration analyzer" was constructed to be used at the testing tunnel for studying the image structure corresponding to varying amounts of third- and fifth-order spherical aberrations and stop diameters adjusted into the system. The two-element system was mounted in micrometer form so that it was possible to

obtain the exact design figures corresponding to any given situation of the spherical-aberration analytical terms and focal ratio. The visual information obtained at the testing tunnel was then applied in the further design work of the period. No photographic testing was accomplished with this instrument, but such testing would even now be very worthwhile.

Another brief but interesting experiment was performed in 1943 along the same lines for the purpose of information on balancing of aberrations for contrast and resolution. Even in 1941 the writer had adopted a rule in design of getting as much light from the aperture into a 20-micron circle as possible and the remaining light as far away as possible, for the purpose of maintaining contrast at a fairly critical resolution level. This rule might be recast into modern language in much more detail. The experiment performed in 1943 tended to confirm this rule adequately well. In this experiment, a well made positive simple lens with spherical surfaces was mounted as a telescope objective and used with a microscope as eyepiece. A large amount of under-corrected spherical aberration was necessarily present.

A boy was stationed about 40 feet away in a dark room with his face illuminated by a sodium-vapor lamp. His face was observed through the telescopic system with the 20X microscope mounted on a micrometer slide. Various focal settings were averaged for the reading at which the boy's face presented the most pleasing compromise between contrast and resolution.

Thereafter, a small pinhole was placed over the sodium light at the same distance and used as a new object. It was observed that as the microscope was moved from decidedly outside focus, the image of the pinhole first formed at a rather definite microscope reading. Here, the pinhole image appeared for the first time to present an edge or smooth outline, but was surrounded by a very large single flare of undercorrected primary spherical aberration.

As the microscope was moved closer to the lens in small increments, the pinhole image measuring about 0.020 mm in diameter became brighter and more sharply defined. This fact means that a larger portion of the lens aperture provided light within the Rayleigh limit with a corresponding increase in the brightness of the pinhole image and sharpness of edge. Finally, just beyond the setting for maximum sharpness of the pinhole image, a slight secondary flare began to grow from the edge of the pinhole. At this setting the Rayleigh limit was so far exceeded by an intermediate zone of the aperture as to cause the resulting caustic to add the furry edge to the pinhole image. The pinhole with the surrounding small secondary flare still lay embedded in a much greater flare of spherical aberration from the outer zone of the lens. *The micrometer reading previously ascertained for the best over-all rendition of the boy's face agreed closely with the focal setting on the pinhole showing the slight secondary caustic flare.* With the pinhole image having a diameter of about 0.020 mm, one might judge the diameter of pinhole image plus flare as about 0.030 mm. Farther inside or outside of this focal setting, one observed that the image of the boy's face grew worse, though on the outside at least the edge of the pinhole image itself was sharper.

All of this can be interpreted to mean that for a particular type of object, such as the face of a boy, best pictorial results are given by a certain degree of contrast and resolution. If the resolution were to

be emphasized, the contrast suffers. If a larger blur circle is used to bring a higher percentage of the light into the expanding circle, the resolution may fall as the contrast increases. *The picture "quality" appears then to be a proper balance between contrast and resolution for a given type of object and scale of object in the focal plane, both depending on the quality of the optical instrument.* A fairly high contrast is required for large and small detail to obtain a pleasing pictorial effect, but there must be enough resolution too for defining what is sought in the picture.

The visual testing of the many different optical systems developed at Harvard consisted mostly of examining the image of a pinhole formed from the 24-inch mirror collimator and of such a size at the focus of the collimator to give a 0.020 mm image in the focal plane of the lens under test. The microscope power usually ranged from 50X to 100X with no attempt made to increase the magnification much beyond the merit of the image. Image characteristics were judged from the appearance of the pinhole image at varying off-axis angles and with different filters. Quite frequently, three-line test patterns were examined for guidance, but most judgments as to the residual defects of the lens performance were derived from pinhole images. Differential corrections to airspaces and occasionally to radii were determined from the pinhole images, via design calculations, and applied to the lens with subsequent rechecking. By this process the best compromises were sought and achieved in the final lens.

It has been the author's experience and perhaps bias that visual testing can be much more exhaustive of a system's ultimate performance than photographic testing in the focal plane. No doubt, photoelectric or photomicrographic methods can be substituted for visual testing in order for one to obtain quantitative data. However, in the focal plane itself use of the turbid photographic emulsion spoils the optical image to such a degree that the subtle aspects of the image structure are lost. Photographic testing in the focal plane for photographic objectives is, nevertheless, highly desirable on the basis that acceptance tests on performance ought to be based on what the instrument is supposed to do. If a photographic objective provides nearly perfect visual images, it may be over-designed and too complicated. If an objective is afflicted by many hybrid aberrations, the observer may find it extremely difficult from a visual examination to judge the photographic performance. The average lens of good quality lies between these two extremes.

Visual testing on a lens bench or in a testing tunnel is often necessary for the designer to determine whether his lens is built to his specifications and whether further improvement can be achieved. Ideally, a design ought to go through a prototype stage and at least one differential correction if necessary before the final optimum performance can be determined. With the introduction of modern high-speed computing and with a better quantitative knowledge of image structure, designers may be able before long to eliminate need for a differential correction on any pronounced scale. As it is, an individual with considerable design and testing experience can minimize the gap between design and performance by judgment alone.

The author feels that it is relatively unproductive for a new type of optical system to receive exhaustive testing before the designer himself has done all he wishes on improving the system. It is easy

to make rather small adjustments that will change the apparent structure of the image very appreciably. The designer can look at the image and read into it how it would look if he made certain alterations in the design. The nondesigner can only accept it for what it is, and may spend too long a time reaching conclusions that may be invalid for the next lens off the production line.

Close attention to detail near the end of fabrication of a lens system may result in a gain in performance by a factor of two or more. Such an adjustment may, for example, convert an effective image 0.020 mm in diameter into one 0.010 mm in diameter without any loss of contrast. There may be compromises possible that will bring about an average high-level performance over the field instead of one excessively high on axis and poor near the edge of the field.

The balancing for maximum contrast at a desired resolution level is a case in point. The designer may be tempted to demonstrate a high visual or even photographic resolution, whereas better pictures might be obtained at a more conservative resolution level by improving the microscopic contrast.

In aerial photography pictures having a resolution of about 10 lines/mm, with good contrast at 5 lines/mm or coarser, appear to be of excellent quality to the uncritical observer. If such a performance is taken as the norm, then one overlooks a whole realm of complex considerations in image structure at a level of 30 lines/mm or finer. Visual study of the subtleties of image structure beyond 30 lines/mm becomes then relatively unimportant. For quality work, however, the question of whether a system has received adequate final differential correction and adjustment is still vital. Otherwise, an elaborate and expensive system inadequately made may perform no better on the average than a simpler system carefully adjusted.

Image Structure

An optical image formed well off-axis in a large fast lens is a hybrid structure resulting from the contributions of many types of image errors or aberrations in varying degrees according to the design. Figure 10.1 gives a schematic array of the many conditions at work, each of which in turn requires a number of power series terms for analytical representation. Moreover, the fifth order indicated is still quite insufficient to give an accurate description of the image errors that occur in such a lens as a 12-inch f/2.5 for 9 by 9 format, and one must draw either on still higher order terms with slow convergence or abandon series methods altogether. In such cases a useful procedure is to make use of the first, third and fifth order for analytical control but to analyze image structure by exact ray-tracing.

In some of the systems designed at Harvard all terms of the first, third, and fifth orders given in figure 10.1 have been satisfied with the exception of a residual oblique spherical aberration and secondary spectrum. In others, the latter two aberrations have been controlled as well. An aberration is spoken of as being satisfied when its contribution to image structure taken together with any other balancing residuals falls within a tolerance based on performance requirements. The coefficient of such an error is necessarily small or close to zero if a fast lens is involved, a fact that often implies a zero in the solution, and some control of the aberration by the designer.

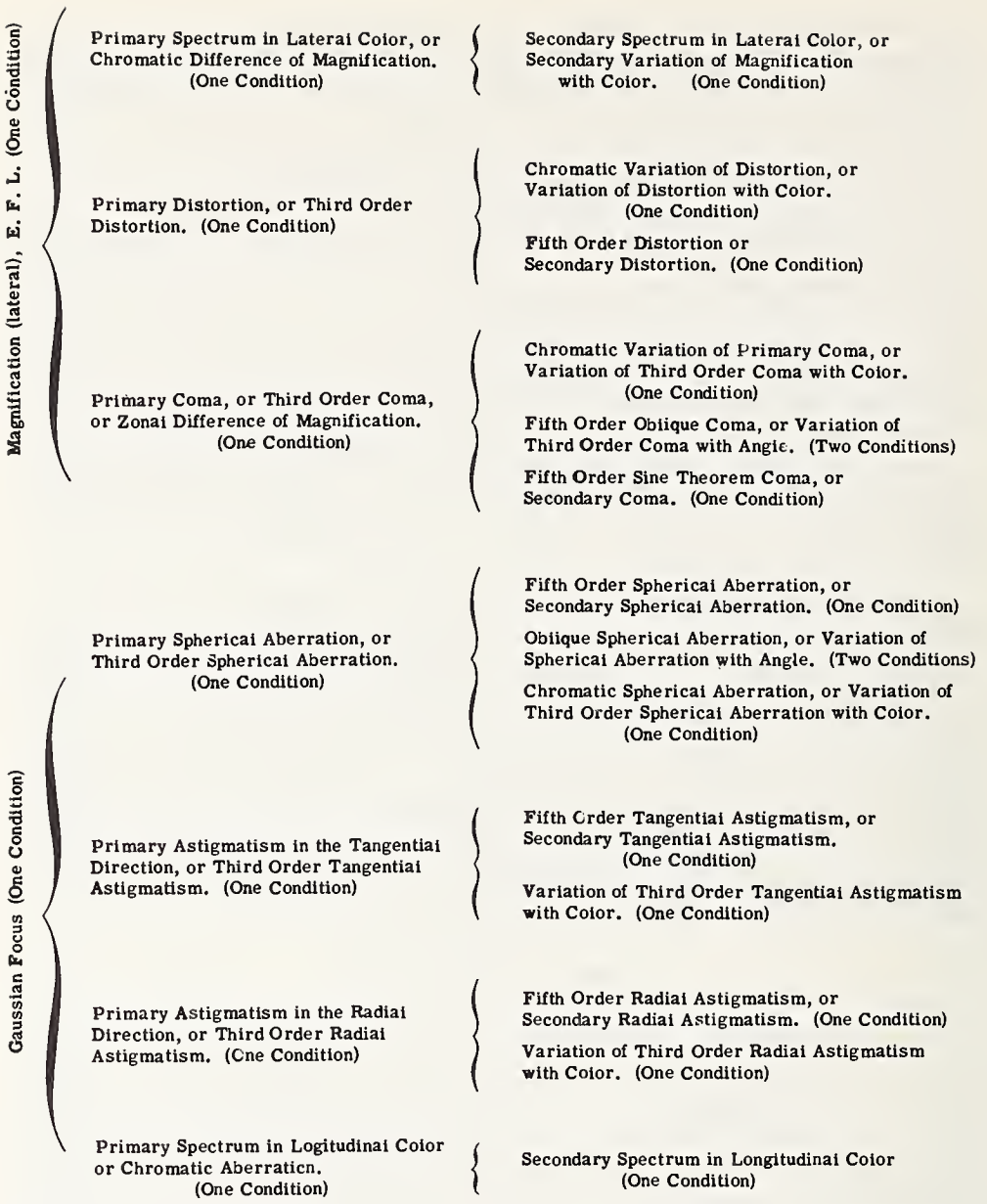


FIGURE 10.1. *A schematic array of the 25 conditions affecting lens performance through the fifth order of approximation.*

Fifth-order field curvature and astigmatism can often be controlled in spite of the complexity of the effort. The asymmetrical terms of the fifth order come out quite often to be small or negligible. In a few cases it has proved possible to eliminate the oblique spherical aberration of the fifth order and leave residuals of the seventh or higher orders. In all cases the residual higher order aberrations have been balanced as well as judgment allows, taking diffraction image formation and vignetting into account, and weighted according to the requirements of the problem.

Figure 10.2 shows a typical array of power series terms for the third, fifth and a portion of the seventh orders, evaluated for the case of a 40-inch f/5 telephoto lens for 9 by 9 format. The coefficients are in millimeters when the normalized field-angle θ is equal to 1 in the

Δy		Δz	
Third Order			
Distortion	$0.186\ \theta^3$		
Astigmatism	$-0.001\ \theta^2 y$	Astigmatism	$-0.000\ \theta^2 z$
Coma	$-0.010\ \theta y^2$ $-0.003\ \theta z^2$	Coma	$-0.007\ \theta yz$
Spherical Aberration	$-0.061\ y^3$ $-0.061\ yz^2$	Spherical Aberration	$-0.061\ z^3$ $-0.061\ y^2 z$
Fifth Order			
Distortion	$(-0.136 + 0.00079\ c_N)\ \theta^5$		
Astigmatism	$(0.334 + 0.00187\ c_N)\ \theta^4 y$	Astigmatism	$(0.111 + 0.00037\ c_N)\ \theta^4 z$
Coma	$(0.495 + 0.00178\ c_N)\ \theta^3 y^2$ $(0.166 + 0.00036\ c_N)\ \theta^3 z^2$	Coma	$(0.332 + 0.00071\ c_N)\ \theta^3 yz$
Oblique Spherical	$(0.084 + 0.00084\ c_N)\ \theta^2 y^3$ $(0.194 + 0.00050\ c_N)\ \theta^2 yz^2$	Oblique Spherical	$(0.194 + 0.00050\ c_N)\ \theta^2 y^2 z$ $(0.304 + 0.00017\ c_N)\ \theta^2 z^3$
Sine Coma	$(0.004 + 0.00020\ c_N)\ \theta y^4$ $(0.005 + 0.00024\ c_N)\ \theta y^2 z^2$ $(0.001 + 0.00004\ c_N)\ \theta z^4$	Sine Coma	$(0.003 + 0.00016\ c_N)\ \theta yz^3$ $(0.003 + 0.00016\ c_N)\ \theta y^3 z$
Fifth Order Spherical	$(0.079 + 0.00002\ c_N)\ y^5$ $(0.158 + 0.00004\ c_N)\ y^3 z^2$ $(0.079 + 0.00002\ c_N)\ yz^4$	Fifth Crder Spherical	$(0.079 + 0.00002\ c_N)\ z^5$ $(0.158 + 0.00004\ c_N)\ y^2 z^3$ $(0.079 + 0.00002\ c_N)\ y^4 z$
Seventh Order			
Distortion	$-0.023\ \theta^7$		
Astigmatism	$0.070\ \theta^6 y$	Astigmatism	$0.022\ \theta^6 z$
Coma	$0.054\ \theta y^6$		
Spherical Aberration	$-0.015\ y^7$		

FIGURE 10.2. *A typical array of coefficients of third, fifth and seventh orders of approximation.*

corner of the format, and where the maximum values of y or z are equal to 1 in the entrance pupil. The quantities Δy and Δz refer to the intercept of a ray, defined by θ, y and z in the entrance pupil, in the focal plane in millimeters relative to the Gaussian point of zero distortion. The quantity c_N refers to an aspheric coefficient of the last surface of the system. The series was evaluated before reintroducing first- and third-order residuals for balancing of aberrations. Figure 10.3 shows the image form calculated from the series of figure 10.2, but with balancing included. The numerical quantities are in millimeters and indicate that the effective image is of the order of 0.020 mm in diameter. The image structure in figure 10.3 is purely geometrical optics, and by physical optics the image is very nearly the usual spurious disk surrounded by diffraction rings with very little extra light in the rings. The vertical elongation of the image in the skew direction was indeed observed in the lenses built, but, neverthe-

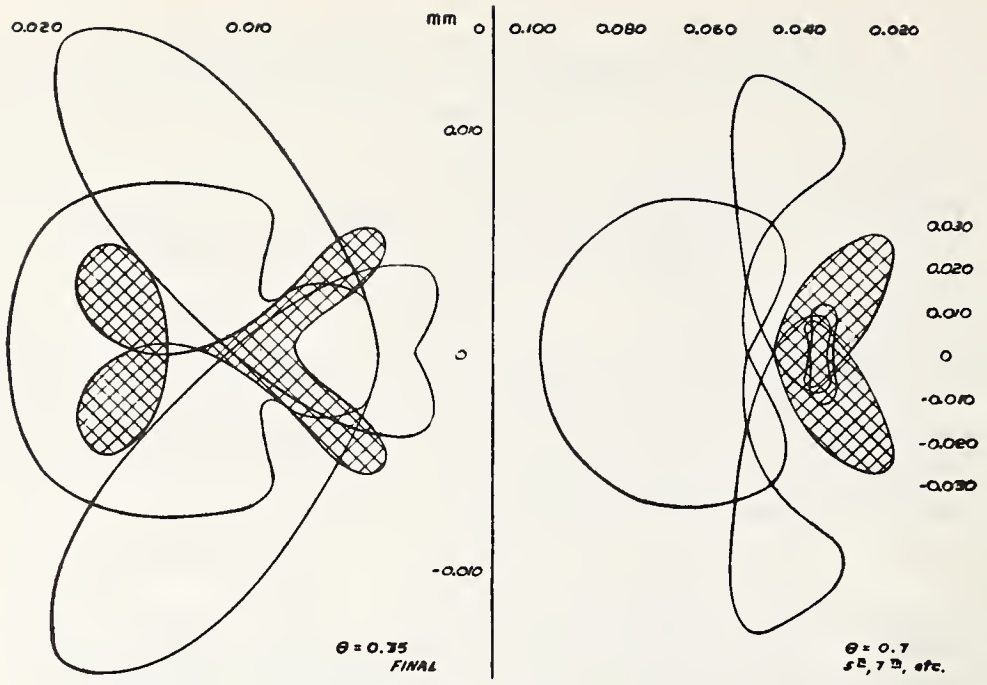


FIGURE 10.3. *Enlarged geometrical image calculated from series for a 40-inch f/5 telephoto lens.*

less, resolution on Super-XX emulsion at high contrast reached a level of 45 lines/mm.

In the meridional plane, owing to maximum deviations of the rays at the rim of any given surface, the aberrational elongations in the image can easily dominate over the skew aberrations. Thus, comatic flares and oblique spherical aberration usually cause a faster deterioration of the image in the meridional plane than do the skew terms. However, owing to the greater ease of designing in the meridional plane, one often finds near the end of his work that the meridional elongations in the image have been brought nearly to zero, but that the image squeezes out still in the vertical direction, above and below the meridional plane. Figure 10.3 indicates that this residual defect occurred in the case of the 40-inch f/5 lens, but other examples at hand show the effect even more strongly. With the advent of faster computing methods by automatic machine, it is probable that the skew direction will come in for more evaluation during the course of a design with improved chances for better image structure in all azimuths.

Much more complicated off-axis images have been encountered than can be shown here. Variations with color have to be considered, and if present in appreciable amount, the effect on the location of focus and image structure must be dealt with during the course of the design. The final balancing of aberrations must also involve contrast-resolution factors, aperture, field, and color errors, weighted according to requirements. It is clear that a designer must be as familiar as possible with the behavior of systems of known corrections in order to determine in advance just what he must do for optimum results. It is probable that German designers of 50 years ago had a very good idea of image structure and balancing, but kept such information only as part of their experience or as company data. Nowadays we may be plowing up much old ground, but this time the growing

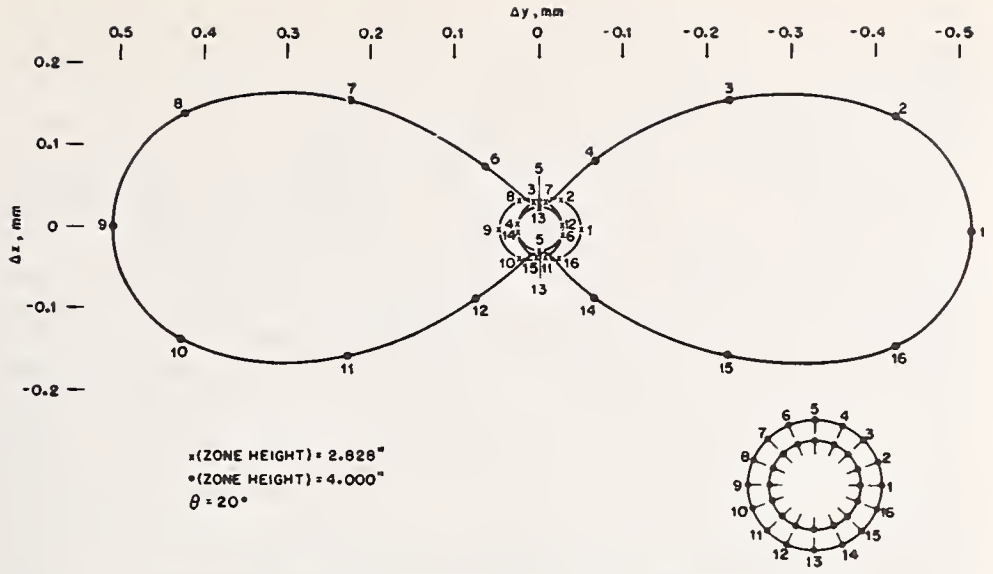


FIGURE 10.4. *Image errors of an f/1 Schmidt camera.*

knowledge of physical optics, photographic effects, psychological factors and the like are undoubtedly going to lead to scientific results available to everyone.

Figure 10.4 shows the oblique errors of a standard Schmidt system as calculated from series and verified by selected rays. The aberration is almost entirely oblique spherical aberration caused by foreshortening of the oblique beam and by the increase in the refractive strength of the plate for inclined rays. Figure 10.4 shows that the unvignetted off-axis image errors of an 8-inch f/1 Schmidt camera are by no means inappreciable. Star photographs made with a 12-inch f/1 Schmidt demonstrate the radial elongation of the images very clearly.

During the summer of 1940 the author made a prolonged attempt to improve the Schmidt by introducing more correcting plates. Figure 10.5 shows an array of terms for the various aberrations of the third and fifth orders for the generalized system of the same figure. Algebraically, the equations hold for a solid glass type of Schmidt system with air-lens aspheric surfaces that may have finite bendings. Aspheric terms on the mirror are also included. One needs only to substitute proper values for the indices of refraction in order to obtain valid equations for a mirror in air, combined with three thin correcting plates that may have finite bendings.

The equations in figure 10.5 show many relationships of interest with respect to the fifth order, and demonstrate how very lengthy an explicit solution through the fifth order for even a Cooke triplet would be. The Schmidt system is about as tractable as any that can be handled in this way. No attempt has been made to find a solution to the analytical expressions, for in order to improve on the ordinary Schmidt, one must have all of the expressions zero simultaneously. Physical considerations can circumvent the use of such complicated expressions, and indeed in 1943 the author came across a rather simple solution to the problem in the form of the "double shell" Schmidt system with a correcting plate of higher order asphericity.

Such a system was developed for television purposes in December, 1943 for the Perkin-Elmer Corp. Shell systems were designed independently elsewhere, as information coming to hand at the end of the war indicated, with some as early as 1940. In the spring of 1946 the double-shell system was further elaborated at Harvard into the Super-Schmidt Meteor Cameras to include an achromatized correcting plate and arranged to have the light coming through the second shell twice with a favorable effect on the image quality.

These "Super-Schmidt Meteor Cameras" now in successful operation cover a field of 52 degrees on specially curved film, have a speed of f/0.65, a color range from 3700-5000 and a clear aperture of 12.2 inches for an 8-inch focal length. The effective star image appears to be of the order of 0.025 mm in diameter over most of the field. The extremely low f/number causes focusing to be extremely critical to the extent that a focal error of 0.010 mm already shows to disadvantage. The secondary spectrum of this system is negligible by ordinary standards, but the focal tolerance here is so critical that the image is enlarged appreciably by the secondary spectrum of the correcting plate and shell combination. There appears to be no easy solution to the defect caused by the secondary spectrum, inasmuch as the requisite crystalline materials or special glasses are either not available or would have achromatizing curves too strong to be useful.

The observed performance of the Super-Schmidt Meteor Cameras is all the more striking, because the central 7.4 inches diameter of the 12.2-inch aperture is blocked out by the film and film-holder. The sharp images are thus formed from an annular entrance pupil, a fact that causes the focal tolerance to be all the more critical. This outer annulus is the region of the worst aberrations, and the core of the image normally formed by the inner zones is missing. Nevertheless, the color error remains the limiting defect in the center of the field, and only slight comatic wedges defined by the vignetting characteristics appear in the outer field.

Figure 10.6 shows a cross-sectional cut-away of a 36-inch f/8 apochromatic lens for a 9 by 9 format, designed in 1943. The apochromatic correction was achieved through a combination of a fluorite element with light crown elements elsewhere. This fluorite element was cemented between two glass elements in the central negative component. The design of this apochromatic system was particularly interesting in view of the first-time use of a large element of fluorite in a lens system. The standards of the design had to be set at a high level of performance. There is no gain in eliminating secondary spectrum if the monochromatic aberrations are made larger thereby. The true apochromat is corrected to have a common focus for three colors and is also corrected for spherical aberration and coma for two colors. This aerial lens also had to have flat field, freedom from astigmatism, and distortion, and had to have an off-axis image sensibly free of oblique spherical aberration.

Figure 10.7 shows the residual meridional aberrations. The inner part of the field shows some outward coma, which at ¾ field has practically disappeared with the up-swing at the right vignetted away. The very corner image comes from the portion of the bottom curve between the vertical lines, the rest being vignetted away, and is evidently astigmatic. Figure 10.8 shows the effective field curves. The tangential astigmatism causes trouble only in the very corner.

FIGURE 10.5. *Third- and fifth-order conditions for a multiple Schmidt camera.*

$y_8 =$ Series Expansion for Solid Glass Schmidt with 3 Aspheric Correcting Air Lenses of Finite Bendings and Aspheric Mirror.

$$x_i = a_i(y^2+z^2) + b_i(y^2+z^2)^2 + c_i(y^2+z^2)^3 + \cdots$$

(Stop at d_0)

$$[\mu_0] = \frac{1}{2n}$$

$$[y_0] = 0$$

$$[\mu_0^3] = \frac{1}{n^3}\left[\frac{1}{16}B_3 + B_5 - \frac{27}{2}S_7\right]$$

$$[\mu_0^2 y_0] = \frac{1}{n^2}\left[\frac{3}{8}B_3 + 3B_5 - 27\,S_7\right]$$

$$[\nu_0^2 y_0] = \frac{1}{n^2}\left[\frac{1}{8}B_3 + B_5 - 9\,S_7\right]$$

$$[\mu_0 y_0^2] = \frac{1}{n}\left[\frac{3}{4}B_3 + 3B_5 - 18\,S_7\right]$$

$$[y_0^3] = \left[B_1 + \frac{1}{2}B_3 + B_5 - 4\,S_7\right]$$

$$[\mu_0^5] = \frac{1}{n^5}\Bigg[+ \frac{1}{8} + \frac{1}{32}n_3 + \frac{5}{64}(n_3+2)\,a_3\,(1+B_3) + \frac{3}{64}(n_3-1)\,a_3^2\,(1+B_3)$$
$$+ \frac{5}{2}(n_5+2)\,B_5 a_5 + 3\,(n_5-1)\,B_5 a_5^2$$
$$+ \left(\frac{3}{32}+\frac{1}{32}n_3\right)B_3 + \left(\frac{33}{16}+\frac{1}{2}n_5\right)B_5 - \frac{27}{16}S_7 - \frac{1}{8}S_8$$
$$+ \frac{3}{16}B_3B_5 - \frac{27}{8}B_3\,S_7 - 27\,B_5\,S_7 + \frac{9}{256}C_3$$
$$+ \frac{1}{2}C_5 - \frac{243}{32}T_7\Bigg]$$

$$d_0 = d_1 = d_3 = d_5 = 0$$

$$d_2 = d_4 = d_6 = -d_7 = \frac{1}{2}$$

$$a_1 = a_2 \quad a_3 = a_4 \quad a_5 = a_6$$

$$B_1 = -2\frac{1}{n}(n-1)(b_2 - b_1)$$

$$B_3 = -4\frac{1}{n}(n-1)(b_4 - b_3) - 1$$

$$B_5 = -2\frac{1}{n}(n-1)(b_6 - b_5)$$

$$C_1 = -6\frac{1}{n}(n-1)(c_2 - c_1)$$

$$C_3 = -\frac{8}{3}\frac{1}{n}(n-1)(c_4 - c_3) - 1$$

$$C_5 = -6\frac{1}{n}(n-1)(c_6 - c_5)$$

$$[\mu_0{}^4 y_0] = \frac{1}{n^4}\Big[+ \frac{3}{4} + \frac{3}{16}n + \frac{5}{8}(n+2)\,a_3\,(1+B_3) + \frac{15}{32}(n-1)a_3{}^2\,(1+B_3)$$
$$+ 10\,(n+2)\,B_5a_5 + 15\,(n-1)\,B_5a_5{}^2$$
$$+ (\frac{3}{8} + \frac{3}{16}n)\,B_3 + (\frac{57}{8} + \frac{3}{2}n)\,B_5 - \frac{63}{2}S_7 - \frac{1}{8}S_9$$
$$+ \frac{3}{2}B_3B_5 - \frac{99}{4}B_3\,S_7 - 117.\,B_5\,S_7 + \frac{45}{128}C_3$$
$$+ \frac{5}{2}C_5 - \frac{405}{16}T_7\Big]$$

$$[\nu_0{}^4 y_0] = \frac{1}{n^4}\Big[+ \frac{1}{8} + \frac{1}{16}n + \frac{1}{8}(n+2)\,a_3\,(1+B_3) + \frac{3}{32}(n-1)\,a_3{}^2\,(1+B_3)$$
$$+ 2\,(n+2)\,B_5a_5 + 3\,(n-1)\,B_5a_5{}^2$$
$$+ (\frac{1}{8} + \frac{1}{16}n)\,B_3 + (\frac{15}{8} + \frac{1}{2}n)\,B_5 - \frac{21}{2}S_7 - \frac{1}{8}S_9$$
$$+ \frac{1}{4}B_3B_5 - \frac{15}{4}B_3\,S_7 - 21\,B_5\,S_7 + \frac{9}{128}C_3$$
$$+ \frac{1}{2}C_5 - \frac{81}{16}T_7\Big]$$

Continued on page 130.

FIGURE 10.5. *Third- and fifth-order conditions for a multiple Schmidt camera—Continued*

$$[\mu_o^3 y_o^2] = \frac{1}{n^3}\left[+\frac{3}{2}+\frac{3}{8}n + \frac{15}{8}(n+2)a_3(1+B_3) + \frac{15}{8}(n-1)a_3^2(1+B_3) + 15(n+2)B_5a_5\right.$$
$$+ 30(n-1)B_5a_5^2 + (\frac{3}{8}+\frac{3}{8}n)B_3 + (\frac{213}{16}+\frac{3}{2}n)B_5 - \frac{273}{2}S_7$$
$$\left.+\frac{75}{16}B_3B_5 - 69B_3S_7 - 201B_5S_7 + \frac{45}{32}C_3 + 5C_5 - \frac{135}{4}T_7\right]$$

$$[\mu_o^3 z_o^2] = \frac{1}{n^3}\left[+\frac{1}{4}+\frac{1}{8}n + \frac{3}{8}(n+2)a_3(1+B_3) + \frac{3}{8}(n-1)a_3^2(1+B_3) + 3(n+2)B_5a_5\right.$$
$$+ 6(n-1)B_5a_5^2 + (\frac{1}{8}+\frac{1}{8}n)B_3 + (\frac{47}{16}+\frac{1}{2}n)B_5 - \frac{61}{2}S_7$$
$$\left.+\frac{17}{16}B_3B_5 - 17B_3S_7 - 43B_5S_7 + \frac{9}{32}C_3 + C_5 - \frac{27}{4}T_7\right]$$

$$[\mu_o^2 y_o^3] = \frac{1}{n^2}\left[+1+\frac{1}{4}n + \frac{5}{2}(n+2)a_3(1+B_3) + \frac{15}{4}(n-1)a_3^2(1+B_3) + 10(n+2)B_5a_5\right.$$
$$+ 30(n-1)B_5a_5^2 + \frac{1}{2}n_1B_1 + \frac{1}{4}nB_3 + (\frac{33}{2}+\frac{1}{2}n)B_5 - 196S_7$$
$$+\frac{3}{8}B_1B_3 + 6B_1B_5 - 81B_1S_7$$
$$\left.+\frac{57}{8}B_3B_5 - 90B_3S_7 - 171B_5S_7 + \frac{45}{16}C_3 + 5C_5 - \frac{45}{2}T_7\right]$$

$$[v_o^2 y_o^3] = \frac{1}{n^2}\Big[+ \frac{1}{4} n + \frac{1}{2}(n+2)\, a_3\,(1+B_3) + \frac{3}{4}(n-1)\, a_3^2\,(1+B_3) + 2(n+2)\, B_5 a_5$$
$$+ 6(n-1)\, B_5 a_5^2 + \frac{1}{2} n_1 B_1 + \frac{1}{4} n\, B_3 + (\frac{7}{2} + \frac{1}{2} n)\, B_5 - 44\, S_7$$
$$+ \frac{1}{8} B_1 B_3 + 2\, B_1 B_5 - 27\, B_1\, S_7$$
$$+ \frac{11}{8} B_3 B_5 - 18\, B_3\, S_7 - 33\, B_5\, S_7 + \frac{9}{16} C_3 + C_5 - \frac{9}{2} T_7 \Big]$$

$$[\mu_o y_o^4] = \frac{1}{n}\Big[+ \frac{5}{2}(n+2) B_1 a_1 + \frac{5}{4}(n+2)\, a_3\,(1+B_3) + \frac{15}{4}(n-1)\, a_3^2\,(1+B_3) + \frac{5}{2}(n+2)\, B_5 a_5$$
$$+ 15(n-1)\, B_5 a_5^2 + \frac{45}{4} B_5 - 115\, S_7$$
$$+ \frac{3}{2} B_1 B_3 + 12\, B_1 B_5 - 108\, B_1\, S_7$$
$$+ \frac{21}{4} B_3 B_5 - 54\, B_3\, S_7 - 72\, B_5\, S_7 + \frac{45}{16} C_3 + \frac{5}{2} C_5 - \frac{15}{2} T_7 \Big]$$

$$[y_o^5] = \Big[+ 3(n-1) B_1 a_1^2 + \frac{3}{2}(n-1)\, a_3^2\,(1+B_3)$$
$$+ 3(n-1)\, B_5 a_5^2 + 3\, B_5 - 24\, S_7$$
$$+ \frac{3}{2} B_1 B_3 + 6\, B_1 B_5 - 36\, B_1\, S_7$$
$$+ \frac{3}{2} B_3 B_5 - 12\, B_3\, S_7 - 12\, B_5\, S_7 + \frac{1}{2} C_1 + \frac{9}{8} C_3 + \frac{1}{2} C_5 - T_7 \Big]$$

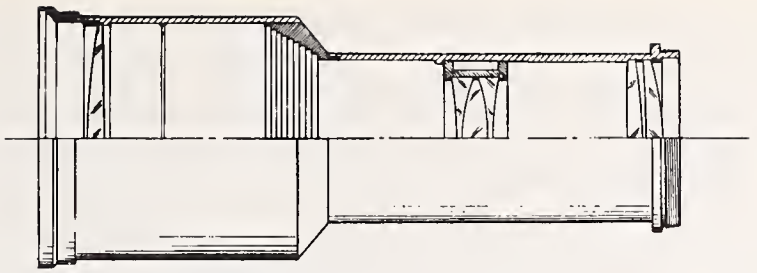


FIGURE 10.6. *A cut-away view of a 36-inch f/8 apochromatic lens in which a single small fluorite element is used.*

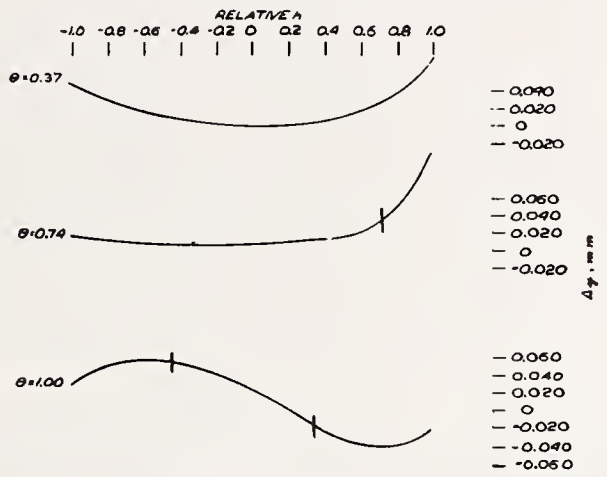


FIGURE 10.7. *The aberrations of the 36-inch f/8 apochromat in the meridional plane.*

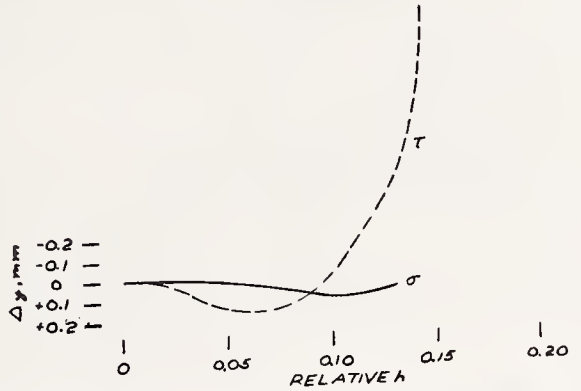


FIGURE 10.8. *The radial and tangential image surfaces of the 36-inch f/8 apochromat.*

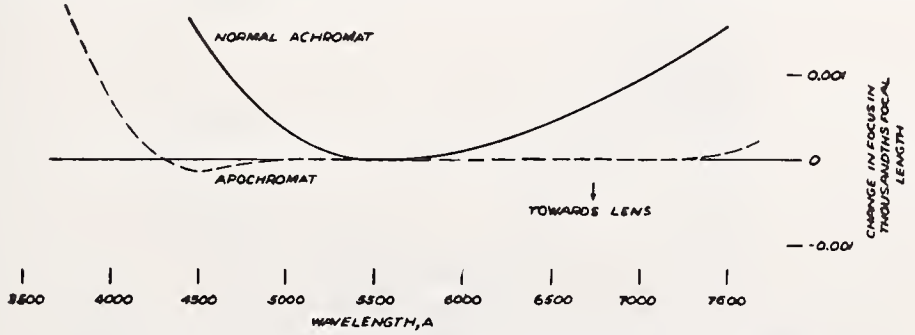


FIGURE 10.9. *The color curve of the 36-inch f/8 apochromat.*

518-743 O - 73 - 23

The performance of this system turned out to be quite satisfactory even for astronomical purposes over a 9-inch diameter circle of the field. The apochromatic correction was fully verified by independent tests at Mount Wilson, and at Eastman Kodak Co. Figure 10.9 shows the color curve of the system as evaluated on the optical bench at Mount Wilson. Diffraction came to the aid of the residual image defects shown in figure 10.7 and figure 10.8, so that the star images remained sharp even to photomicrographic magnifications up to 80X.

Figure 10.10 shows the color curve of a 100-inch f/10 astronomical lens of four elements, designed for optimum performance in red light. It can be seen that the color curve, though normal for an achromat, is very pronounced in the yellow, green and blue, owing to the location of the minimum in red light. This lens was actually built and tested thoroughly. The visual image without filter showed a striking color flare around a red "core", enhanced by the sensitivity of the eye to green light. However, the photographic image with yellow filter showed almost no effect of the color curve on Super-XX film. While color pictures taken with such a lens would no doubt be deficient, a lens of this kind seems entirely adequate for black and white pictures on panchromatic emulsion with yellow or red filter. The tests in the laboratory proved that the sharp core of the image in orange and red light accomplished more than the large flare in yellow and green light detracted. Star images photographed in red light appear fully round and sharp over a 14 by 14 photographic plate.

Many of the systems designed by the author have received similar tests. The astronomical systems are in use and much data are at hand on their performance. However, this report need not be burdened with more information of this kind, information that might better appear in the form of a manual or compendium of results. It is necessary here only to show that the large variety of optical images designed, produced and studied by the writer have built up a certain fairly consistent set of experience that make image interpretation reliable and further design work more precise.

As a general rule, the lower order image errors must be satisfied properly in order to draw on the liberal tolerances of the higher order errors. In this sense, one must expand the image errors in terms of the aperture, rather than of the field. Thus, fifth-order astigmatism must be regarded as a first-order error in the aperture. The aberration is just as important in the outer field as the simple focal error is on axis. At each point in the field, then, the image quality must be

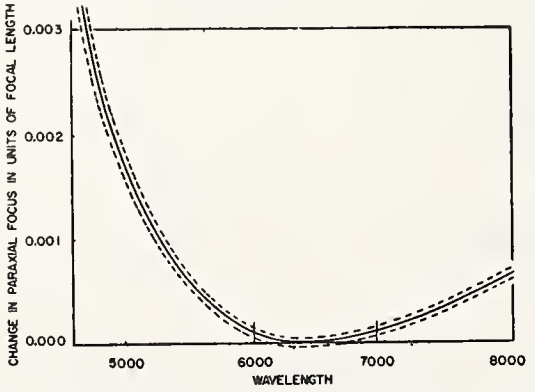


FIGURE 10.10. *Color curve of a 100-inch f/10 Ross lens for red light.*

studied in powers of the aperture. The focal errors such as curvature of field and astigmatism must be satisfied first. Next, the variations of these errors with color must be satisfied, together with lateral color, which is of first-order importance. Next come coma-like errors, varying as the square of the aperture. Thereafter come errors of spherical aberration, easy to control on axis but very difficult to control at far off-axis angles.

Lens-Film Performance

Figure 10.11 is reproduced through the courtesy of the Eastman Kodak Company. The several characteristics of Super-XX and Panatomic-X roll films are shown all on one type of graph. The target contrast is given as the log of the ratio of high light to low light intensity. It will be seen that at a contrast level of 2:1, the resolving power is still as good as 25 lines/mm. Figure 10.11 also shows how the resolving power determined from conventional three-line patterns is a function of exposure. The peak resolving power occurs at a density of about 1.7 for Super-XX and about 1.4 for Panatomic-X for high target contrast. It is significant that the maximum resolving power

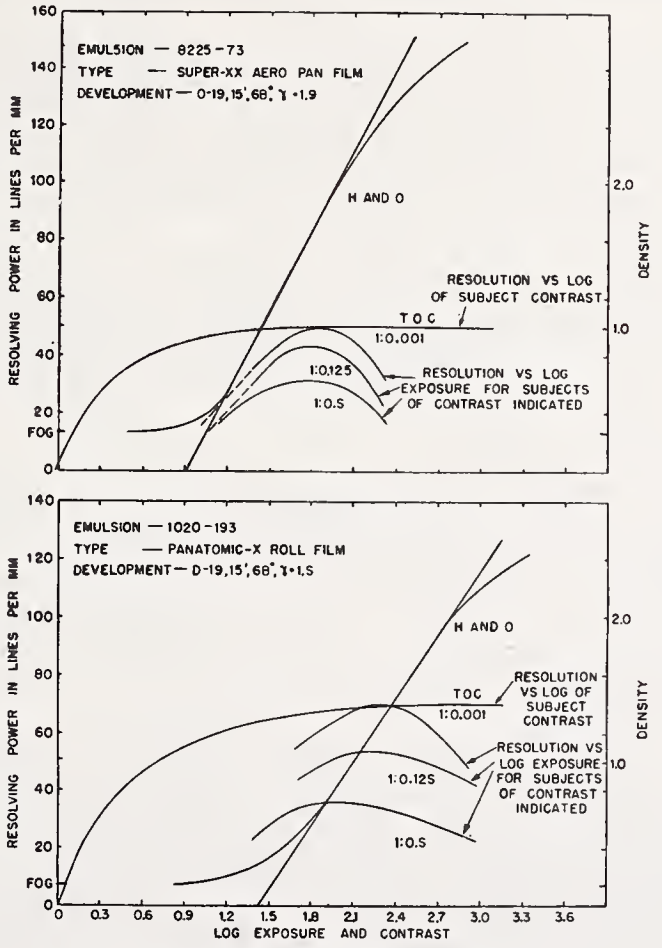


FIGURE 10.11. *Photographic properties of Super-XX and Panatomic-X aerial film.* (Courtesy of the Eastman Kodak Company).

for lower contrast shifts to lower density levels for both emulsions, an effect more pronounced for Pan-X than for Super-XX.

In view of the nature of the curves in figure 10.11, combined with experimental microphotometer tracings made on test patterns, the author was led to propose late in 1945 a type of testing involving the concepts of *macroscopic* and *microscopic* contrasts. The term macroscopic contrast refers to the contrast of large areas on the emulsion. The term microscopic contrast refers to the measured contrast at a given line number in a sequence of decreasing line widths and separations. Microscopic contrast is therefore a function of line number and goes into macroscopic contrast in the limit for large areas. Microscopic contrast goes to zero at the limit of resolution of the photograph, if measured in terms of density difference.

It is evident that the visibility of resolving-power lines depends, to a large extent, on density difference between lines and spaces on the test emulsion. As far as the emulsion is concerned, it does not matter whether the reduction in density difference in the finer lines is caused by aberrations of the lens system, by general fogging, or by turbidity within the emulsion. The final microscopic contrast of a given lens-film combination will depend functionally on the intrinsic characteristics of the emulsion, the target contrast, the lens quality, the exposure, development, and means for measuring the contrast.

For each value of target contrast one can plot a curve of microscopic contrast, as measured in terms of density difference, against line number from macroscopic contrast down to zero, in any given test of lens-film under standardized conditions. The resultant curves are then to be related to performance characteristics that can be considered of direct importance to the ability of the combination to reveal detail and contrast.

Such a family of curves can be plotted for a "perfect lens-film combination", and then for the test lens-film combination, all other conditions being held constant. A good test lens will have the microscopic contrast as close as possible to the ideal limiting curve all the way down the progression of line number. One can interpret from such curves the performance of the lens under varying conditions of haze, and target contrast.

While this procedure was described in 1945, no one has carried it out systematically so far as the author is aware. However, the new resolution-contrast test pattern of the Bureau of Standards accomplishes essentially the same thing, except for differences produced by the use of long lines. The differences may be marked where the aberrations are large. The Bureau of Standards results may indicate higher resolution values over tests conducted with the standard three-line patterns.

If we view the limiting resolution from the standpoint of the emulsion, we can relate the target contrast needed for the imperfect lens-film to attain the stated resolution, to the target contrast needed for the perfect lens-film to attain the same resolution. In this way the loss of picture quality can be expressed in terms of the *microscopic equivalent target contrast.* Thus, at a given resolution limit the inferior lens will require more contrast in the target for resolution than will the perfect lens, and the equivalent target contrast will be a measure of the loss of quality.

TABLE 10.1.

Macroscopic ground target	Super-XX	Pan-X	Ground target contrast	Haze	Microscopic equivalent target contrast
High contrast					
	Lines/mm	*Lines/mm*		%	
Perfect lens	50	70	1/0.001	None	1/0.001
	34	43	1/0.001	50	1/0.33
Imperfect lens	26	32	1/0.001	None	1/0.50
	23	27	1/0.001	50	1/0.60
Low contrast					
Perfect lens	26	32	1/0.50	None	1/0.50
	20	23	1/0.50	50	1/0.67
Imperfect lens	16	17	1/0.50	None	1/0.74
	14	14	1/0.50	50	1/0.80

Table 10.1 gives a tabulation of contrast values worked out from typical results of aerial photography. It is evident that the poor lens gives lowered microscopic contrast, which is the equivalent of added haze for the perfect lens. Increased gamma of development may restore the contrast in the coarser patterns but cannot affect the finer patterns appreciably. Excessive macroscopic contrast produced by prolonged development for the purpose of increasing the microscopic contrast will then distort the tonal values of the photograph.

It is also evident from table 10.1 that the perfect lens has its performance lowered more rapidly by the addition of haze than does the poor lens, though at all times the perfect lens stays systematically superior to the poor lens. Another way of stating the case is to say that the law of diminishing returns sets in when the image quality is perfected beyond a certain point, and that one pays dearly for a slight increase in contrast and resolution. Much aerial photography occurs at low contrast where the superiority of the perfect lens is less clearly defined. All of such troubles are caused essentially by the diffusing nature of the emulsion.

Table 10.1 proves that an increase in focal length is more important than an increase in lens performance, where we can assume a certain reasonable quality to every professional lens system. If haze is so bad that even a perfect lens-film will yield only 16 lines/mm, say, the imperfect lens-film of average performance may still resolve 12 lines/mm. If the focal length of the imperfect lens is increased by 33 percent, the resulting performance referred to the target will be approximately identical to that of the smaller perfect lens, inasmuch as the haze factor is independent of focal length. In aerial photography at a given altitude under such bad haze conditions, either one can use a greater focal length to achieve ground resolution or else fly lower with the same camera, all to achieve a larger scale that in the presence of haze may permit the desired object to be resolved.

It is well known that comparative photographs taken with large and small lenses have about the same macroscopic contrast in the presence of haze, but have ground resolution approximately proportional to focal length. There are many other factors, such as innate lens quality, the mounting, shutter, filter, used, etc. However, if

all these things are held equal, and if the laboratory performance in lines/mm remains independent of focal length as is largely true of modern aerial lenses for the USAF, then the focal length is the most important factor in achieving ground resolution. Desirably, one should have as perfect a lens as prudence and economy of construction can supply, but the focal length is still the most important factor.

Where attention is given on a high technical level to every detail of the aerial lens, photography and flight conditions, one can readily achieve an average of 25 lines/mm in the air with a top quality lens. The imperfect lens may still produce an average of 18 lines/mm, if all other factors are optimized. These figures are obtainable only in the absence of haze and for high target contrasts on the ground.

It has proved possible under excellent conditions to reach a resolving power in the air of 42 lines/mm in a direct test run over ground targets, a figure that the author believes to be the peak so far recorded anywhere. This one test run proves that on excellent days the pure gaseous atmosphere remaining contributes only a slight amount of haze if a red filter is used. However, over much of the world, the atmosphere on the average is very hazy, whereby the differences between good and bad lenses in practice are much obscured.

Focal Problems in Aerial Photography

Possibly the most direct cause of inferior aerial photographs is simply an inadequate focal setting. A "poor" lens may have a greater depth of focus than a "perfect" lens. Therefore, if owing to errors of usage, the lens is out of focus, it can easily happen that the poor lens will return a better picture than the perfect lens. The perfect lens has a higher peak resolution but may have a shorter base to the curve of resolution against focal setting. In the limit with a perfect lens the depth of focus curve can be calculated from known emulsion properties and the nature of the target. The agreement is exact. With an imperfect lens one finds it difficult to calculate the curve of resolution against focal setting, owing to the influence of color aberrations, zones of the lens, etc.

The problem of focusing a lens system depends greatly on the nature of the images produced and on the light source, filter and emulsion. Usually, for an imperfect lens there will be a focal position of best resolution and another of best contrast, though the latter may change with the target contrast. For the perfect lens the optimum focal settings for best resolution and best microscopic contrast coincide. At moderate levels of resolution such as 10 lines/mm, the limitation being produced by vibration, image motion and/or by haze, *focusing for best microscopic contrast* in the laboratory may prove to be the best answer. If vibration and image motion are eliminated, and if haze is moderate, at a level of 20 lines/mm, *focusing for best resolution* probably will yield optimum results. An experienced photographer in the field might consider the situation and set his focus accordingly. It is unlikely that routine observers will have sufficient training, and hence fixed focus cameras are still necessary.

Many workers in aerial photography are not acquainted with the fact that the focal position of a lens is not necessarily stationary. A lens may be said to be *factory-focused*. This statement implies that there exists a focus and that the only problem is to find its position.

Once such is accomplished, the aerial pictures taken thereafter are always supposed to be in focus.

The smaller aerial lenses have such slight changes of focus within the large tolerance permitted by the observed depth of focus curves as to cause little or no focus troubles in practice. Large aerial lenses, however, exhibit a noticeable shift in focus caused by: (1) Temperature changes between equilibrium conditions; (2) thermal gradients or transients; (3) air density effect; (4) ground distance, according to vertical or oblique; (5) simple flexure of the camera and its component parts; (6) change in focal setting with change of filter and/or emulsion; (7) change in focus caused by stopping down, if lens imperfect. We are concerned here with only the first four causes.

Figure 10.12 shows the results of observations made in a cold-chamber test of a 40-inch lens. The total range of focal changes amounts to nearly 1 mm, owing to thermal gradients and to changes between equilibrium temperatures. The depth of focus at the level of 40 lines/mm is of the order of 0.15 mm. A focal error of the order of 0.5 mm can already cause a loss of resolution to perhaps 20 lines/mm or so.

Figure 10.13 shows the combined effects of ground distance for vertical photographs and focal changes caused by loss of air from the lens at the various altitudes. Here again appreciable errors are

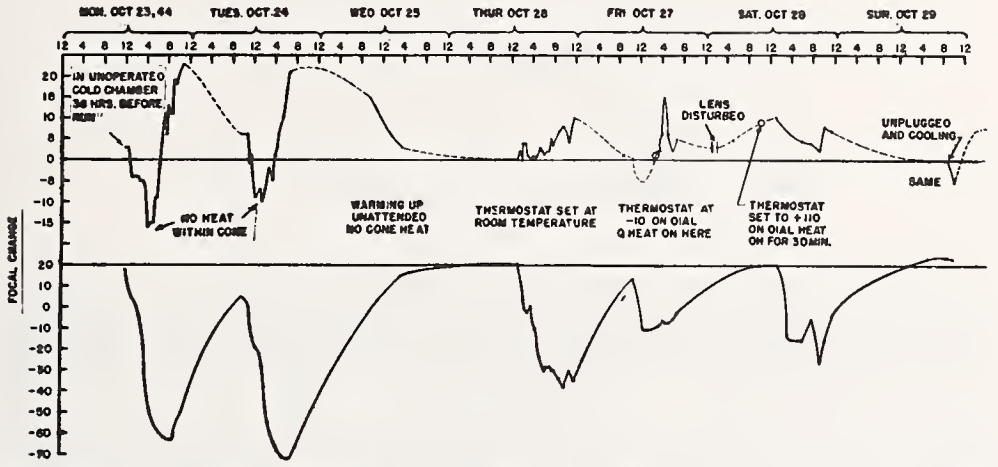


FIGURE 10.12. *Focal changes of a 40-inch f/5 telephoto lens as a function of temperature and temperature gradients.*

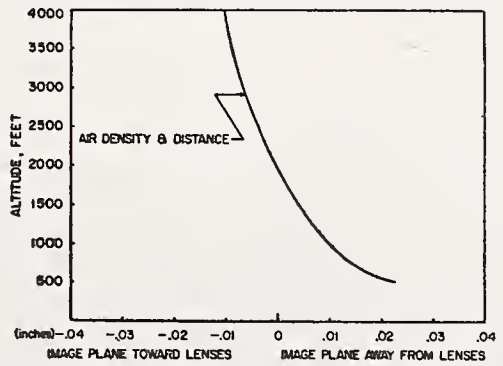


FIGURE 10.13. *The air-density effect and ground-distance image shift for a 40-inch f/5 telephoto lens.*

encountered for both effects. At very low altitudes, of course, the ground distance effect becomes very large.

Table 10.2 contains calculated shifts for the case of a 24-inch and of a 144-inch focal length. It is clear that the focal errors are so large as to represent a serious problem to the aerial photographer. Thermal-gradient changes are even more serious than equilibrium changes, as shown by figure 10.12.

TABLE 10.2. *Change in equilibrium temperature*

A change of −80° C causes the image in the case of a 24-inch focal length to move 0.90 mm beyond the film plane away from the lens. The corresponding shift in the case of a 144-inch focal length is 6.30 mm.

Air-density effect		
Altitude	Focal length	
	24-inch image shift	144-inch image shift
ft	*mm*	*mm*
Infinity	−0.65	−3.44
40,000	− .50	−2.60
20,000	− .32	−1.71
10,000	− .19	−1.00
5,000	− .11	−0.54
2,000	− .05	− .24
Ground	.00	.00
Ground-distance effect		
Infinity	0.00	0.00
40,000	.03	1.10
20,000	.06	2.20
10,000	.12	4.39
5,000	.24	8.78
2.000	.61	21.95

Ideally, large lenses ought to be focused in the air immediately before a picture run is made. Short of this, the next best procedure is to prepare tables of focal changes from laboratory cold- and pressure-chamber observations. There is little justification in one's guessing at the focal setting for a large lens unless such tables have been prepared. A trained observer should also take thermal gradients into account. If the focal problem is carelessly handled, pictures as poor as 3 lines/mm might result and one may as well make use of a lens of shorter focal length.

Several of the lenses discussed above have been tested thoroughly at the Eastman Kodak Co. through the work of L. A. Jones, R. N. Wolfe, and associates. A very few of the voluminous and careful test results are reproduced here by permission of the Research Laboratories, for which acknowledgment is hereby made.

Figure 10.14 and figure 10.15 show the results of resolving power against focal setting for a 6-inch f/3.5 spherically-concentric lens. Because of the spherical symmetry no single optical axis exists. The very slight deterioration of the tangential resolving power in figure 10.14 is brought about by the fore-shortening of the aperture in the oblique beam with the consequent loss of diffraction resolving power. This lens was corrected for use with a red filter, and the spherical correction at f/3.5 was just slightly beyond one Rayleigh limit. The curves at high contrast show peak resolutions as much as 70 lines/mm

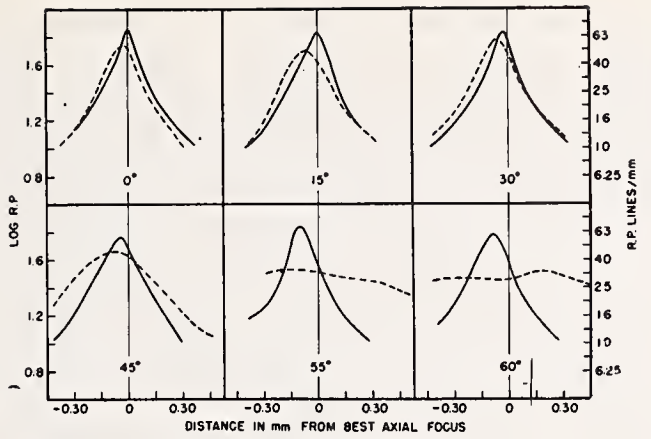


FIGURE 10.14. *Resolving power versus focal setting for a 6-inch f/3 wide-angle lens with red filter, Super-XX Aero Pan, and high-contrast 3-line test object.*

---, radial; —, tangential.

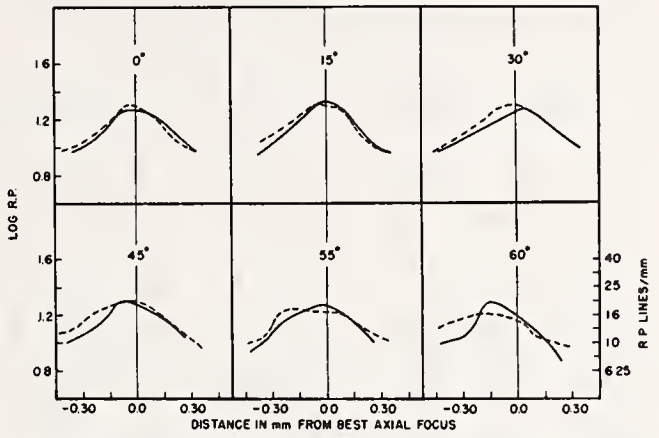


FIGURE 10.15. *Resolving power versus focal setting for a 6-inch f/3 wide-angle lens with red filter, Super-XX Aero Pan, and low-contrast Cobb test object.*

---, radial; —, tangential.

on Super-XX. The apparent large depth of focus is brought about by the broadening of the bright lines of the test pattern at the expense of the dark spaces between the lines. That is, at a position 0.3 mm outside of focus, the dark spaces were only 0.020 mm wide on the test negative but still one could see three separate lines. This effect increases the depth of focus of the three-line pattern over and beyond what one would normally expect from the size of an out-of-focus star image.

Figure 10.15 shows the peak resolution to be expected with the Cobb 2-line test chart at low contrast (log contrast equals 0.17). This f/3.5 lens is nearly as good a lens as might be required to test the film itself. Hence, the loss of resolution down to about 20 lines/mm is an emulsion property. It is clear again that good image correction can be swallowed up at low contrast by the turbidity of the emulsion.

Testing at low contrast provides only a compressed scale for distinguishing between good and poor lenses. Table 10.1 is typical of what happens. Figure 10.15 as compared to figure 10.14 points it out again. If testing at low contrast is to be accomplished without additional testing at high contrast, the observer must content himself

with considering that no lens is very good and that every *designed* lens is fair.

The author believes that the high peak resolution observed may be brought about by the f/3.5 speed rating of the spherical lens, along with a near absence of spherical aberration and color. The 40-inch f/5 telephoto lens in the controlled models at least was just as well corrected for spherical aberration but in red light gave somewhat lower peak resolutions. The probable slight dependency on f/number ought to be kept in mind as more data become available. The dependence is shown to some extent by the many Eastman tests.

Figure 10.16 shows wedge photographs made with the 40-inch f/5 telephoto. The figure is more or less self-explanatory. The ordinates are logarithmic and hence the length of the sharp peak indicates that most of the light is where it belongs. For comparison purposes figure 10.17 shows similar wedge patterns for the standard 24-inch f/6 aerial lens, which is of Tessar construction, and about at the limit of what a Tessar can be expected to accomplish. The sharp peaks of figure 10.16 are intended to produce high resolving powers with good microscopic contrast at the level of 30 lines/mm. The shallow

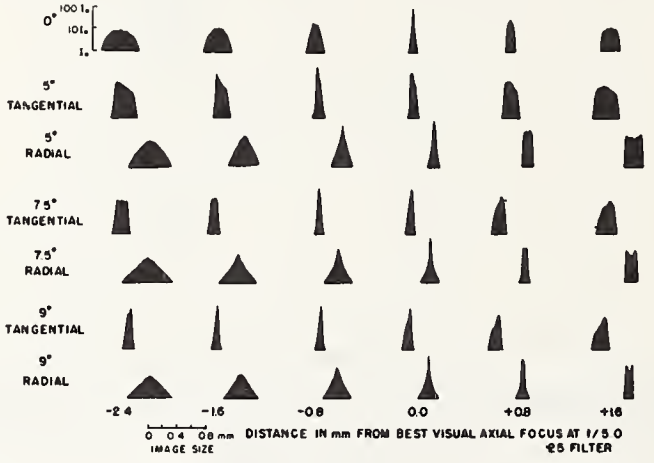


FIGURE 10.16. *Wedge tests on images from a 40-inch f/5 telephoto lens.*

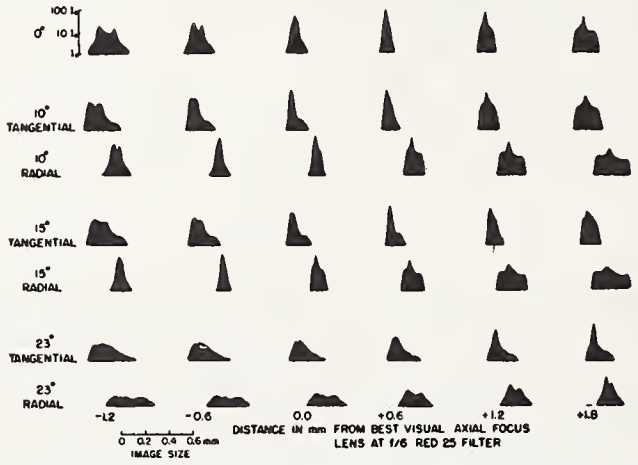


FIGURE 10.17. *Wedge tests on images from a 24-inch f/6 Tessar lens.*

peaks of figure 10.17 in cases of heavy exposure produce "muddy" photographs so far as the microscopic contrast is concerned, at levels around 20 lines/mm. Figures 10.18 and 10.19 show the resolution curves for this same f/6 standard Tessar for high and low contrast.

Figure 10.20 shows the wedge photographs for a 100-inch astronomical Ross lens. The peaks are sharp, even though there is very

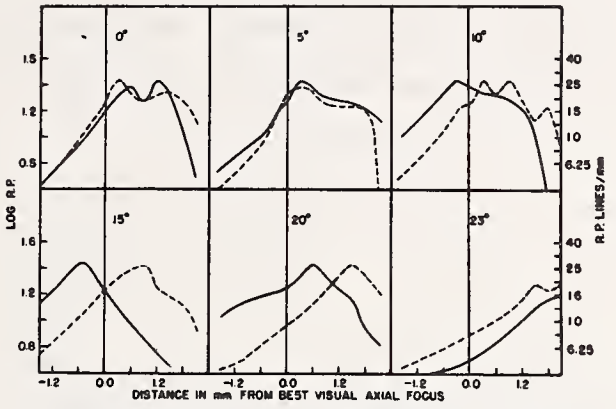


FIGURE 10.18. *Resolving power versus focal setting for a 24-inch f/6 Tessar lens, with Super-XX Aero Pan, tungsten, no. 12 filter, and high-contrast 3-line test object.*

---, radial; —, tangential.

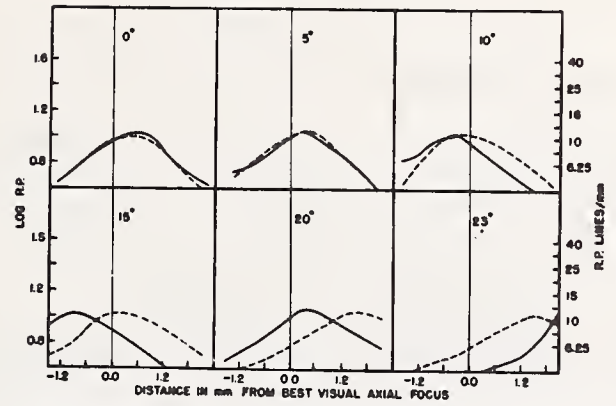


FIGURE 10.19. *Resolving power versus focal setting for a 24-inch f/6 Tessar lens, with Super-XX Aero-Pan, tungsten, no. 12 filter, and low-contrast Cobb test object.*

---, radial; —, tangential.

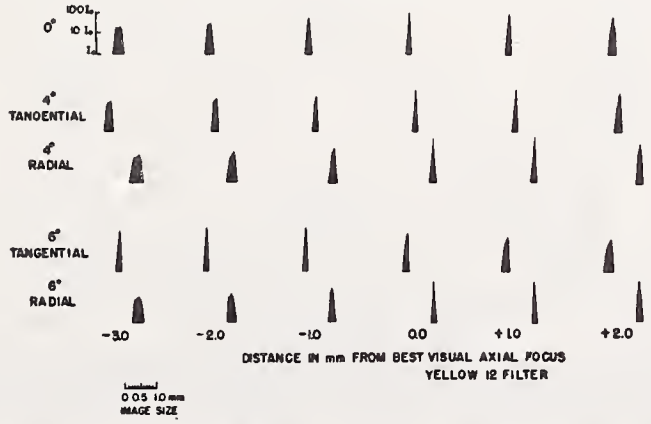


FIGURE 10.20. *Wedge tests on a 100-inch f/10 Ross lens.*

considerable secondary spectrum. Figures 10.20 and 10.22 show the observed resolving powers at high and low contrasts. It should be noted that the optical glass of this lens was deficient. The observed axial visual image in red light showed no "clean" edge, but rather a hazy patch of small size. The loss of contrast in the edge causes a loss of microscopic contrast for the finer lines. This loss is barely discernible in the wedge photographs of figure 10.21, where the approach to the peak is slightly broader than would have been the case if the lens had been fully corrected. Another lens of the same design made since shows resolving powers as high as 55 lines/mm at high contrast.

Figures 10.23 and 10.24 show the resolving power results for the

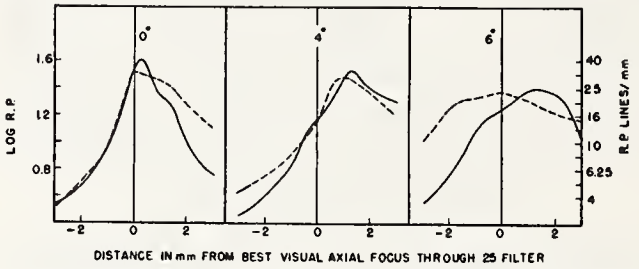


FIGURE 10.21. *Resolving power versus focal setting for a 100-inch f/10 Ross lens with Super-XX Aero Pan, tungsten, no. 12 filter, and high-contrast 3-line test object.*

---, radial; —, tangential.

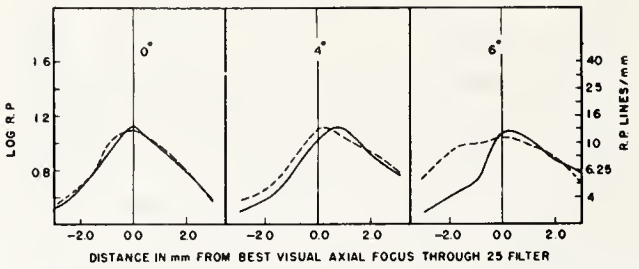


FIGURE 10.22. *Resolving power versus focal setting for a 100-inch f/10 Ross lens with Super-XX Aero Pan, tungsten, no. 12 filter, and low-contrast Cobb test object.*

---, radial; —, tangential.

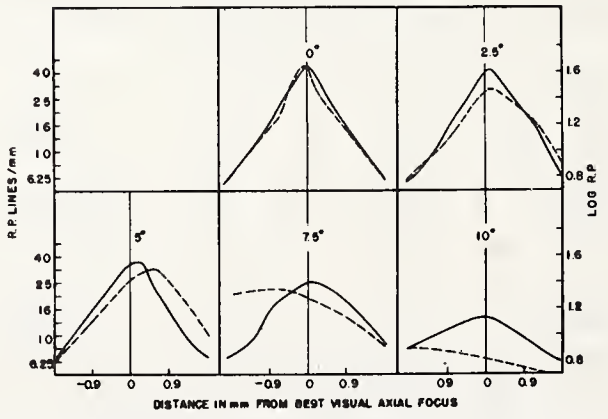


FIGURE 10.23. *Resolving power versus focal setting for a 36-inch f/8 apochromat with Super-XX Aero Pan, "daylight", no. 12 filter, and high-contrast 3-line test object.*

----, radial; ——, tangential.

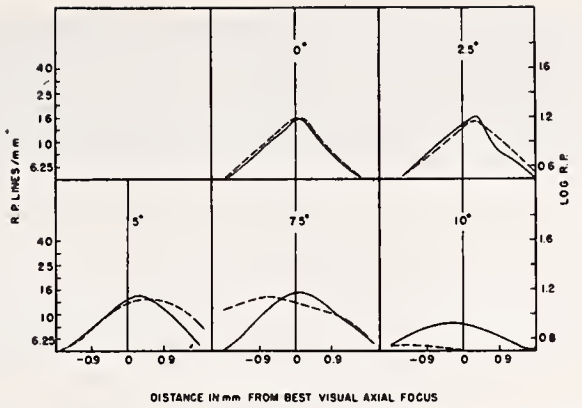


FIGURE 10.24. *Resolving power versus focal setting for a 36-inch f/8 apochromat with Super-XX Aero Pan, "daylight" no. 12 filter, and low-contrast Cobb test object.*

Object distance=28.8 focal length; — — — —, radial; ——, tangential.

36-inch f/8 apochromat discussed above. The effect of the apochromatism is not evident in the resolving powers as such but would show up more in the level of microscopic contrast around 30 lines/mm. Such lenses as the 36-inch apochromat would take aerial pictures of good microscopic contrast around the average expected resolving power in the air of 25 lines/mm., where other factors have been minimized.

The Calculated Photographic Image

With the coming use of electronic calculating equipment in optical design, consideration must be given as to whether the photographic image can be calculated. The cost of a large lens is so considerable as to make it desirable to go as far as possible on paper during the design stages. To a considerable extent the views described above have been drawn on repeatedly by the author in designing photographic systems and in predicting by judgment alone what might resonably be expected of the lens in photographic test. However, if it should prove possible to calculate the photographic image with accuracy, a step in quantitative analysis will have been achieved.

The author with the valued assistance of W. Randolph Angell, Jr. has made a preliminary attempt to calculate the photographic image in a special case. A simple plano-convex lens of barium crown glass with its normal complement of primary color was set up in the laboratory. Infrared photographic resolving power focusing runs were made at a 10-fold reduction between collimator and lens. A minus-blue filter was used, but otherwise the standard aerial infrared film was exposed to the spectral colors from a 3,000° tungsten source. Figure 10.25 reproduces a portion of the photographs taken. The second pattern seems resolved and corresponds to a resolution of 16.6 lines/mm When one considers that the primary spectrum from 5000 to 9000 angstroms is altogether uncorrected, that the lens has some spherical aberration, and otherwise is f/4, the observing of 16.6 lines/mm indicates to some extent how inadequate a lens can be and still give fair photographs. Of course, the microscopic contrast is low and even at a level of 5 lines/mm the photograph resulting would appear "muddy"

FIGURE 10.25. *Enlarged view of test images produced by a simple f/4 lens of 10-inch focal length.*

Pattern 2 is 16.6 lines/mm.

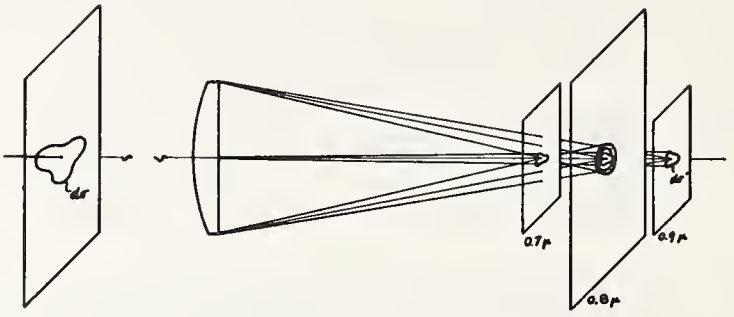


FIGURE 10.26. *Schematic view of the problem of calculating the photographic image in the presence of large chromatic aberration.*

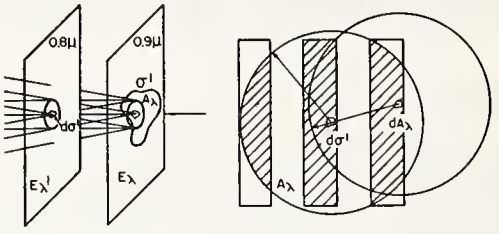


FIGURE 10.27. *Enlarged view of the image region in the problem of calculating the photographic image in the presence of large chromatic aberration.*

and unsatisfactory. Some of the out-of-focus colors have blur circles considerably larger than the entire test pattern.

Figure 10.26 shows the initial circumstances of the attempt to calculate the image. A mean focal plane was chosen to lie at 0.8 micron, though the results of figure 10.25 are not necessarily found optimum at 0.8. A small object element of area providing the illumination images into a similar area, $d\sigma'$. The illumination on the mean focal plane at 0.8 micron consists of the in-focus 0.8-micron light plus the inside and outside focus neighboring colors.

The illumination of any element $d\sigma'$ in its own focal plane is given by

$$E_\lambda d\lambda d\sigma' = k\pi \sin^2 \theta' B_\lambda d\lambda d\sigma,$$

which is the standard formula. We need it here only to show that the distribution of the illumination E_λ with wavelength is proportional to the similar emission from the object. The *dilution factor* k need not be evaluated, and we can work directly with E_λ, instead of B_λ.

Figure 10.27 shows an enlarged view of the space around the image. element. E'_λ (not a derivative) is taken as the illumination function on the chosen focal plane at 0.8 micron, resulting from the illumination E_λ on the displaced focal plane at another wavelength, λ. If the ele-

ment of area $d\sigma'$ is taken to be extremely small compared to the color aberration, then the illumination in the focal plane at 0.8 micron will be

$$E'_\lambda d\lambda d\sigma' = \iint_{A_\lambda} \frac{E_\lambda d\lambda d\sigma'}{(A_\lambda)_{\max.}} \, dA_\lambda,$$

where $(A_\lambda)_{\max.}$ is the total area of the out-of-focus blur circle striking the 0.8-micron focal plane, and where the integration is taken over the entire illuminated area of the test pattern, whatever its shape, within the radius of the blur circle of area $(A_\lambda)_{\max.}$.

This integration can be performed at each wavelength, and in fact

$$E'_\lambda d\lambda d\sigma' = \left[\frac{\iint_{A_\lambda} dA_\lambda}{(A_\lambda)_{\max.}} \right] E_\lambda d\lambda d\sigma'.$$

If we multiply by the dependence of filter transmission on wavelength, F_λ, and by the sensitivity of the emulsion relative to wavelength, S_λ, we have

$$F_\lambda S_\lambda E'_\lambda d\lambda d\sigma' = \left[\frac{\iint_{A_\lambda} dA_\lambda}{(A_\lambda)_{\max.}} \right] (F_\lambda S_\lambda E_\lambda) \, d\lambda d\sigma'.$$

To convert to the characteristic curve, we have simply

$$\log \mathscr{E} = \log \int_0^\infty \left[\frac{\iint_{A_\lambda} dA_\lambda}{(A_\lambda)_{\max.}} \right] (F_\lambda S_\lambda E_\lambda) \, d\lambda + \log \kappa,$$

where κ is a constant. Where the blur circle is completely illuminated for a macroscopic area, σ', then the photographic density can be set at any desired value by an adjustment to the abscissa scale. In particular, for infrared aerial film we can set this maximum density that would be obtained by photographing a uniformly illuminated surface at 1.6 on the standard characteristic curve for this emulsion. For any other pattern to be calculated, we have only to evaluate

$$\log \int_0^\infty \left[\frac{\iint_{A_\lambda} dA_\lambda}{(A_\lambda)_{\max.}} \right] (F_\lambda S_\lambda E_\lambda) \, d\lambda.$$

This procedure has been carried out for the case actually photographed, as in figure 10.25, but only for the 10 lines/mm resolution level. Figure 10.28 shows the variation of focal position with color, which represents a substantial change. Figure 10.29 shows a graph of B_λ, which we can take directly as being equivalent for our purposes to E_λ. Figure 10.30 is the spectral sensitivity curve obtained from

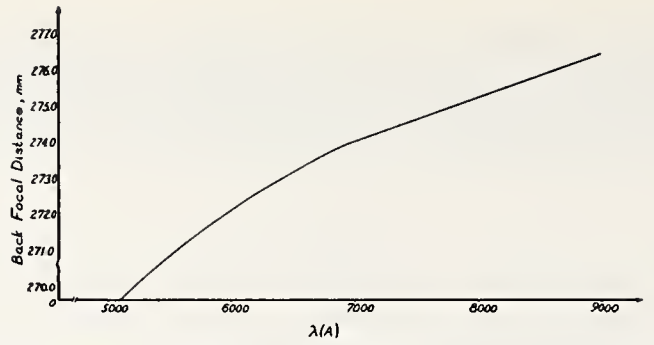


FIGURE 10.28. *Variation of back focal distance with color.*

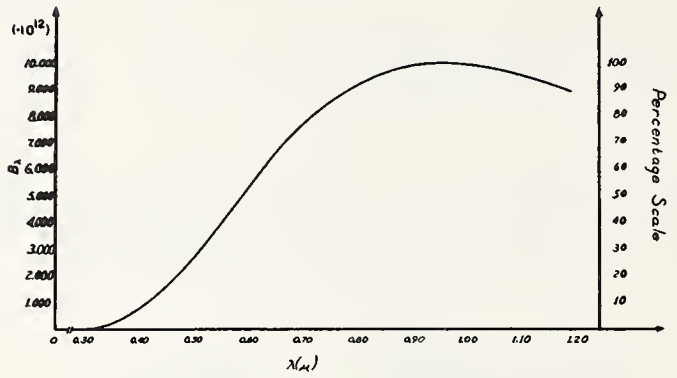


FIGURE 10.29. *Black-body radiation curve for 3,000° K.*

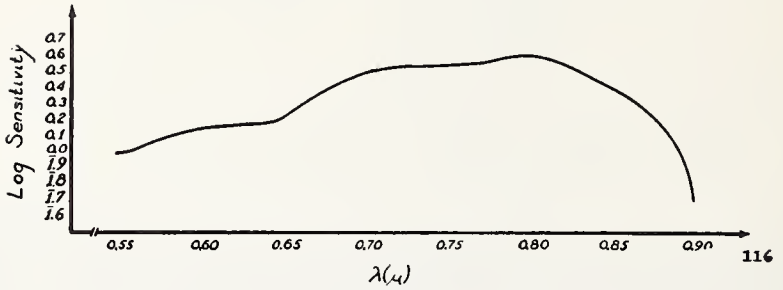


FIGURE 10.30. *Spectral sensitivity curve for I–N emulsion.*

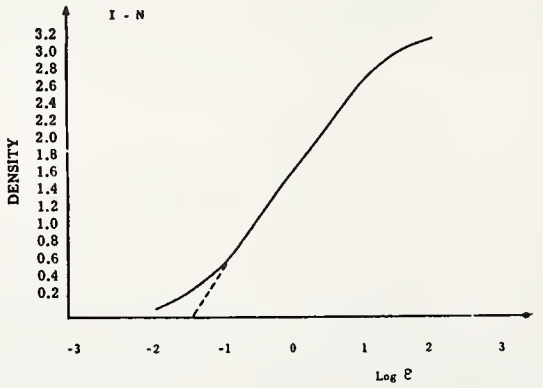


FIGURE 10.31. *Characteristic curve adopted for I–N emulsion.*

the Kodak Handbook of Photographic Plates. Figure 10.31 is the characteristic curve adopted for type I–N infrared emuslion. Figure 10.32 is the product of the filter function F_λ (minus-blue), S_λ, and E_λ. The total area under the curve represents the maximum exposure. Integration was performed by planimeter, in convenient units. The logarithm of the area was then set at density 1.6 on the characteristic curve.

Figure 10.33 represents typical curves for the indicated points on the pattern. The solid curve represents the illumination at point A on the inserted diagram, which is within a bright line. The peak illumination comes then from the in-focus 0.8-micron light and neighboring colors. Similarly, at point B where there would be no illumination if the lens were good, one can see from the dashed curve that at 0.8-micron there is indeed no illumination. Other outlying colors do contribute to the surface brightness, however, mostly from colors not far distant.

It is to be noted that the blur circles include all three lines of the pattern. The integration is performed by moving the chief ray or point of the planimeter over the outline of the bright areas of the best pattern within the requisite distance of $d\sigma'$. The color aberrations of this particular lens are so large as to have the entire out-of-focus pattern contributing to the photographic image at each point.

Figure 10.34 is the final result, where the calculated densities are tabulated on an enlarged diagram of the three-line pattern. The largest calculated density is 1.20, whereas the adopted density for macroscopic areas is 1.6. However, a density of 1.20 still represents

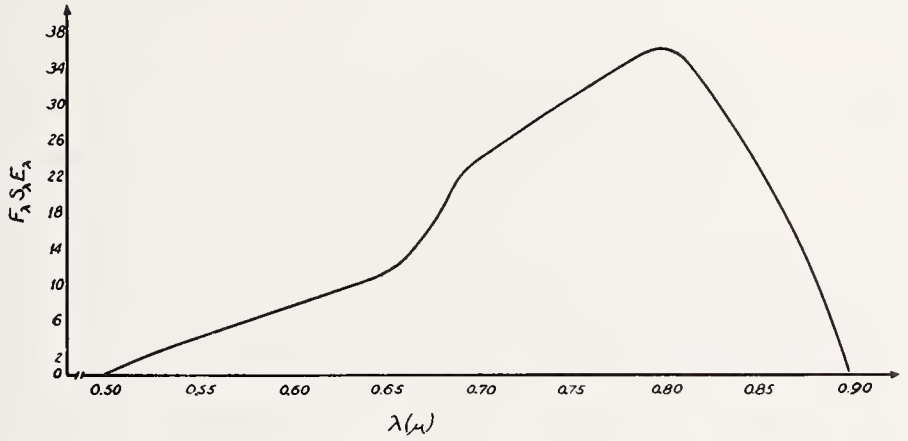


FIGURE 10.32. *Maximum effective exposure for filter, emulsion and light source.*

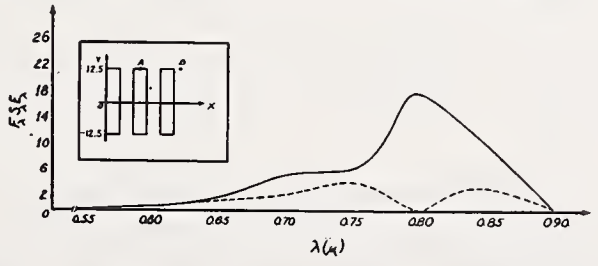


FIGURE 10.33. *Sample curves of calculated illumination versus wavelength.*

Back focal distance=275.27 mm at $\lambda=0.8\,\mu$. —, exposure A at $X_0=12.5$ ul., $Y_0=12.5$ ul.; - - -, exposure B at $X_0=27.5$ ul., $Y_0=12.5$ ul. 1 ul=1 unit length=10 microns Lines and spaces of resolving-power pattern are each 50 microns wide.

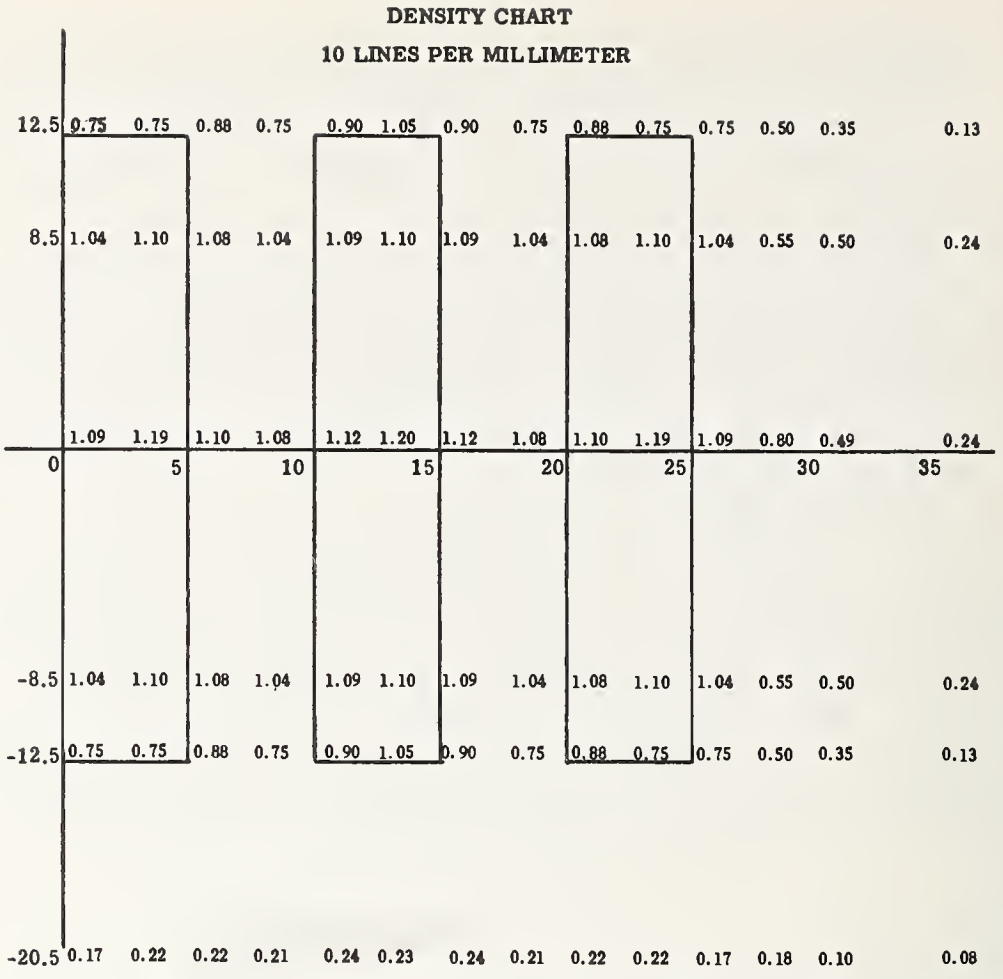


FIGURE 10.34. *Calculated photographic image of a three-line resolution pattern at 10 lines/mm level for test case of unachromatized simple lens.*

good blackness of the image, and indeed this density can be changed by exposure time and development.

The density differences between the spaces and the centers of the bright lines seem small, but experience with measured targets indicates that the eye would call this pattern well resolved. The microscopic contrast would be low, but the three lines could easily be seen.

It is difficult to determine whether the calculated and observed patterns agree. We plan to repeat the experiment with a photomicrographic enlargement as well, in order to have some control on the photographic factors. Also, in redoing the work, we plan to determine the actual best-focus position of the emulsion against millimeters of back focal length, as calibrated from the sodium image. Also, the lens will be stopped down enough to minimize the added effects of spherical aberration.

An Objective Method of Testing

During the summer of 1944 in the wartime laboratory at Harvard, the author initiated a form of testing designed to eliminate the personal equation from evaluation of lens performance. Special cardboard charts were printed for the purpose and copied to target size

FIGURE 10.35. *A view of the large testing tunnel.*

at varying contrasts onto lantern slide plates. These charts contained lines of thoroughly scrambled block letters, the letters in each line being smaller than in the preceding line in a geometrical progression. Test negatives from the tunnel were read off by an uninformed observer against a check list. The lens performance was scored according to the line where the observer made fewer than 50 percent errors. This particular testing method was used for awhile in 1944 and then set aside for other more urgent activities.

In recent months the author has taken up this type of test procedure once again, the purpose being to determine what image properties of a lens facilitate recognition of various types of objects at varying contrasts. Time has not permitted more than a preliminary series of observations, but enough can be presented here to indicate the nature of the work.

Figure 10.35 shows a view of the testing tunnel in the author's laboratory. The tunnel is 28 feet long, 4 feet square, and so constructed that the sides of the tunnel may be fastened to the ceiling or lowered at will. A 16-inch paraboloidal mirror is set up as a collimator at the far end of the tunnel. This mirror delivers the collimated rays from a target about in the center of the tunnel to the test lens at the other end. Ordinary focal-plane test photographs or photomicrographic enlargements can be made for any lens up to 100-inches focal length or so.

In order to avoid small photographic targets with uncertain properties for the smallest resolution lines, the writer has prepared to use a *live* target, shown in figure 10.36. The target cabinet is made of oak and houses 52 light bulbs of *daylight* type rated at 60 watts. There is an intermediate diffusing screen and then an interchangeable target panel. The present target measures approximately 24 by 24

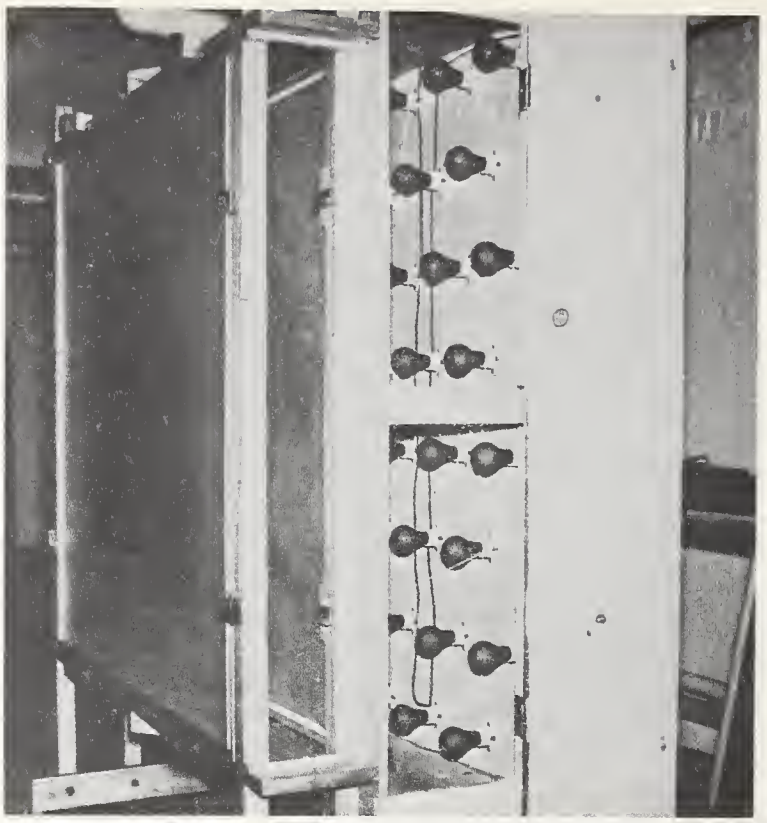


FIGURE 10.36. *A view of the test target and cabinet.*

inches and is handpainted onto a 48- by 48-inch ground glass. The unpainted surface serves as a macroscopic photometric standard. The surface brightness corresponds to that of the ground on a cloudy-bright day.

In testing lenses of short focal length one can photograph the target directly from distances up to 50 feet in the laboratory itself. For testing lenses of greater focal length, one reimages the target by means of a suitable reducing lens of high quality onto the focal plane of the collimator mirror. The resulting small image is collimated by the mirror and the parallel rays sent to the test lens. In this way one can be sure that the surface brightness of even the smallest target will remain the same as for the macroscopic photometric areas.

Varying contrasts are introduced into the collimated beam by means of a beam-splitter placed over the reducing lens near the focus of the mirror. The lowered contrast will then be brought about by a superposition of uniform illumination onto the whole target area including the photometric area and test patterns. Temporarily, the writer has had to approximate the equivalent by adding fogging exposures to the target exposures and for this paper has made use of a 3-inch f/3.5 Tessar test lens.

Because the target panels are interchangeable, one can copy large-scale transparencies as *live* objects down to the smallest size used for test purposes. Thus, different surface objects of interest, such as roads, houses, streams, railroads, etc., can be copied at varying scale with a given lens and the range of its effectiveness for each kind of target determined. Variable haze and color of haze can easily be introduced.

Figure 10.37 shows the target used in the present preliminary investigation. The picket-fence pattern is intended for microdensitometry. The standard three-line patterns decrease in size according to the cube root of 2, although admittedly, the sixth root of 2 is more desirable. Similarly, the letters are selected for uniqueness from the alphabet and in each line have been scrambled to confuse the observer. The block letters are as large as the three-line patterns, and the line width of the letter in every case is equal to the line width of the corresponding three-line pattern. There are 10 *shape* objects selected for distinctness but basic to the problem of *recognition*. Also, there are 10 *figure* objects, selected for recognition tests of orientation and shape.

Every test line and shape in the entire target has clean-cut edges and sharp intersections, produced by ruling pen and hand correction. Thus, every edge and corner delivers a pure *square wave* to the test lens via the collimator. There is no intermediate photographic process to cause a crumbling away of edges and corners of the smaller patterns, and the minifying lens used at the focus of the collimator at f/15 gives results of microscopic quality. The only real deficiency in this *live target* set-up arises from small air turbulence within the test tunnel. Experience to date indicates that the air in the tunnel limits performance to approximately 1 second of arc.

Five test films have been taken. The results of the fifth run are given in figures 10.38, 10.39, 10.40, and 10.41. The author reduced

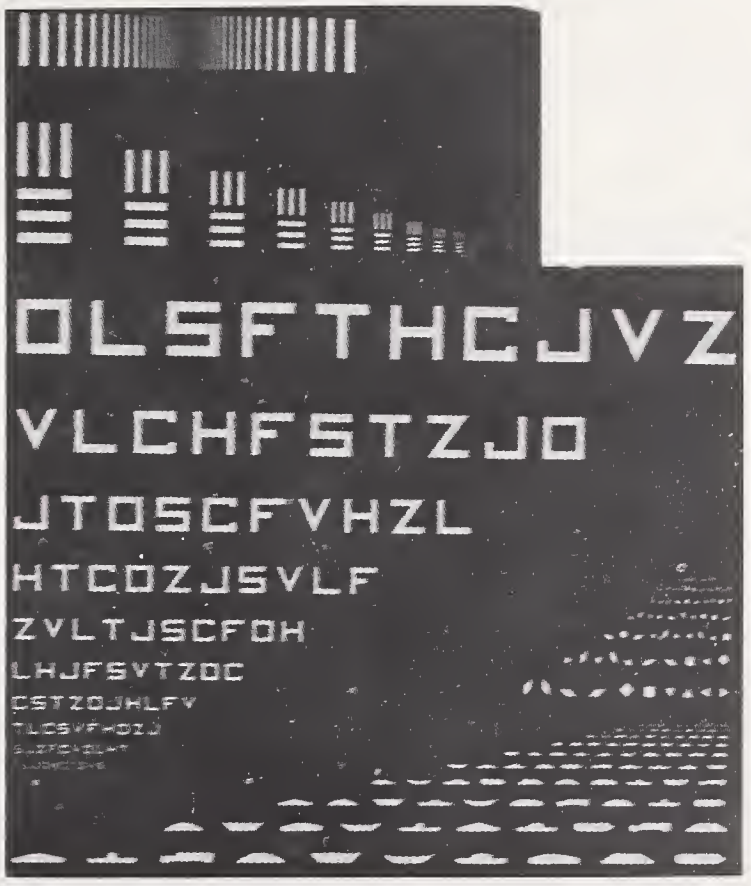


FIGURE 10.37. *The test target for letters, shapes and figures as used at present.*

the negatives personally, making use of a 60X microscope and calling off each shape, letter, or figure to a recorder who scored against a key. The order of reading was altered frequently, nor did the author examine the key prior to the microscopic examination of the test negatives. Nevertheless, it was apparent that slight effects of memory do interfere with the complete objectivity of the test, and arise in the unavoidable process of elimination.

In figures 10.38 through 10.41 the first column gives the resolving power at the image plane of the test lens, corresponding to the pattern number given in the second column. However, these resolving-power figures are not the observed values but are the progression used. Row 1 is the finest pattern, and also represents the finest patterns among the shapes, figures and letters. The scale, then, runs from 5 to 40 lines/mm. The check marks mean that the particular test object was read correctly against the key.

Roll 5 - SUPER-XX, 18 min. Microdol$_3$, 70°C, tray.

R.P.	Row	Shapes	Figures	Letters	#
40	1			x x x x x x x x x x	5-1
32	2			x x x x x x x x x x	
25	3	x x x		x x x x x x x x x x	f/11
20	4	x x	x x x x	x x x x x x x x x x	
16	5	x x x x x	x x x x x x x	x x x x x x x x x x	1/500
13	6	x x x x x x x	x x x x x x x	x x x x x x x x x x	40/32
10	7	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	20′
8	8	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	0.45 / 0.06
6	9	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	
5	10	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	
40	1			x x x x x x x x x x	5-2
32	2			x x x x x x x x x x	
25	3	x x x x	x	x x x x x x x x x x	f/11
20	4	x x x x x	x x x x	x x x x x x x x x x	
16	5	x x x x x x x x x x	x x x x x x x x	x x x x x x x x x x	1/250
13	6	x x x x x x x x x x	x x x x x x x x x	x x x x x x x x x x	40/32
10	7	x x x x x x x x x x	x x x x x x x x	x x x x x x x x x x	20′
8	8	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	0.53 / 0.07
6	9	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	
5	10	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	
40	1			x x x x x x x x x x	5-3
32	2			x x x x x x x x x x	
25	3	x x x		x x x x x x x x x x	f/11
20	4	x x x x x x x x	x x x	x x x x x x x x x x	
16	5	x x x x x x x x x	x x x x x x x	x x x x x x x x x x	1/125
13	6	x x x x x x x x x x	x x x x x x x	x x x x x x x x x x	40/32
10	7	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	20′
8	8	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	0.51 / 0.06
6	9	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	
5	10	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	
40	1			x x x x x x x x x x	5-4
32	2			x x x x x x x x x x	
25	3			x x x x x x x x x x	f/11
20	4	x x x x		x x x x x x x x x x	
16	5	x x x x x x x	x x x x x x x	x x x x x x x x x x	1/61
13	6	x x x x x x x x x x	x x x x x x x	x x x x x x x x x x	40/25
10	7	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	20′
8	8	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	0.80 / 0.07
6	9	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	
5	10	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	

FIGURE 10.38. *A tabulation of individual test results, 5–1 to 5–4.*

Roll 5 - Super-XX, 18 Min., Microdol$_3$, 70°C, tray.

R.P.	Row	Shapes	Figures	Letters	#
40	1			x x x x x x x x x x	5 - 5
32	2			x x x x x x x x x x	
25	3	x x		x x x x x x x x x x	f/11
20	4	x x x x x x x x x	x x x x	x x x x x x x x x x	
16	5	x x x x x x x x x	x x x x x x x x x	x x x x x x x x x x	1/31
13	6	x x x x x x x x x x	x x x x x x x x x	x x x x x x x x x x	40/25
10	7	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	20′
8	8	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	
6	9	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	1.12 / 0.08
5	10	x x x x x x x x x x	x x x x x x x x x x	x x x x x x x x x x	
40	1			x x x x x x x x x x	5 - 6
32	2			x x x x x x x x x x	
25	3	x x x	x x x	x x x x x x x x x x	f/11
20	4	x x x x x x x	x x x x x x x	x x x x x x x x x x	
16	5	x x x x x x x x x	x x x x x x x x	x x x x x x x x x x	1/16
13	6	x x x x x x x x	x x x x x x x x	x x x x x x x x x x	32/32
10	7	x x x x x x x x x	x x x x x x x x x	x x x x x x x x x x	20′
8	8	x x x x x x x x x x	x x x x x x x x x	x x x x x x x x x x	
6	9	x x x x x x x x x x	x x x x x x x x x	x x x x x x x x x x	1.46 / 0.33
5	10	x x x x x x x x x x	x x x x x x x x x	x x x x x x x x x x	
40	1			x x	5 - 7
32	2			x x x x x x x x x x	
25	3			x x x x x x x x x x	f/11
20	4			x x x x x x x x x x	
16	5			x x x x x x x x x x	1/16
13	6	x x x x x x		x x x x x x x x x x	25/25
10	7	x x x x x x x x x	x x x x	x x x x x x x x x x	20′
8	8	x x x x x x x x x	x x x x x x	x x x x x x x x x x	
6	9	x x x x x x x x x	x x x x x x x x x	x x x x x x x x x x	1.57 / 1.42
5	10	x x x x x x x x x	x x x x x x x x x	x x x x x x x x x x	
40	1			x x	5 - 8
32	2			x x x x x x x x x x	
25	3			x x x x x x x x x x	f/11
20	4			x x x x x x x x x x	
16	5			x x x x x x x x x x	1/31
13	6	x		x x x x x x x x x x	25/20
10	7	x x x x x x	x x x x x	x x x x x x x x x x	20′
8	8	x x x x x x x x x x	x x x x x x	x x x x x x x x x x	
6	9	x x x x x x x x x x	x x x x x x x x x	x x x x x x x x x x	1.17 / 1.00
5	10	x x x x x x x x x x	x x x x x x x x x	x x x x x x x x x x	

FIGURE 10.39. *A tabulation of individual test results, 5–5 to 5–8.*

In the last column on the right, we have from top to bottom in each group the film number and picture number. The f/number is then given, and next the exposure time as marked on the lens. It is obvious that these exposures are not correct, because there is no recognizable difference in density between 1/500th and 1/125th. However, the first three observations can be compared for they should be equivalent. 40/32 means that the lens resolved 40 lines in the horizontal line pattern, and 32 lines in the vertical pattern. The pictures indicated that the lens possesses either a slight decentration or that the shutter introduced vibration, even though the camera was clamped to a concrete pier. The next figure is the distance to the target in feet, focused by rangefinder. The figure 0.45/0.06 represents the observed densities of the high light and low light as measured by visual densitometer, and therefore represents the contrast in terms of density differences.

Roll 5 - Super-XX, 18 Min., Microdol$_3$, 70°C, tray.

R.P.	Row	Shapes	Figures	Letters	#
40	1			x x x x x x	5 - 9
32	2			x x x x x x x x	
25	3			x x x x x x x x x	f/11
20	4			x x x x x x x x x	
16	5			x x x x x x x x x	1/61
13	6	x x x x	x x x x	x x x x x x x x	32/16
10	7	x x x x x x x x x x	x x x x x	x x x x x x x x x	20′
8	8	x x x x x x x x x x	x x x x x x x	x x x x x x x x x	
6	9	x x x x x x x x x x	x x x x x x x	x x x x x x x x x	0.92 / 0.64
5	10	x x x x x x x x x x	x x x x x x x x	x x x x x x x x x	
40	1′			x x x x x x x x x	5-10
32	2			x x x x x x x x x	
25	3			x x x x x x x x x	f/11
20	4			x x x x x x x x x	
16	5	x x	x x	x x x x x x x x	1/125
13	6	x x x x x x x	x x x x x x	x x x x x x x x	32/25
10	7	x x x x x x x x x	x x x x x x	x x x x x x x x x	20′
8	8	x x x x x x x x x	x x x x x x x x	x x x x x x x x x	
6	9	x x x x x x x x x	x x x x x x x x	x x x x x x x x x	0.68 / 0.44
5	10	x x x x x x x x x	x x x x x x x x	x x x x x x x x x	
40	1			x x x x x x x x	5 - 11
32	2			x x x x x x x x	
25	3			x x x x x x x x	f/11
20	4			x x x x x x x x	
16	5	x x x	x	x x x x x x x x	1/250
13	6	x x x x x x	x x x	x x x x x x x x	32/25
10	7	x x x x x x	x x x x x x x	x x x x x x x x	20′
8	8	x x x x x x x x x	x x x x x x x x	x x x x x x x x	
6	9	x x x x x x x x x	x x x x x x x x	x x x x x x x x	0.76 / 0.49
5	10	x x x x x x x x x	x x x x x x x x	x x x x x x x x	
40	1			x x x x x x x x	5 -12
32	2			x x x x x x x x	
25	3			x x x x x x x x	f/11
20	4			x x x x x x x x	
16	5	x x	x	x x x x x x x x	1/500
13	6	x x x x	x x x	x x x x x x x x	32/25
10	7	x x x x x x x x x	x x x x x x x	x x x x x x x x	20′
8	8	x x x x x x x x x	x x x x x	x x x x x x x x	
6	9	x x x x x x x x x	x x x x x x x x	x x x x x x x x	0.74 / 0.47
5	10	x x x x x x x x x	x x x x x x x x	x x x x x x x x	

FIGURE 10.40. *A tabulation of individual test results, 5–9 to 5–12.*

Roll 3- Plus -X, 13 min. Microdol$_1$, 70°C, tray.

R.P.	Row	Shapes	Figures	Letters	#
40	1				3 - 12
32	2			x x x x x x x	
25	3			x x x x x x x x	f/11
20	4			x x x x x x x x x	
16	5			x x x x x x x x x	1/100
13	6			x x x x x x x x x	20/20
10	7	x	x x x	x x x x x x x x	20′
8	8	x x x x x	x x x	x x x x x x x x	
6	9	x x x x x x x x	x x x x x x x	x x x x x x x x	0.89 / 0.75
5	10	x x x x x x x x	x x x x x x x x	x x x x x x x x	

FIGURE 10.41. *A tabulation of individual test results ,3–12.*

The first six pictures of roll 5 (5–1 through 5–6 in fig. 10.38 and 10.39) are for high contrast and were taken by single direct exposures of the infinite contrast target at the marked exposure times given. The next six pictures of roll 5 (5–7 through 5–12 in fig. 10.39 and 10.40) were taken in reverse order of exposure time, and contained the direct exposure as marked and a fogging exposure of the photometric area, also of the exposure time marked. Hence, in this case the total exposure given the emulsion was in the ratio of 2/1, if we neglect the superposition of exposures. The recorded densities of high light and low light show what happened, and again we note from 5–10 to 5–12 no change in low light density in spite of a supposedly shorter exposure time for the last picture.

Figure 10.41 shows a similar set of observations made with a single direct exposure and two fogging exposures made for 3–12 at 1/100th of a second. The observed densities are 0.89/0.75. Figure 10.42 reproduces another low-contrast picture of densities 0.62/0.47. The patterns look well resolved in spite of the lowered contrast, but the picture *quality* is far from pleasing.

Table 10.3 gives a summary of the test results on roll 5 and on 3–12, and represents therefore three contrast groups. In order to make this tabulation the author arbitrarily assumed a criterion from figures 10.38 through 10.41 that at least 6 out of 10 test patterns had to be read correctly. The line so qualifying is then recorded in table 10.3 as the resolving-power equivalent. Thus, on 5–2 under *Shapes*

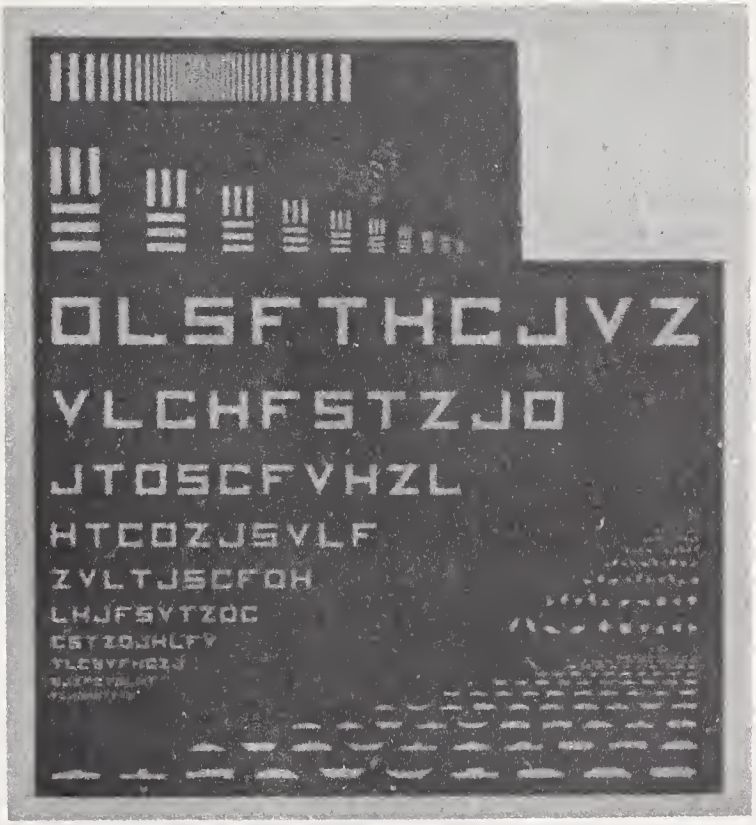


FIGURE 10.42. *Reproduction of a sample low-contrast target photograph.*

TABLE 10.3. *Target resolving power corresponding to 60-percent score*

Densities	Shapes	Figures	Letters	R.P.	Number
0.45/0.06	13	20	>40	36	5–1
0.53/0.07	20	16	>40	36	5–2
0.51/0.06	20	16	>40	36	5–3
0.80/0.07	16	16	>40	32	5–4
1.12/0.08	20	16	>40	32	5–5
1.46/0.33	16	20	>40	32	5–6
0.74/0.47	13	13	40	28	5–12
0.76/0.49	13	10	>40	28	5–11
0.68/0.44	13	13	>40	28	5–10
0.92/0.64	10	10	40	24	5–9
1.17/1.00	8	8	32	22	5–8
1.57/1.42	13	8	32	25	5–7
0.89/0.75	8	6	32	20	3–12

the equivalent resolving power 20 is recorded. This means that the three-line pattern at 20 lines/mm has a single line width equal to 0.025 mm and a height to each line of 0.125 mm. The corresponding *shapes* all have a longest dimension of 0.125 mm on the test negative and an average width of 0.025 mm. Under these conditions at least 6 out of 10 were recognized while at the same time the mean resolving power in the two directions on the same test negative 5–2 was observed to be 36 lines/mm. One would conclude that when the photograph resolves 36 lines/mm, shapes measuring 0.125 by 0.025 mm can be recognized with at least a 60-percent probability of correctness from among 10 varying objects of similar nature if the contrast is high on the photograph.

Table 10.3 indicates that on the whole the *Shapes* and *Figures* correspond well for equal difficulty of recognition. The letters on the other hand are so easily recognizable as to require a complete revision of the target.

It was obvious to the author in reducing the test negatives that there is a considerable variation in difficulty of recognition among the various shapes and figures in the same rows. After a few thousand readings have been made from test negatives taken with the best lens available at different scales, one could make up a new target with the shapes and figures on a line arranged for equal statistical difficulty, and similarly for the letters. This new target would then serve for testing inferior lenses and the criterion of a 60-percent score would become much more meaningful and more sharply defined between lines. As it stands, the diamond figure with its longest diagonal tilted at 45 degrees was recognizable down into the smallest lines of patterns. Likewise, among the shapes the so-called *truck* was recognizable, even when only a grain or two represented the wheels. Practice of the observer is also a real factor. The author became rapidly more skillful in interpreting the test negatives as he worked. Figures 10.38 through 10.41 were observed after the author had practiced for about an hour in a preliminary reading. The most difficult patterns to distinguish were those having *rounded* corners versus those with sharp corners, such as the shapes representing the basic profiles of a boat versus a canoe. The decay in the image quality caused by the lens-film combination does round off corners and therefore interferes with recognition.

Recognition Factors

Different types of objects have certain characteristics that lead to their recognition, as distinct from resolving power and detectability. Near the limit of the emulsion where the recognition factors become suppressed, recognition begins to fail. An imperfect lens in general will require a larger scale compared to the perfect lens before recognition can be achieved. Similarly, an increase in contrast of the image will improve the chances for recognition, whether this increase in contrast arises from use of a better lens or from higher target contrast or both. For example, the good lens may be able to show recognizability for a truck in the shade of a tree. The bad lens will fail here, but may nevertheless be able to provide recognizability for the truck at exactly the same scale, if the truck is out in sunlight. The laboratory target discussed above contains *recognition factors* for each of the shapes and figures involved.

Figure 10.43 shows these *recognition factors* in the case of standard block letters. The particular choice of the factors in the second line was made by the author, whereas other workers might adopt different criteria. However, an attempt is made here to have an arc as short as can be distinguished from a straight line segment, and to give only a minimum number of factors. Every letter can be distinguished clearly from every other letter in the scheme adopted.

The third line shows a message that appears entirely unfamiliar, and seems to be only a hodge-podge of letters. The fourth line shows a message that becomes legible upon close examination. The fifth line reproduces the fourth with fewer recognition factors, but is still recognizable once the fourth line has been read. Finally, the third line is simply the fourth line reproduced backwards without spaces. A trained interpreter could easily read the message in the third line. The point is that training is necessary in deciphering various types of objects, and the trained observer needs fewer recognition factors.

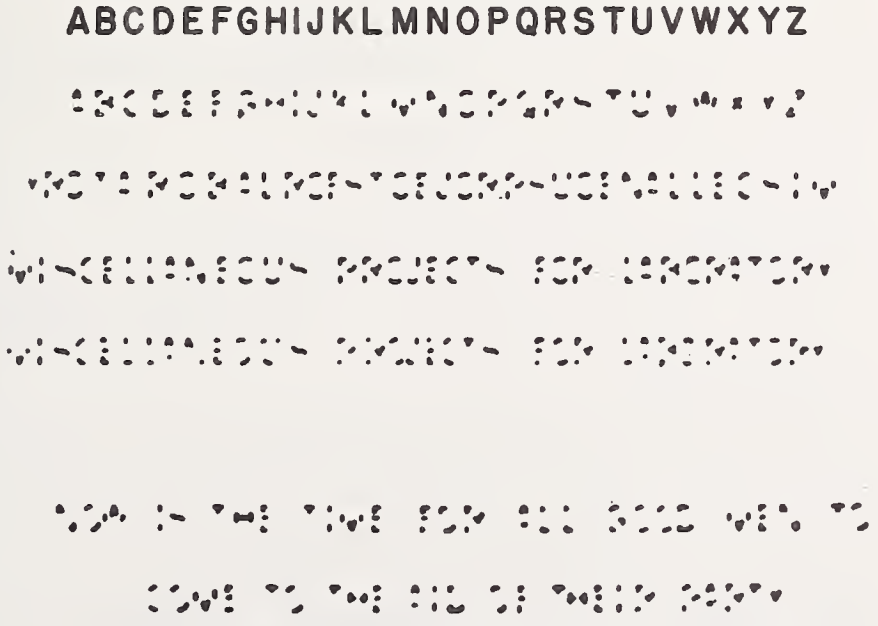


FIGURE 10.43. *Recognition factors for letters of the alphabet.*

The last double line presents a statement familiar to most people. These are ample clues for almost anyone to determine the meaning. A trained observer could read this line with fewer than half the clues. Context is so effective that clues required to decipher letters or words can be suppressed in the entire message. Hence, if the message is viewed as a unit, there will exist a minimum number of recognition factors for the message, as distinct from the individual letter. That is, each object or unit of information required will have this minimum number of recognition factors in context with the surroundings. A blur on a railroad track is likely to be a locomotive or a car. The same blur on a road is likely to be a truck. If on a runway, the same blur is likely to be an airplane, etc.

The type of lens testing described above is to be applied toward measuring the ability of the lens to produce recognition of standard objects and forms. Once correlations have been established with ordinary three-line patterns, the latter may continue to serve as a quick measure of lens performance. In the final analysis it should be possible to draw up a list of object types, and tabulate the scale versus contrast any given lens must employ in order to assure recognizability. A good lens should be able to provide such information at a smaller scale than a poor lens. Hopefully, the correlations between resolving power and contrast on the one hand, and the objective targets on the other, will be so distinct as to render the laboratory study sufficient for a broad range of applications in the field.

Discussion

Mr. D. P. Feder, National Bureau of Standards, Washington, D. C.: This question is addressed to Dr. Coleman. What about the difficulty of calculating the interferometer patterns that I notice seem to fit quite well to the observed pattern? How long a calculation is this?

Dr. Coleman: That takes about 8 hours.

Mr. Feder: Is that for any degree?

Dr. Coleman: Yes. You would have to know the aberrations to know the magnitude.

Dr. Baker: Is this with a test system?

Dr. Coleman: This is to calculate.

Dr. Baker: On a test instrument?

Dr. Coleman: Any system we considered.

Chairman: Are there other questions, please?

Dr. K. V. Pestrecov, Bausch & Lomb Optical Co., Rochester, N. Y.: This question is to Dr. Baker. You mentioned a beautiful wide-angle lens that resolves 60 or 70 lines/mm. Why isn't it used for aerial photography?

Dr. Baker: That is a spherically symmetrical lens with a curved field that I developed for aerial photography in 1941. The system has been used successfully. However, the curved field introduces awkward problems and as a consequence the lens has been little used. The first one was designed and built at the Harvard Observatory in 1941. Two more were built at Harvard during the war along with a rectifying printer projection lens designed by Dr. Grey.

Dr. Pestrecov: Were the calculations on the photographic image resulting from large color aberrations published?

Dr. Baker: The work will be published. It is part of our activities at Harvard Observatory right now.

Dr. Pestrecov: Is it contract work?

Dr. Baker: Yes, it is contract work. This portion of our work is declassified, as most of it probably will be. It is of interest, I think.

Dr. Pestrecov: It is very interesting. Do ou hope to publish it soon?

Dr. Baker: Well, I think so, as part of the proceedings of this Symposium.

Prof. B. O'Brien, University of Rochester, Rochester, N. Y.: I would like to ask Dr. van Heel if he has been successful in using ordinary photographic materials for recording these interesting patterns. Granted the color film is ideal but there are inconveniences as has been emphasized; can you not make rather successful measurements from black and white photographs?

Dr. van Heel: I did not understand you.

Dr. Pestrecov: Can you not make rather successful measurements in black and white?

Dr. van Heel: I am sorry. I have not been clear. The whole method depends on the fact that you can discern colors.

Dr. Pestrecov: Must this necessarily be true? For example, suppose you admit two narrow wavelength bands. Can you not use these as appropriately?

Dr. van Heel: Assume that you use two filters. The image of a red line and of a green line are next to each other on the plate. If the limit of resolution is infinitely small, you can get anything you want, but in practice it is hardly feasible to obtain that precision in black and white. I tried it with monochromatic light and with the photographic plate. It is difficult to locate the right position. I don't contend it is impossible, but why not use your ability to discriminate colors?

Dr. Pestrecov: I was thinking only of the practical difficulties of the time—was it two weeks?—required for the return of the processed films from Paris.

Mr. J. M. Naish, Optics Section, Instrument and Photographic Dept., Royal Aircraft Establishment, Hants, England: Would Prof. van Heel please say a little about the actual making of the grating? How accurate is it necessary to maintain the spacing and thickness of the wires?

Dr. van Heel: It should be perfect—a grating of that spacing would be like this (illustrating). I ought to know all of these positions within 2 microns, and it would be fine without recalibrating every time, but I have not succeeded in making such a grating so I have to calibrate for all positions and all the lines and to use my calibration table.

Mr. Naish: And what is the ratio of line to space?

Dr. van Heel: It is unimportant. In any case you get colors.

11. Geometrical and Interferential Aspects of the Ronchi Test

By G. Toraldo di Francia [1]

The Ronchi test, when performed with a low-frequency ruling, presents a purely geometrical character; ray optics is then fully sufficient to explain the appearance of the patterns. On the contrary, when a high-frequency grating is employed, the result is undoubtedly a phenomenon of interference and must be dealt with by means of wave optics. Still the interference patterns retain a close resemblance to the geometrical ones. It is interesting to investigate the cause of the resemblance and this is done very easily by means of the eikonal function. As the frequency increases the transformation of the shadows of the grooves into interference fringes is followed step by step. The theory is applied to the third-order aberrations and in that case it is found that, apart from a displacement, the interference pattern is identical to the geometrical one.

Introduction

One of the most sensitive tests for the quality of the image of a point is the Ronchi test. As is well known, this test consists in studying the light wave transmitted by an optical system by means of a diffraction grating. In spite of the fact that diffraction plays an important role in the observed phenomena, there can be no doubt that the ultimate result is the production of an interference pattern. For this reason it is customary to give the apparatus the name of "grating interferometer". On the other hand there are many cases when the test can be considered from a third point of view, that is, the merely geometrical one.

Since the early days of the Ronchi test there has been much speculation regarding its theory and interpretation. In this connection the question arose as to the simplest way of approaching the problem of interconnecting the three different aspects we have just mentioned. Many bulky calculations were carried out on this purpose, especially by the researchers of the Istituto di Ottica at Arcetri.

In almost all previous researches on the subject the simplifying assumption was made that the grating merely splits the impinging wave into many identical waves, rotated through a given angle with respect to one another. However this procedure did not bring about an actual simplification; on the contrary it led invariably to extremely lengthy calculations.

The purpose of this paper is to show how much the problem is simplified by a more rigorous approach that takes into account the actual phenomenon of diffraction. It will become apparent that in this way a better insight can be obtained of the passage from the pure geometrical phenomenon to the interferential one.

---

[1] Istituto Nazionale di Ottica, Arcetri, Firenze, Italy.

Some Geometrical Relations

The shape of the wave under test will be determined by means of the mixed eikonal function, or what amounts to the same, by means of its wave aberration at infinity. In other words the wave will be thought of as coming from infinity and its wave aberration will be computed with respect to a spherical wave of infinite radius, whose center we shall call C_r. Taking the axis of the wave as the x-axis, any point on the wave front at infinity can be defined by means of the two last direction cosines β,γ of the corresponding ray. The wave aberration $w(\beta,\gamma)$ will also be a function of these two cosines.

As is well known, from the theory of the eikonal function, if we call y and z the coordinates of the point where a given ray meets the plane through C_r perpendicular to the axis, we have

$$y=\frac{\partial w}{\partial \beta}, \quad z=\frac{\partial w}{\partial \gamma}. \tag{1}$$

From these relations it follows easily that, if w has the form

$$w=\frac{x}{2}(\beta^2+\gamma^2), \tag{2}$$

the wave is spherical and C_r is located at a distance x apart from the center of the wave.

Astigmatism corresponds to the wave aberration

$$w=\left(\frac{x}{2}+\frac{a}{4}\right)\beta^2+\left(\frac{x}{2}-\frac{a}{4}\right)\gamma^2. \tag{3}$$

If we assume the plane $\gamma=0$ to be the meridional plane, a is the distance from the radial focus to the tangential one and x is again the distance of C_r from the middle focus.

Third-order coma for a centered system is given by

$$w=\frac{x}{2}(\beta^2+\gamma^2)+c(\beta^3+\beta\gamma^2), \tag{4}$$

the plane $\gamma=0$ being taken as the meridional plane and C_r being at a distance x from the paraxial focus.

Third-order spherical aberration is given by

$$w=\frac{x}{2}(\beta^2+\gamma^2)+s(\beta^2+\gamma^2)^2, \tag{5}$$

where x has the same meaning as before.

Now let us take a thin rectilinear wire and put it in the path of the light wave. If the observer places his eye directly behind the wire, he sees the shadow of the wire projected at infinity. The shape of the shadow is an indication of the type of aberration that is present.

Now it is very easy to find an angular equation for this shadow. Let us suppose that the wire is parallel to the y axis and is placed at a distance z_0 from the x axis. We take C_r to be the point of the x

axis that has minimum distance from the wire. According to eq 1 the wire will meet with all rays satisfying the equation

$$\frac{\partial w}{\partial \gamma}=z_0. \tag{6}$$

This is the angular equation of the shadow. In practice, after writing eq 6 explicitly, one can replace β and γ by two rectangular coordinates Y and Z, respectively, and obtain the equation of the shadow as projected on some very distant screen. This is only an approximation, but a very good one, meeting all the requirements of practical application. Accordingly we shall rewrite eq 6 in the form

$$\frac{\partial w(Y,Z)}{\partial Z}=z_0. \tag{7}$$

In this way one obtains for a spherical wave from eq 2

$$xZ=z_0, \tag{8}$$

which is a straight line parallel to the y axis. This is represented in figure 11.1a, where the circle represents the exit pupil of the instrument.

In the case of astigmatism we assume that the meridional plane of the wave is rotated through an angle φ with respect to the y axis, that is, with respect to the wire. Thus we replace β with $\beta \cos \varphi - \gamma \sin \varphi$ and γ with $\beta \sin \varphi + \gamma \cos \varphi$. With this substitution the wave aberration (3) takes the form

$$w=\left(\frac{x}{2}+\frac{a}{4}\cos 2\varphi\right)\beta^2-\frac{a}{2}\beta\gamma\cos 2\varphi+\left(\frac{x}{2}-\frac{a}{4}\cos 2\varphi\right)\gamma^2.$$

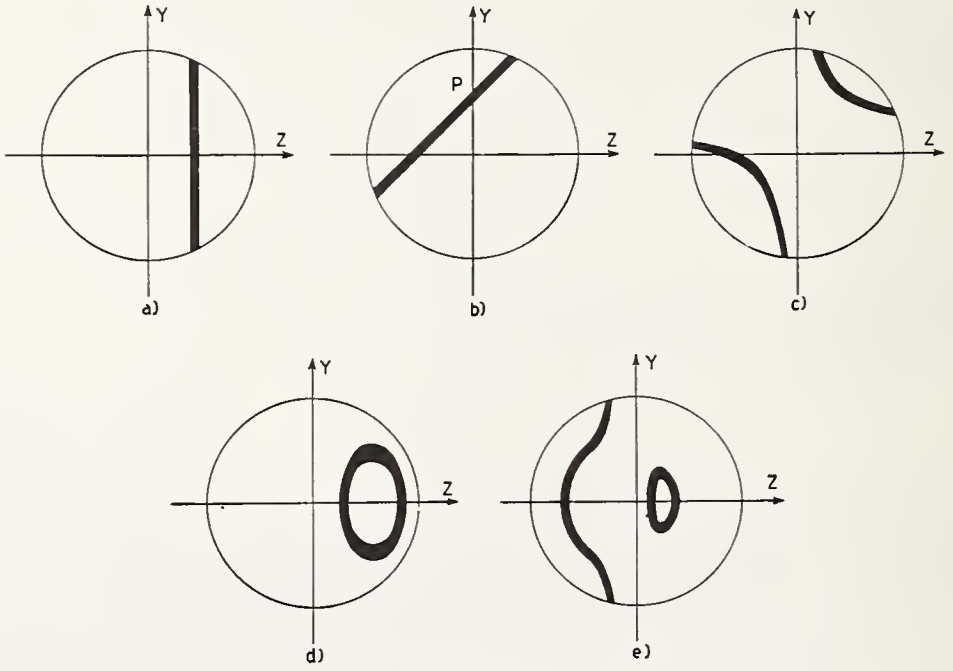


FIGURE 11.1. *The shadow of a thin straight wire.*

a, Perfect spherical wave; b, astigmatism; c, coma (wire parallel to the meridional plane); d, coma (wire perpendicular to the meridional plane); e, spherical aberration.

By applying eq 7 we easily arrive at

$$Y=\frac{2x-a\cos 2\varphi}{a\sin 2\varphi}Z-\frac{2z_0}{a\sin 2\varphi}. \tag{9}$$

This equation represents a straight line. When x is made to vary, the line turns around a fixed point P of the Y axis (fig. 11.1, b).

Let us consider the case of coma. By applying (7) to (4) one easily gets

$$Z(2cY+x)=z_0. \tag{10}$$

The shadow has, therefore, the shape of an equilateral hyperbola, with the asymptotes parallel to the Y and Z axes, respectively (fig. 11.1, c). The center of the hyperbola is located on the Y axis (that is on the meridional plane) at a distance $-x/2c$ from the axis. It is interesting also to consider the case where the wire is perpendicular to the meridional plane, instead of being parallel to it. One has only to write (4) in the transposed form

$$w=\frac{1}{2}x\,(\beta^2+\gamma^2)+c\,(\gamma^3+\gamma\beta^2), \tag{11}$$

where the roles of β and γ have been interchanged, and to apply the eq 7. Thus, one obtains

$$xZ+3cZ^2+cY^2=z_0. \tag{12}$$

as the equation of the shadow. This is an ellipse, with its axes parallel to the Y, Z axes (fig. 11.1, d.) The center of the ellipse is located on the Z axis at a distance $-x/6c$ from the axis of the wave.

Finally in the case of spherical aberration, we use eq 5 to obtain the equation

$$Z[x+4s(Y^2+Z^2)]=z_0. \tag{13}$$

This represents a cubic, very well known in optics (fig. 11.1, e).

The Grating Interferometer

Let us now consider the grating interferometer. If the spacing of the ruling is large, we can apply ray optics, considering the pattern as consisting of the shadows of the grooves. It would then be very easy to discuss the patterns, by means of the results of the preceding section. However, we will take another way of approach that will prove much more instructive.

According to Huygens principle, each point of the wave front at infinity sends out a spherical wave, which arrives at the center of the reference sphere as a plane wave. The phase of this plane wave can be computed by means of the wave aberration at infinity and is evidently $-2\pi w/\lambda$.

Thus, we arrive at the decomposition of the impinging wave into a bundle of plane waves, each having its own orientation and phase. Each one of these plane waves, when incident upon the grating is

518-743 O - 73 - 25

split into a set of diffracted plane waves. We assume the grating to be perpendicular to the x axis and the grooves to be parallel to the y axis.

According to the laws of diffraction by a grating, a diffracted wave of the k'th order will have the direction cosines β, γ if the original wave had the direction cosines β', γ', given by

$$\beta' = \beta$$

$$\gamma' = \gamma - k' \frac{\lambda}{p},$$

p being the period of the grating. In the same direction we find the diffracted wave of k''th order, corresponding to an incident wave whose direction cosines are

$$\beta'' = \beta$$

$$\gamma'' = \gamma - k'' \frac{\lambda}{p}.$$

According to what we have stated above, the phases of these two waves will be, respectively,

$$\varphi'' = -\frac{2\pi}{\lambda} w\left(\beta, \gamma - k' \frac{\lambda}{p}\right)$$

$$\varphi'' = -\frac{2\pi}{\lambda} w\left(\beta, \gamma - k'' \frac{\lambda}{p}\right).$$

There will be positive or negative interference, accordingly, as the difference of the phases is a multiple of 2π or an odd multiple of π. Therefore, the angular equation of the nth bright fringe at infinity will be

$$w\left(\beta, \gamma - k'' \frac{\lambda}{p}\right) - w\left(\beta, \gamma - k' \frac{\lambda}{p}\right) = n\lambda.$$

If only the two waves considered were present, an observer, placing his eye directly behind the grating, would see an interference pattern whose nth bright fringe on an infinitely distant screen would have the equation

$$w\left(Y, Z - k'' \frac{\lambda}{p}\right) - w\left(Y, Z - k' \frac{\lambda}{p}\right) = n\lambda. \tag{14}$$

This is the fundamental functional equation of the grating interferometer.

The Passage from Geometrical Optics to Wave Optics

Let us now assume that the spacing p is very large with respect to the wave length. In this case $k'\lambda/p$ and $k''\lambda/p$ are very small and we can expand the functions w of eq 14 in Taylor series, retaining

only the first derivatives. Thus we obtain

$$\frac{\partial w}{\partial Z}=\frac{np}{k'-k''}.$$

For two consecutive waves $k'-k''=1$, and we get

$$\frac{\partial w}{\partial Z}=np. \tag{15}$$

A comparison with eq 7 shows that the nth bright fringe has the same equation as the projection of the nth transparent groove of the grating; the dark fringes are therefore coincident with the shadows of the opaque groove of the grating.

This result is valid for the interference of two consecutive waves. But we can easily extend it so as to take into account all the waves diffracted by the grating. Indeed, we can divide the entire set of diffracted waves into pairs of consecutive waves, each pair forming its dark fringes on the geometric shadows of the opaque grooves. By superimposing the interference patterns of all these pairs we get evidently as a net result the shadows of the grooves.

In this manner, although we started from wave optics, we have accounted for a purely geometric phenomenon. We can therefore utilize the results of our discussion of the shadow of a wire. The patterns of figures 11.1 now become, respectively, the patterns of figure 11.2.

Let now the spacing p become smaller, so that a better approximation is needed. In the Taylor expansion of the functions w of eq 14

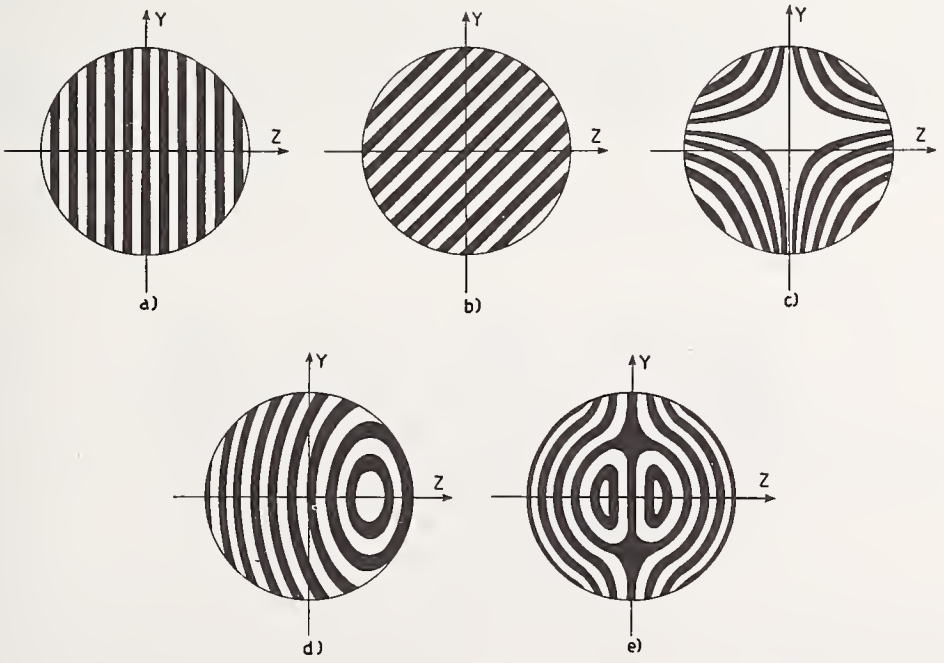


FIGURE 11.2. *Interference patterns with a low-frequency grating.*

a, Perfect spherical wave; b, astigmatism; c, coma (meridional plane parallel to the lines of the grating); d, coma (meridional plane perpendicular to the lines of the grating); e, spherical aberration.

one must retain also the second derivatives. Assuming for the sake of simplicity $k'-k''=1$, one obtains

$$\frac{\partial w}{\partial Z}-\frac{k'+k''}{2}\frac{\lambda}{p}\frac{\partial^2 w}{\partial Z^2}=np. \tag{16}$$

Let us put

$$\frac{k'+k''}{2}\frac{\lambda}{p}=\delta. \tag{17}$$

We see at once that in the present approximation eq 16 can be written

$$\left[\frac{\partial w}{\partial Z}\right]_{(Y,\,Z-\delta)}=np, \tag{18}$$

where the left side is $\partial w/\partial Z$ computed at the point $Y,Z-\delta$. It is easy to interpret eq 18 if we notice that δ is the arithmetic mean of the angles of diffraction $k'\lambda/p$ and $k''\lambda/p$ of the two waves k' and k'', respectively. A comparison of eq 18 with eq 15 shows that the entire pattern has been shifted by the amount δ along the Z axis with respect to the preceding approximation. This happens because the scattering of the different waves with respect to the central wave has now become apparent and the field of interference of the waves k', k'' is centered around the point $Y=0$, $Z=\delta$. The situation is illustrated in figures 11.3a, 11.3b, 11.3c, 11.3d, and 11.3e, each case corresponding to that designated with the same letter in figures 11.1 and 11.2. Apart from the shift, the patterns are still identical to those of the shadows of the grooves. Of course, they are repeated many times; once for each possible pair of waves. If the successive interference fields partially overlap, the result is a very complicated phenomenon of multiple interference. However, in figure 11.3 we have represented the case where p is already sufficiently small for being out of this intermediate stage of confusion.

Finally we pass to a third approximation. First we shift the origin to the point $Y=0$, $Z=\delta$; owing to the definition (17) of δ, eq 14 is replaced by

$$w\left(Y,Z+\frac{k'-k''}{2}\frac{\lambda}{2}\right)-w\left(Y,Z-\frac{k'-k''}{2}\frac{\lambda}{p}\right)=u\lambda. \tag{19}$$

We put $k'-k''=1$, and

$$\epsilon=\frac{\lambda}{2p}. \tag{20}$$

Then we expand the functions w of eq 19 in a Taylor series, up to the third derivatives inclusive. We find thus

$$\frac{\partial w}{\partial Z}+\frac{\epsilon^2}{6}\frac{\partial^3 w}{\partial Z^3}=np. \tag{21}$$

The new feature with respect to eq 15 is represented by the term containing $\partial^3 w/\partial Z^3$. This term has no influence when only such aberrations are present as are described by eq 2, 3, and 4, where no power of γ higher than the second appears. In the casd of eq 11, the term

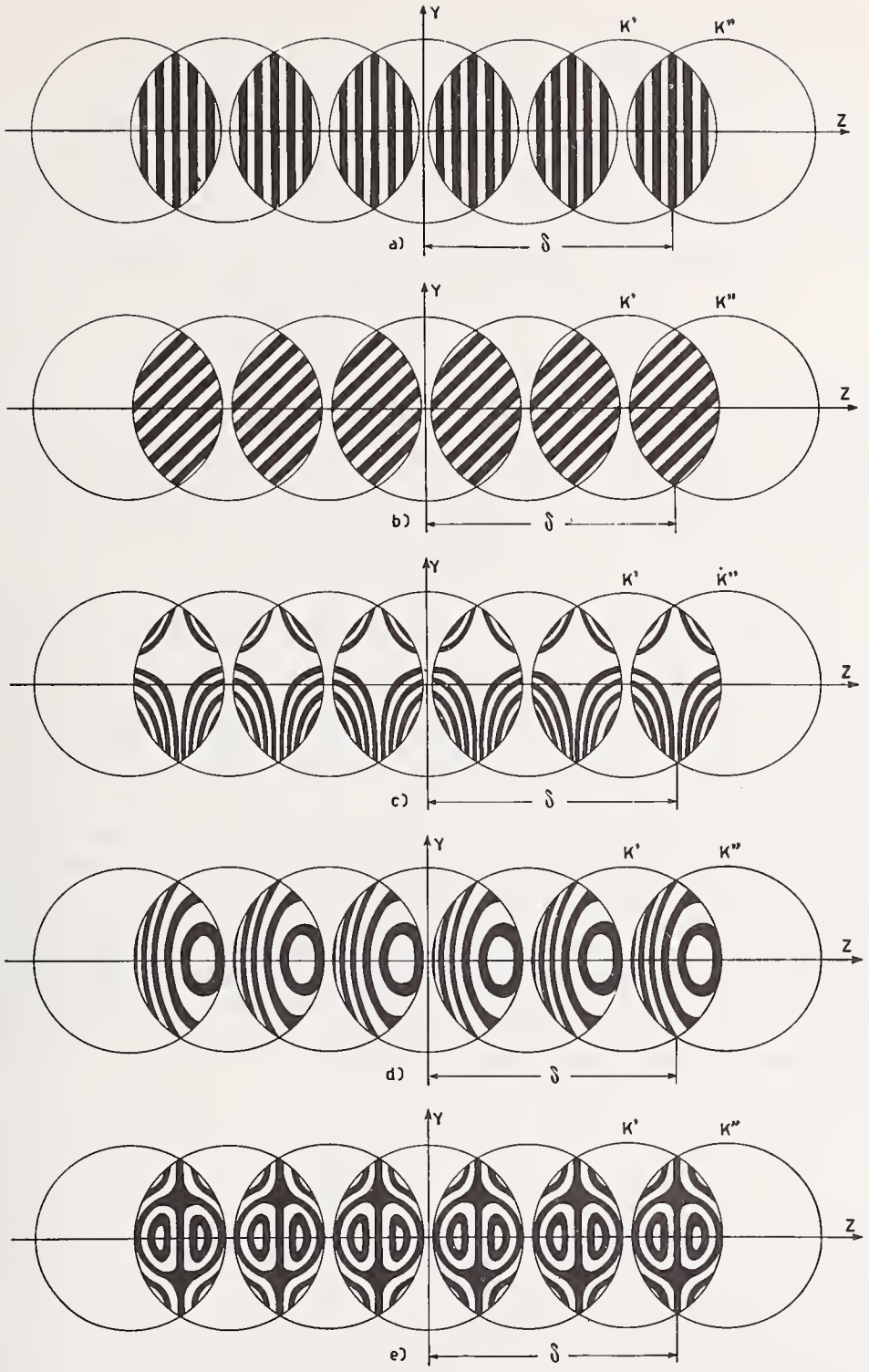


FIGURE 11.3. *Interference patterns with a high-frequency grating.*
a, b, c, d, and e, correspond respectively to the same cases of figure 11.2.

$c\gamma^3$ would only bring a constant in the left side of eq 21; this would amount to the same as changing n, i. e., the order of each fringe, and would by no means alter the structure of the pattern.

We conclude that, apart from the lateral shift, the interference patterns of the third-order aberrations up to coma are identical to the corresponding shadow of the grating, repeated as many times as there are pairs of consecutive waves. This result evidently holds good in any further approximation, depending on higher derivatives of w.

As regards spherical aberration, insertion of (5) into (21) brings to

$$Z[x+4s\epsilon^2+4s(Y^2+Z^2)]=np. \tag{22}$$

A comparison with eq 13 shows that the difference is represented by the constant term $4s\epsilon^2$ in the square brackets. This term has the same effect as a change of origin for x. It follows that in the present approximation we have exactly the same patterns as in the preceding ones; but they occur at different locations of the grating along the axis of the wave. This result holds in any higher approximation.

Conclusion

We have thus given all the elements that are needed for a discussion of the Ronchi test as applied to third-order aberrations. We have purposely not dwelled on the many rules that can be given for the quantitative evaluation of actual aberrations by means of the interference patterns. These rules are very easy to derive once the basic equations given in the present paper are known and their meaning is understood.

We have shown that both the geometric and the interferential versions of the Ronchi test give rise to the same type of patterns; but this does not mean that they are equivalent. Of course the interferential one is much more sensitive and must be applied in all the cases where a high degree of accuracy is required. One must not be deterred from making use of it by the fear of all the fine adjustments that are inherent in every interferential test. Indeed, the grating represents the only interferometer that need not be adjusted; it could be termed a self-adjusted interferometer.

12. A Combined Test Procedure for Camera Lenses, and Photoelectric Examination of Intensity Distribution in Line Images

By Erik Ingelstam and Per J. Lindberg [1]

Lens Tester

The first part of this paper deals with a simple and convenient instrument for testing lenses, which we believe to be novel. By means of this instrument it is possible to determine the tangential and radial resolving power (estimated from a photographic negative), the corresponding image surfaces, and the distortion, by a method that is rapid and inexpensive. For a more complete description reference should be made to an earlier report from our laboratory [1].[2]

Apparatus

This instrument is shown in figures 12.1 and 12.2. The characteristic feature is a collimator provided with a "tilted test plate" making an angle of 60° with the focal plane and shaped as shown in figure 12.2. Its function is evident from the figure. An advantage of the method is that the test is made with the lens mounted in the camera to which it is fitted, with its own film holder and in conjunction with a filter if desired. The test can be made at different field angles, β, from the axis and on the resulting negative one can read the point of greatest sharpness. This point is related to the distance from the emulsion to the position of greatest sharpness by the quadratic equation shown in figure 12.1. The patterns of line groups on the plate shown in figure 12.2 have been distorted to produce the known square contour of the Foucault groups on the test negative as shown in figure 12.3. The determination of the x positions is very rapid, and in the legend comments are made concerning the presence in this case of one region of best resolution and one of best contrast as discussed by Kingslake in this symposium.

This type of test plate has been of good use for different purposes. The requirements of the Royal Swedish Air Force brought about the construction of most of the apparatus reported here, and it has been used for most of the qualification tests of new lens designs as well as for a routine collimator to be used in the over-all program of the Swedish Air Force for controlling and adjusting the cameras for air reconnaissance. The authors are indebted to O. Hagsten, Chief of the Photographic Department of the Royal Swedish Airboard and his staff, for excellent collaboration and advice concerning the many practical problems in aerial photography.

[1] Optics Laboratory, Royal Institute of Technology, Stockholm 70, Sweden.
[2] Figures in brackets indicate the literature references at the on p. 182.

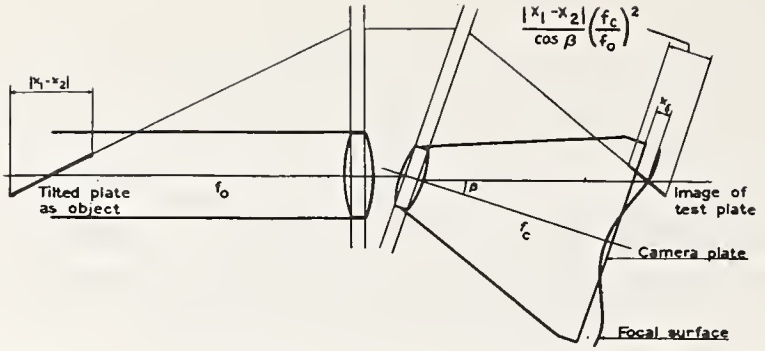


FIGURE 12.1. *Arrangement with the "tilted test plate" in the collimator.*

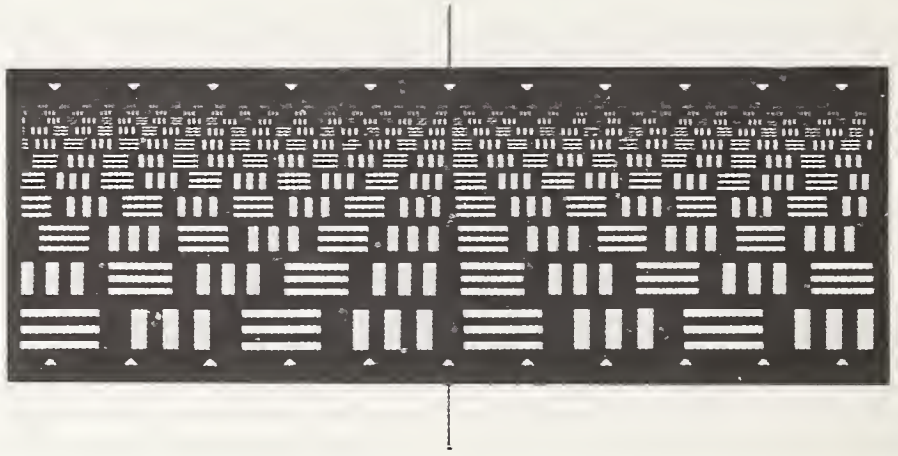


FIGURE 12.2. *Shape of the "tilted test plate," the consecutive line groups being in the scale ratio of* $2^{1/3}$.

FIGURE 12.3. *An enlarged record of a tilted test plate with interrupted line groups.*

The smallest groups are best resolved at $x\delta$, which coordinate indicates the intersection between the test-plate image and the camera plate, so the focusing defect is the distance between $x\delta$ and $x=0$.

At the point x_2 out of focus there arise line groups with reduced number of lines, which is sometimes a form represented in the record. At x_3 large groups are imaged without halation, and at first one would say that the best sharpness is situated there, but $x\delta$ indicates the best "micro-sharpness."

Another application of the tilted test plate is the testing of the optics of fluorographic cameras, which is reported elsewhere [2].

In order to obtain the distortion of a camera simultaneously with the test of its resolving power, we have made use of a support, of which the main features are shown in figure 12.4. Not having the several fixed collimators used in the larger laboratories, we turn the camera support around an axis, the angles being read with the precision of a few seconds of arc. Besides the plate (fig. 12.2), the collimator is provided with indices in the focal plane designed to photograph well with different resolutions. In the series of records taken as the camera is turned to equally spaced fixed angular positions, the positions of these indices can be exactly measured in the comparator. In order to avoid reading the micrometers on the precision scale every time, the support has fixed angular positions, maintained by the ball and spring positioner shown in detail in figure 12.5. It has been proven that the positions determined by a ball-bearing ball positioning into conical holes are, because of freedom from hysteresis, reproducible within less than 1 micron. Therefore, in certain cases, when it is not necessary to know the absolute value of the distortions, but

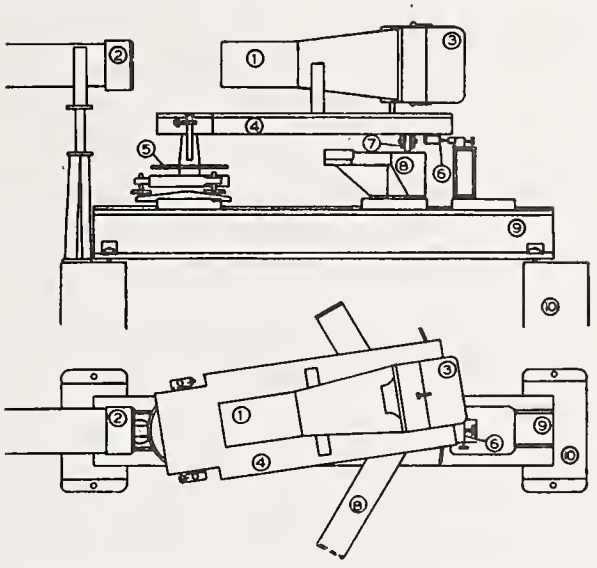


FIGURE 12.4. *Side and upper view of rotatable camera support.*

1, Camera-lens system; 2, collimator lens; 3, film or plate holder; 4, rotatable camera support; 5, goniometer precision scale; 6, ball and spring positioner; 7, ball bearing of the rotatable support; 8, V-shaped path for the ball bearing; 9, optical prism bench; 10, box-like base.

FIGURE 12.5. *Ball and spring positioner.*

only relative displacements across the image field, these positions form the necessary fixed points. This is the case when lateral chromatic aberration is examined; parallel adjacent rows of records are taken with six different interference filters inserted between the test plate and the light illuminating the collimator, the successive rows being displaced by slightly tilting the collimator. A cross-table measurement of the indices obtained in this way is sufficient to evaluate this type of aberration. The corresponding axial chromatic aberrations in the radial and tangential directions are, of course, at the same time determined from the records of the tilted test plate.

It should be clear that all photographically achievable information about a camera can be recorded on one single photographic plate by means of this procedure. The values of the resolving powers obtained are relative ones photographically recorded and visually estimated. Because of this lack of information the instrument described in the second part of this paper was constructed.

Image Scanner

In our laboratory we have found it valuable also to construct equipment for measuring intensities in optical images, and accordingly, a photoelectric scanning device has been constructed. The arrangement is suitable for measuring across a line, as the light passes through a linear slit to the phototube. The sensitivity is sufficiently great to permit the use of a pinhole aperture, e. g., of a few tenths of a micron in diameter if this should be necessary, but the "indefinitely long" linear slit is appropriate for most purposes. Of former constructions of this class the interesting method of Jones and Wolfe [3] may be mentioned. The intensity distribution perpendicular to the slit was obtained by photographing through a gray wedge, thus obtaining logarithmic curves that gave much information about image intensity distribution in a lens system. Further, Herriott [4] has devised a registration device utilizing a photocell, its current being coupled to the y-coordinate of an oscillograph whose x-axis was synchronized with the movement of the entrance slit across the image, the movement being maintained by the electromagnetic system of a loudspeaker. In this way the oscillograph directly shows the intensity curve. Surely, other laboratories in optics must also have devised similar equipment utilizing photoelectric measurements of aerial images.

The main characteristics aimed at in our construction are : (1) Wide range of sensitivity with true response (10^6), which makes it necessary to have the power supply for the photomultiplier tube, the amplifier contained in the system, and the recorder, all well and precisely constructed and quickly adjustable by switches to different ranges of sensitivity; (2) freedom from extraneous light in the system itself, a feature that is absolutely necessary in order to cover the large range to which reference has been made (achievable only by avoiding the use of any type of subsequent magnifying system after the image, such as a microscope); (3) very narrow entrance slits, as we still want a true shape of the intensity function perpendicular to the slit.

Photocell carriage. The photomultiplier tube is an RCA 931–A, its mounting with base above being shown in figure 12.6. The adjustment of the device before inserting the phototube is obtained by the

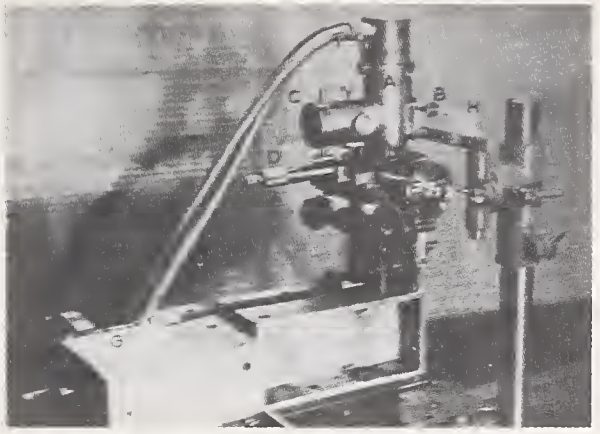


FIGURE 12.6. *Photoelectrical unit.*

A, Tube housing; B, screw for slit adjustment; C, tube for insertion of adjusting microscope; D, micrometer screw for axial slide displacement; E, micrometer screw for transversal slide displacement; F, gear box for E screw; G, synchronous motor; H, glass under test.

microscope inserted into a tube at the rear. The upper part of the carriage is adjustable in all directions; if the slit is to traverse a plane which is obliquely cut by the light rays, the tube housing can be turned before clamping. A micrometer serves for the displacement in the axial direction. A precision advance laterally is obtained by a micrometer screw pushing a slide. The synchronous motor is coupled to the gear box giving to the slide a variety of speeds easily changed during the recording. The most commonly used speeds are 25, 5, 1, and 0.2 microns per second. Of course, everything must be very substantially built in order not to be disturbed by occasional vibrations in the building during the recording.

Production of the slit. The slits are formed by shadowed metal evaporation onto glass plates. Before evaporating the aluminum the intended slit area is covered by a quartz thread drawn as described in Strong's Handbook [5], and fixed in position by tapes. A selection of such slit plates from 1-micron slit width up to about 10 microns for use as will be reported later was prepared. Made by this method they are precisely uniform, the measurement of their widths being determined by viewing the positions of their Fraunhofer diffraction minimas. One eye looks at a slit illuminated with monochromatic light, the other eye observes where the diffraction minima fall on a metric scale. In this way the uniformity of a slit is rigorously controlled, and a precise measure of its width is achieved. The slit side of the glass plate is always turned toward the incident light. Between the slit plate and the photomultiplier tube a diffusing screen is mounted to equalize the intensity over the photosensitive area thus securing a constant gain.

Electric equipment. The potential applied to the multiplier tube is supplied by a rectifier and stabilizing unit shown in figure 12.7 as P. It gives variable dynode voltages, of 50, 70, and 100 volts. The photomultiplier current is amplified by a balanced d-c amplifier, A, with decade input resistances, making it possible to vary the sensitivity in decades by means of simple switching. The output of the A unit leads to the Speedomax recorder, ordinarily adjusted, however, for a balancing time of 1.5 seconds.

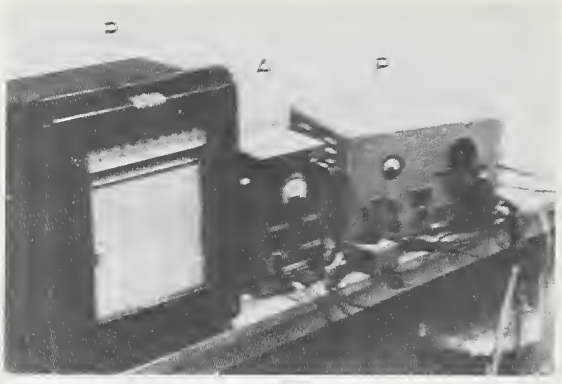


FIGURE 12.7. *Electric equipment.*

P, Power supply; S, switch; A, amplifier; R, recorder.

Records of test line figures, etc. We need not go in any detail into the troublesome questions related to resolving power in this symposium. It should be sufficient to point out that most of the practical men dealing with and buying photographic lenses still have nothing more than a specification that the resolving power is e. g., 38 lines/mm at the image center, though we know that such a specification is so incomplete as to have little value because it is based on a subjective estimation of the limit of visual recognition of structure by means of magnification on a photographic film, possibly also with unknown graininess. In view of the large amount of basic research by Selwyn, Arnulf, Coleman, and others we feel that there is now a great need to base measurements upon real physical quantities and to separate a *physical resolving power* which only deals with the intensity distribution in the aerial image.

Even in this idealized case of an infinitely narrow long object slit the specifications are, as well known, not unambiguously defined.

In spectroscopy the conventional measure of resolving power is, in spite of criticism, the well-known Rayleigh criterion. The application of this criterion in spectroscopy generally requires that the natural line width be substantially infinitely small. Therefore, very narrow spectrum lines and appropriate slit widths are used. The criterion corresponds to a drop in the intensity curve between the two maxima of 19 percent, this figure being partially conventional. As is well known, the eye easily resolves closer line objects, which often leads to confusion. The Sparrow criterion [6] which demands that between two or more lines, no drop shall be present in the intensity curve, or the slope shall be zero at one intermediate point, thus defines a limit of resolving power less severe. The criterion has also some further advantage.

When one tries to define physical resolving power for line images for photographic and similar optical apparatus, it seems to be proper to consider the infinitely narrow line. From this ideal object, by integration, one can derive the common test figures with a rectangular intensity shape, as the Foucault targets, and also the desirable Selwyn test objects with a sinuoidal intensity function. This is possible by means of known mathematical procedures [7, 8]. The criterions mentioned, however, relate to the shape of the curves at the top, or to the slope, and they are insensitive to the presence of stray light.

The stray light is taken into account in Selwyn's tests but only when the tests are performed in such a manner that the whole area or a known area of the light field is occupied by the test figures, which in practice is often neglected. The test figures with rectangular shape and decreased contrast now often in use are no doubt valuable for routine practice, but there is a link missing between what they give and the problem of the influence of stray light on a physically ideal image.

If one can obtain a well-defined physical conception of the resolving power, and by means of recording devices such as the one reported here analyze experimentally the true intensity shape of lines, it would be possible to relate this physical resolving power to visual resolving powers and the photographic resolving power under the various circumstances as investigated by Arnulf, Selwyn, and others. In the general case it demands recalculations which may possibly be inconvenient, but there is still a large interest in exactly knowing the intensity distribution in one-dimensional optical images. In designing and constructing this highly precise image scanner, the chief purpose has been to provide a means for making objective physical measurements of resolving power and contrast. The experimental results here presented illustrate its suitability for this purpose.

Special Records

Experimental results. Several series of records have been made with test plates of the ordinary rectangular shape in our device. Figure 12.8 shows records taken with lines in groups of three in geometric progression. They were taken with commercial photographic objectives. Without any large uncertainty, the resolving power according to the Sparrow criterion can be determined from the first group where the undulations at the top disappear. It must, however, be understood that one deals here with broad lines, their widths being equal to the space between them, and not with an ideally narrow line.

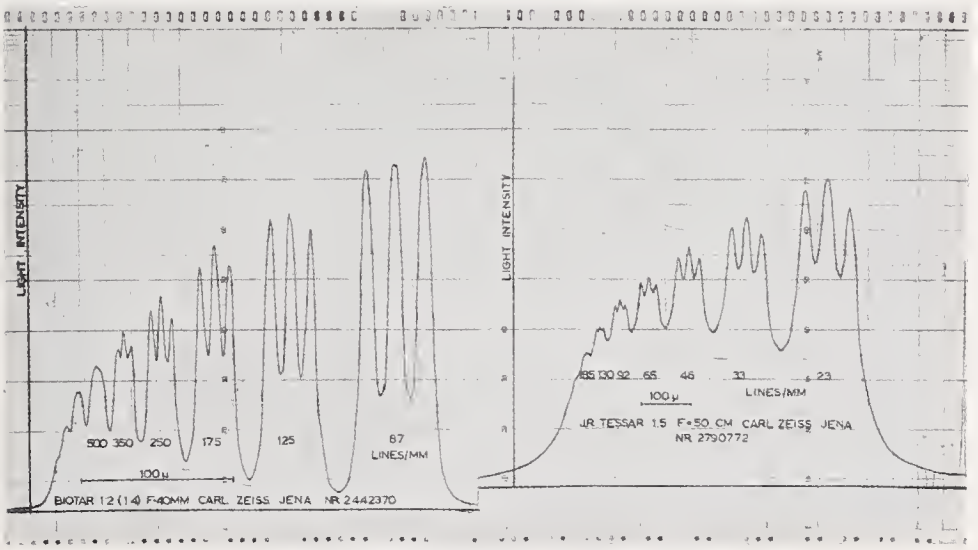


FIGURE 12.8. *Records of three-line Foucault groups.*

Accordingly, a recalculation of the ideal narrow line is necessary; such integration procedures being possible by means of mechanical or optical devices. Collimators with different focal lengths were used, the objects being placed in their focal planes. When the focal length of the objective to be tested was short, the object was placed about 15 meters distant without a collimator. In these records white light was used.

A better solution for the study of a physically ideal line is achieved by examining a single line imaged by the lens under test. A special purpose in taking up this research is the examination of the influence of poor polish of glass surfaces on the intensity distribution.

When examining the image intensity distribution in the outer part of the line image where the intensity has dropped to 10^{-5} or 10^{-6} complete freedom from scattered light and other disturbances must be achieved. For example, in the plate forming the entrance slit of the photocell there may be some very small pinholes in the evaporated layer that permit the passage of stray light. In order to be free from this, an additional opaque screen is placed between the slit and the source with a slit approximately 20 microns wider than the definitive slit.

The apparatus is arranged for this purpose as shown in figure 12.9. The object slit consists of a uniformly broad straight slit. It is illuminated by a monochromator of special type. This monochromator consists of a plane grating and a lens, this device imaging a monochromatic part of the light from a ribbon-filament lamp onto the slit. This arrangement provides many times the illumination that is provided by interference filters, which have also been used. The lens whose intensity distribution was primarily studied was a plano-convex lens of F=1,000 mm and a diameter of 45 mm, the edges being shielded to prevent reflection of any noticeable amount of light.

The special investigation taken up is the influence on the intensity distribution of different polishes on a glass surface situated near the focal plane. For this purpose two nearly equal thick plano-parallel plates, one of smooth, one of rough polish, were subsequently placed as shown in figure 12.9. Because of the short distance between surface and focal plane, only about one-hundredth of the focal length, a bad polish causes mainly stray light; when the distance is larger, the

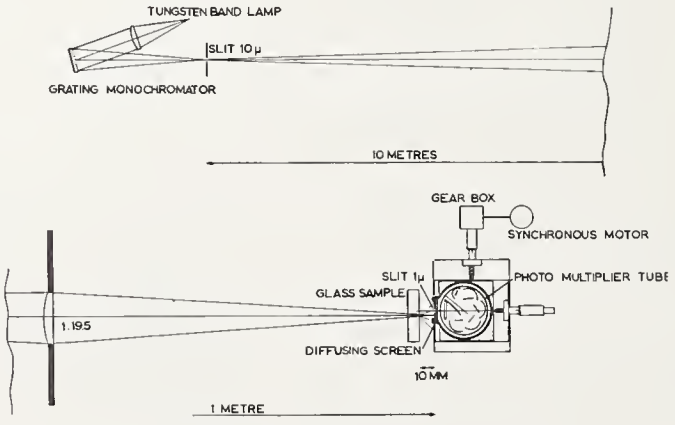


FIGURE 12.9. *Apparatus outline.*

central part of the image will be affected causing loss of resolving power. The additional spherical aberration caused by the plate, as well as the original spherical aberration, is of minor importance for the total shape of the curve and can be calculated with sufficient precision. The area where the rays traverse the glass plates is precisely determined by records on photographic paper to permit their examination with multiple beam interferometry; they were in advance selected by examination in a Michelson interferometer and had sensibly perfect planeness and homogeneity.

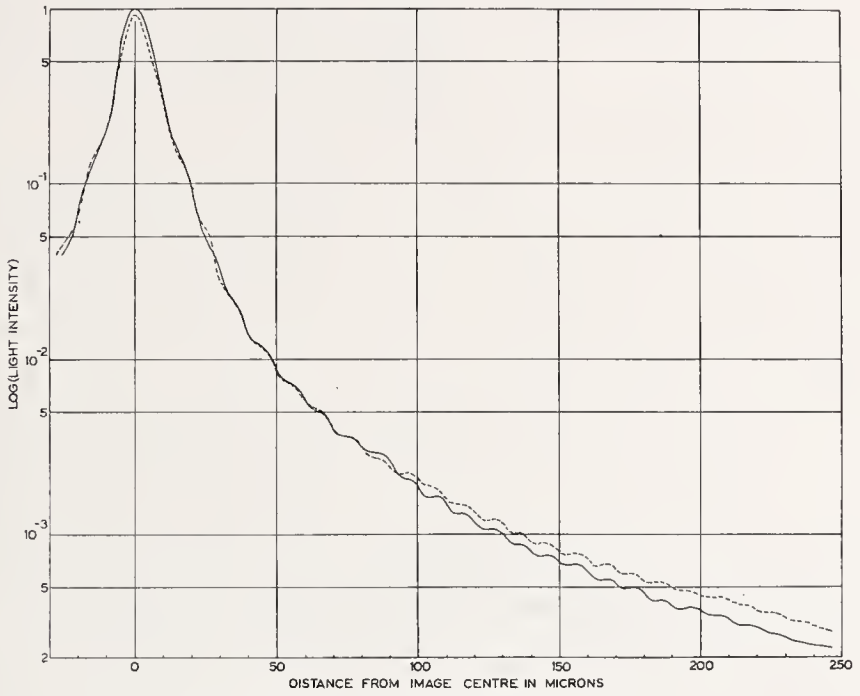


FIGURE 12.10. *Intensity distribution.*

Solid curve, fine-polished glass; dotted curve, rough-polished glass.

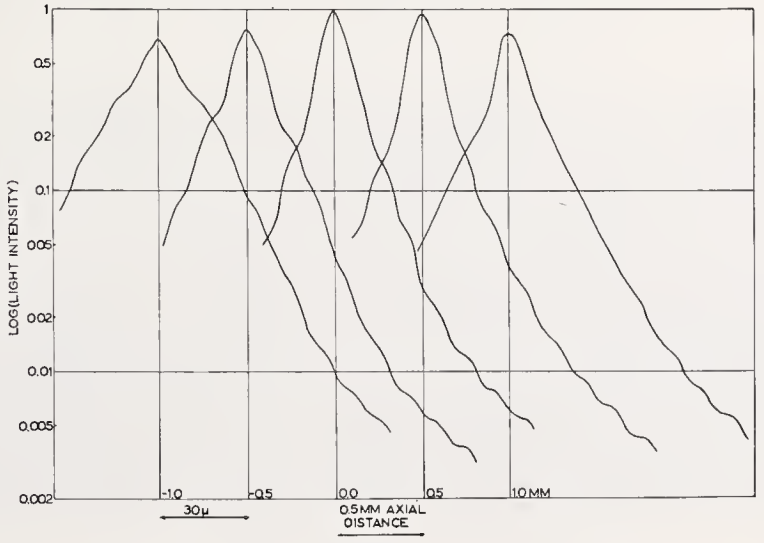


FIGURE 12.11. *Different axial cuts of image.*

Figure 12.10 gives two typical curves for the central part of the image transformed to a logarithmic scale of intensity. The x-coordinate is linear. An interesting feature is the true record of the outer parts of the image, the undulations up to about the 20th being clearly recorded. A test of the correct performance of the carriage micrometer drive is afforded by the coincidences of the fringes. In figure 12.11 the central image is represented, down to about $10^{-2.5}$ of the maximum intensity. The different curves are displaced laterally in this diagram for different positions along the axis. There are different parallel sections of the ray about the plane of the best focus. Figure 12.12 shows an original record for illustrative purposes. The sensitivity range has been changed three times, namely, in the ratios of 10, 10, and 19, the exact values of these ratios having been determined separately. The fluctuations of the last record is due to electronic noises and thus limit the accuracy.

In order to pass over to still larger angles, the intensity was increased about 100 times by widening the slits. This affords less detail and accuracy in the central part, which is of no great importance as this second record covers a region extending some hundreds of microns beyond the first record. These recordings have been extended to approximately 4 mm from the center, thus covering scattering angles up to about 20 degrees.

In figure 12.13 the corresponding curves for the best focal position in the two cases, with high-polished glass and with rough-polished glass are traced, logarithmically in the x-direction. The curves without glass are, within the accuracy, close to those for the highly polished glass. A comparison shows that the centers are somewhat changed in the sense that the rough-polished glass gives a lower intensity. The outer regions are still more different, the light spread resulting from the fine structure of the wavefront due to the undulated profile curve is quite evident and well measurable within desired accuracy.

Quantitative relations between the amount of stray light and the shape of the wavefront as studied by means of the multiple-interference or phase contrast may, we hope, be established when more data are gathered and treated in detail. The interest of such an investigation seems to be indicated by, on one side, the possibility nowadays, thanks to phase contrast and other methods, to examine precisely

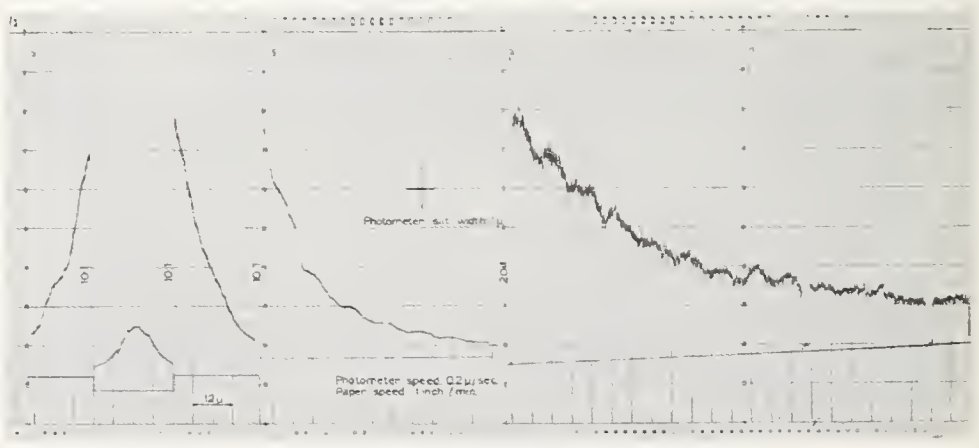


FIGURE 12.12. *Original record.*

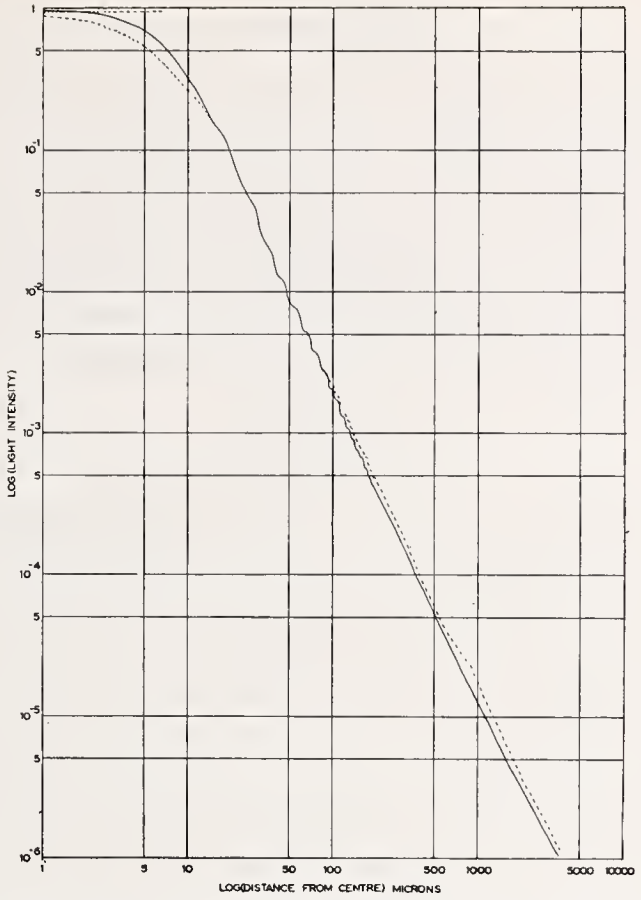


FIGURE 12.13. *Intensity distribution far from the center.*
Solid curve, fine-polished glass; dotted curve, rough-polished glass.

such fine structure of wavefronts, and, on the other side, the troubles in some instruments due to scattered light for the reason mentioned. Illustrative examples to be studied are the front lenses of microscopes, and spectroscopic instruments, where this type of stray light is often known as *foot light.* A recent investigation with phase contrast on such types of irregularities in the wavefront [9] should be supplemented by recordings of the intensity distribution as performed here.

The experimental procedure used here is a means for studying the real performance of a lens system comparing the intensity with that predicted from theoretical calculations. The curves shown here have not yet been analyzed in this way, but the slope of the curves, figure 12.13, seems to agree with the corresponding functions in this region (W, more than one hundred diffraction units) remote from the line center.

Note added in proof (Oct. 1953): A subsequent development of this scanner is reported in P. Lindberg, Measurements of Contrast Transmission Characteristics in Optical Image Formation, Optica Acta (in press).

518-743 O - 73 - 26

References

[1] E. Ingelstam and P. J. Lindberg, Rep. Lab. Optics; Roy. Inst. Technol. (Stockholm) No. 9 (1951).
[2] E. Ingelstam and P. J. Lindberg, J. Opt. Soc. Am. **41,** 346 (1951).
[3] L. A. Jones and R. N. Wolfe, J. Opt. Soc. Am. **35,** 559 (1945).
[4] W. Herriott, J. Opt. Soc. Am. **37,** 472 (1947).
[5] J. Strong, Modern Physical Laboratory Practice (Blackie & Son, Ltd., London, 1938).
[6] C. M. Sparrow, Astrophys. J. **44,** 76 (1916).
[7] A. Gray, G. B. Mathews, and T. M. MacRobert, A Treatise of the Bessel Functions and their Applications to Physics (McMillan Co., London, 1939).
[8] G. Lansraux, Rev. Opt. **26,** No. 24, 278 (1947).
[9] E. Ingelstam and E. Djurle, Arkiv Fysik **4,** No. 27 (1952).

13. Measurements of Energy Distribution in Optical Images

By R. E. Hopkins,[1] Howard Kerr,[1] Thomas Lauroesch,[1,2] and Vance Carpenter [1,3]

Introduction

The Photographic Branch of the Signal Corps is constantly faced with the problem of how to write specifications for lenses to be used in their various applications. They were well aware of the shortcomings of the standard methods of specifying a lens' performance, and requested that our laboratory study the problem and recommend testing and specification procedures. In contrast to the Air Force they were primarily interested in lenses of moderate focal length ranging from 1 to 20 inches.

The first part of this study included an investigation of the literature on the subject. One point that seemed to crystallize out of the study was that most of the work had been done by users of equipment. The tests performed were usually designed to measure the final performance of the lens, and most of the interest was centered around the validity of the testing procedures.

During and after the second war several workers began describing methods for testing the lens itself [1,2,3].[4] Jones and Wolff [4] described a method for determining the distribution of energy in an image with a photographic process. Herriott [1] described a photoelectric method for measuring the intensity contour of a line image. Hansen [5] performed an important basic experiment. He studied an Elmar 8-cm focal-length lens at several apertures by photographing a periodic grating and a transparency of a building. He found that the lens resolved the best when used at the full aperture of f/3.5. On the other hand, observers when asked to pick the sharpest picture of a building selected pictures taken at f/6.3. The spherical aberration of the lens was measured on an interferometer, and showed that the minimum diameter of the blur circle at f/6.3 was less than for f/4.5.

Prior to reading this article Dr. Robert Wolfe at Eastman Kodak suggested to us during a conference a similar approach to the problem of image evaluation.

S. Huber [6] wrote a paper in 1943 that described a logical approach. He designed and constructed a lens with known spherical aberration, and calculated the distribution of light in the image. He then took photographs of a test chart and correlated the results. He found that the resolving power could be related to the diameter of the central core of the image that contained 25 percent of the total light.

[1] Institute of Optics, University of Rochester, Rochester, N. Y.
[2] Now located with Eastman Kodak Co.
[3] Now located with American Optical Co.
[4] Figures in brackets indicate the literature references on p. 198.

It was clear from the trend of papers that more information on the image energy distribution of lenses was needed. It was also evident that purely physical measurements on the image would not be sufficient. A physical measurement would answer many of the problems of resolution and contrast but, in order to set up tolerances on energy concentration, experiments like Huber's and Hansen's would be needed.

With this background it was decided that the problem would be studied as follows. (1) A purely physical method for measuring the characteristics of the lens would be developed; (2) studies similar to Huber's and Hansen's would be made to relate image characteristics with performance on film.

Another point that came out of the literature was that many of the papers on the subject dealt with the problem of how to treat unsymmetrical off-axis images. If one accepts the concept that the object is made up of a series of points and that it is the function of the lens to form a faithful image of the point, the problem then is to measure how well the lens does form a point image and no particular weight should be given to structure of the image in anything but a radial direction. Therefore, it was decided to measure only the radial energy distribution. This approach is probably valid only if the objects being photographed have random orientation of all edges. It is obvious that if one intends to photograph railroad ties or picket fences then radial and azimuthal distribution should be known.

Methods for Measuring Energy Distribution

Interferometer

The first approach was to use the interferometer. The instrument and testing procedures have been adequately described in the literature [7, 8]. In general, the interferometer has been rejected as a testing instrument on two counts, which are believed to be not entirely justified.

Several authors have described detailed methods for analyzing an interferogram to determine the coefficients of the various Seidel third-order and high-order coefficients. These methods are difficult and time consuming. However, if one merely uses the interferogram to determine the radial energy distribution the interpretation is simple and straightforward. For example, figure 13.1 shows the interferometer pattern of an f/2 photographic lens. The radial energy distribution of energy can be determined by marking off equal slope contours. Measuring the areas included within the contour lines and dividing them by the total area gives a radial energy distribution plot. This can be done accurately if necessary, but for many purposes merely a rough estimate of area is sufficient. For example, figure 13.1 shows at a glance that there is considerable vignetting at 20° off-axis and that only about 30 percent of the energy is well concentrated into a spot.[5]

The other complaint about the interferometer is that it is too sensitive. This is not justified if one attempts to gather no more information than other less sensitive devices like a lens bench test. The interferometer is sensitive because it is capable of giving detailed

[5] Coleman [8] made the valuable suggestion of using the area of the largest inscribed circle free from fringes as a measure of quality.

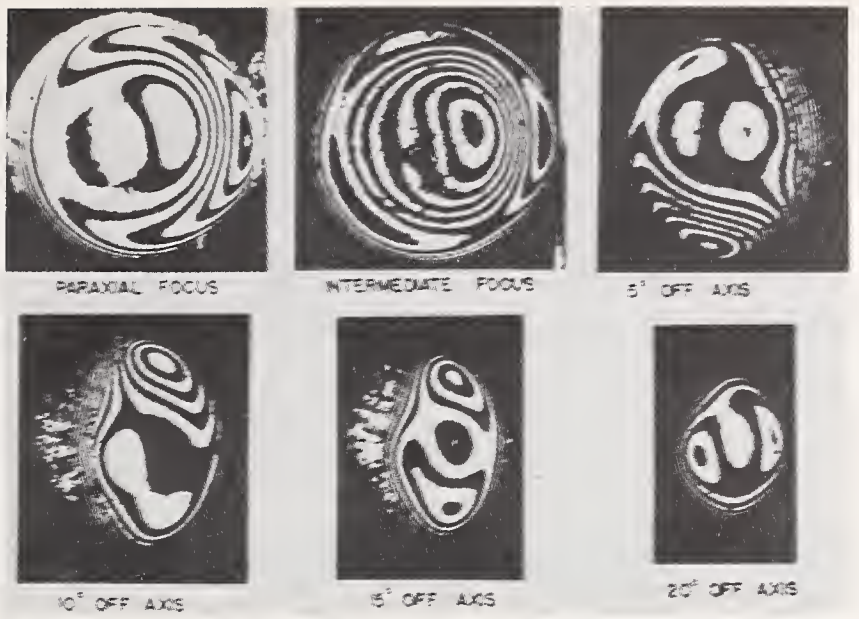


FIGURE 13.1. *Interferogram of F/2–50-mm focal-length photographic lens.*
Wave length 5461 A.

information much of which is not needed for analyzing a photographic lens.

The interferometer is the perfect instrument from a lens designer's viewpoint. It dissects the image for him. It gives him a quantitative picture of the wave front. It gives him a feel for the design.

It does have the following disadvantages: (1) It does not indicate directly the effect of diffraction; (2) the interferogram does not show the effects of scattered light.

Further studies relating interferograms and the diffraction effects are needed in order to make the interferometer more useful.

Several measurements of energy distribution were made on lenses with the interferometer. Reasonably good agreement was obtained between this method and a photoelectric method to be described. However, the measurements on the interferometer and the photoelectric device were made on separate pieces of equipment and it was not possible to be sure of the focus accurately enough to establish whether the differences were real or due to diffraction.

Photoelectric Method

There are several methods described in the literature for measuring the performance of a lens photoelectrically [1, 2]. Most of the methods described use a slit source and the image formed by the lens is scanned across the narrow dimension. If one accepts the concept that the radial energy distribution is a measure of the performance of the lens, then the image of a point source may be photometered by allowing the light to pass through successively larger circular apertures.

The experimental arrangement to be described measures the radial energy distribution. A schematic arrangement of parts is shown in

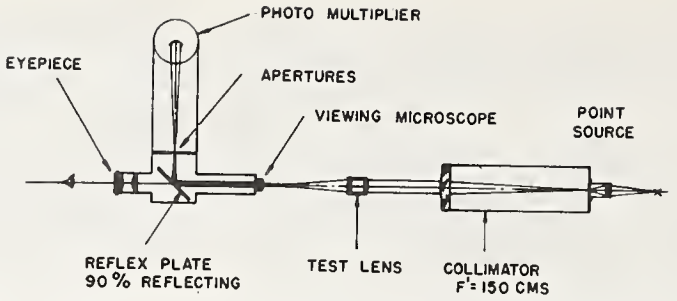


FIGURE 13.2. *Schematic arrangement for photoelectric testing of photographic lenses.*

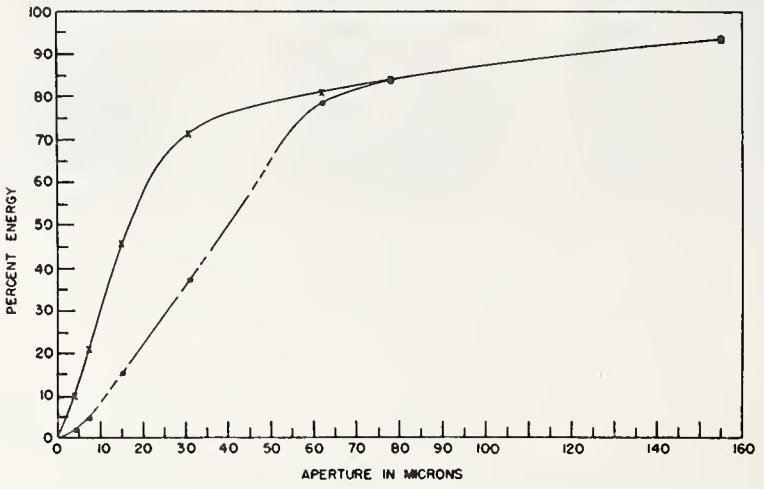


FIGURE 13.3. *Method of plotting energy distribution*

Lens name, Sonnar 50 mm; focal number, F/2; focal setting, best visual; filter, Wratten 58. × —, 0°; ● - - - , 15°.

figure 13.2. The lens to be tested is set up to form an image of a point source, imaged at infinity by a collimator. The image is then reimaged with a microscope objective onto an aperture slide. The aperture slide contains a series of small apertures that may be slid into the image plane. The light that passes through the aperture is received by a photomultiplier. A measurement of the energy distribution is made by sliding in the smallest aperture. The current through the photomultiplier is recorded. Larger and larger apertures are then slid into place until there is no further increase in photomultiplier current. The percentage of flux for any particular aperture is then obtained by dividing the current for each aperture by the current for the largest aperture. The data obtained for a lens is then plotted as shown in figure 13.3. The interpretation of this plot for the 0° image is as follows: 30 percent of the total energy in the image of a point source is contained within 10 microns; 80 percent is confined to 60 microns. Note that these curves are not intensity-distribution curves. In order to find the intensity at any radius it is necessary to differentiate the curve.

Experimental Details

1. The photomultiplier used was an RCA 931A. This tube is definitely red insensitive, but no attempt was made to balance the

photomultiplier with photographic response. Most of the measurements reported in this paper were made using filtered light. The light source was a 50-cp automobile-headlight bulb.

2. The photomultiplier was battery operated and was followed by a low-noise a-c amplifier. The final signal was measured with a Ballantine volt-meter. The light was chopped at 900 c/sec.

3. Special Bausch and Lomb [6] coated microscope objectives were used to view the image. These objectives were painted dull black and special care was taken in blackening the mount and spacing rings in order to eliminate scattered light.

4. Ground glass was used between the apertures and the photomultiplier.

5. The size of the holes and their apparent sizes when in use by the microscope objectives is shown in table 13.1.

6. The reflex viewer was used to provide a view of the image being photometered.

TABLE 13.1. *Size of holes in aperture plate versus size appearing in focal plane of lens being tested*

Apparent size with —	Size of hole (μ)						
	100	200	400	800	1600	2000	4000
16-mm objective (μ)	7.8	15.5	31.0	62	124	155	310
8-mm objective (μ)	3.9	7.8	15.5	31.0	62	78	155

Test on Reliability of Equipment

The linearity of the amplifier was checked electrically by putting in a calibrated electrical signal from a Jackson signal generator. The overall equipment was checked as a photometer in the following manner. A well corrected telescope objective was used as a test lens. The largest size aperture was then put in place. A green Wratten No. 58 filter was inserted in front of the point source. Neutral filters were then also inserted in front of the point source and their transmissions were measured with the equipment. The filters were calibrated previously using a Bausch and Lomb Martens photometer. The results are shown in table 13.2. The errors included in the table are the average deviation from the mean of four readings. The

TABLE 13.2. *Transmission measurement made with photoelectric equipment and compared with measurements made on a Martens polarizing photometer*

Martens polarizing photometer	Photoelectric photometer
%	%
78.2 ± 2	73.9 ± 0.7
62 ± 1	62.9 ± 0.4
46 ± 1	44.7 ± 1
36.6 ± 0.7	38.6 ± 1
28.9 ± 0.6	28.3 ± 0.1
25.3 ± 0.5	26.0 ± 0.4
18.5 ± 0.4	18.4 ± 0.3
14.9 ± 0.3	15.2 ± 0.4
9.1 ± 0.2	10.4 ± 0.6

[6] Recommended for our use by James Benford.

greatest discrepancy occurs at the high transmission values. The error may be due either to the Martens or the photoelectric reading. The largest discrepancy between the two methods amounts to 4 percent.

The complete reflex viewer and photoelectric equipment was checked by measuring the energy distribution of the image formed by a perfect lens. A well corrected telescope objective stopped to f/15 was set up directly on axis. Several attempts were made to measure the image from a point source and when compared with the Airy disk the measured image was always larger. Several refinements were made. The lens was centered more accurately, the tungsten source was replaced with a mercury source with a filter isolating the 5461 A line. The energy distribution was measured at several positions of focus. The pinhole was made as small as feasible. With all these refinements the agreement with the theoretical curve improved. The best agreement found is shown plotted in figure 13.4. The results are tabulated in table 13.3. The curve is made up of points using a 16-mm and an 8-mm objective. The agreement with the theoretical curve is only fair but we believe the difference is not due to an error in measurement. There are two reasons why the experimental curve could be broader than the theoretical curve. (1) There is scattered light. The lens is an air-spaced doublet. The surfaces were not tested but they may be imperfectly polished or imperfectly cleaned. (2) The air between the source and lens was not completely homogeneous. One could observe slight image boiling. This effect would broaden the image.

Attempting to fit the measured curve with the theoretical curve actually strengthened our confidence in the reliability of the equipment. Each improvement made in technique showed up as a better check. In order to get still better agreement it would be uecessary to check the lens surfaces and probably repolish them. Also steps would have to be taken to ensure perfectly homogeneous air between the source and the lens.

Since we were interested in testing lenses and it was not known at the time what accuracy of measurement was needed it was decided not to check this point any further. As can be seen, there are differ-

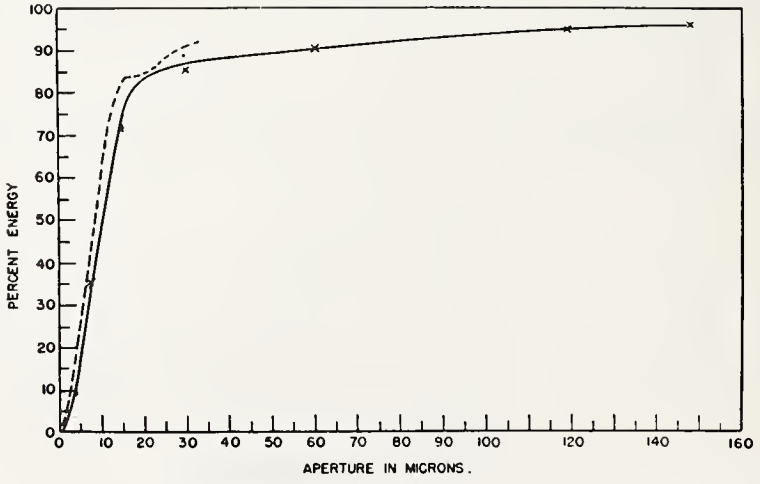


FIGURE 13.4. *Energy-distribution curve for an f/15 telescope objective.*

Lens name, Telescope Objective; focal number, 15; focal setting, best visual; filter, Mercury 5461. - - -, Theoretical; ———, experimental; x, 16-mm objective; ●, 8-mm objective.

TABLE 13.3. *Energy content for a perfect telescope objective*

The theoretical energy content for ξ point source on axis is compared with the measured energy content for a well-corrected F/15 (F′=35 cm) air spaced telescope objective. Monochromatic 5461 A light was used for this test.

Spot diameter	Energy		
	Airy disk theoretical	Measured with 8-mm objective	Measured with 16-mm objective
μ	%	%	%
3.75	14.5	9.6 ± 6.04	----------
7.50	47.2	35 ± .6	35 ± 1.5
15.0	83.0	73 ± .9	71 ± 1.5
30.1	90.8	89 ± 1.3	85 ± 1.0
60.0	94.0	92 ± 3.1	90 ± 2.0

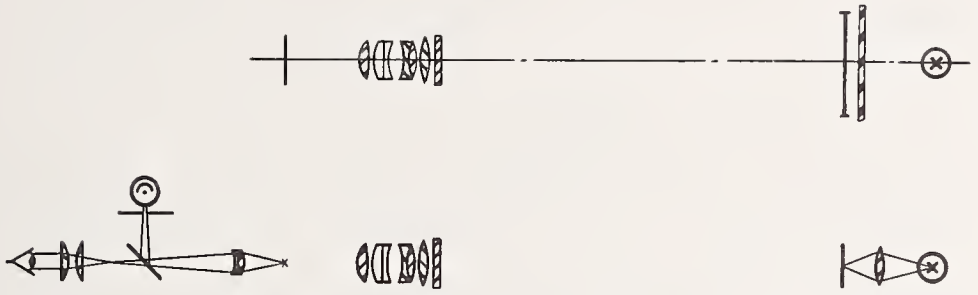


FIGURE 13.5. *Experimental setup for image evaluation.*

ences between the theoretical curve and the measured curve as large as 12 percent, but it is our belief that the difference is not entirely an error.

Image Evaluation Experiment

Following the suggestion of Wolff, Huber and Hansen the following experiment was performed to relate the energy-distribution curve with pictures taken on film. The basic experiment performed and the equipment is illustrated in figure 13.5.

The point source image of a 7-inch Aero Ektar lens was photometered. The lens was then shifted to a new position and a transparency of a scene was photographed. The film plane was located in the same position as viewed by the photometer. The scene that was photographed is shown in figure 13.6. Finally, on a separate film, a photograph of a $\sqrt[6]{2}$ line test chart was taken. In this manner the energy distribution in the image, the resolving power, and a picture of a scene were obtained for several focal settings and f-number.

Experimental Details

1. The 7-inch Aero Ektar was used because it provided a large variation of spherical aberration with change in f-number.

2. The scene was placed 8 feet from the lens, and it subtended 5°. The image of the scene on the film was 13 mm by 11 mm. The long focal length was used to obtain a reasonable size image on the film without using the lens more than 2½° off-axis. The image quality over the entire picture was then uniform.

3. The Aero Ektar was mounted on a Kine Exacta camera body. A special stiffened focal plane was built into the camera. For each picture the film was pressed against the film plane by a pressure plate that could be screwed down tight from the outside of the camera.

4. The lens was always focused at a definite focus position. The settings used were 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5, and 11.5. The distance between each setting was 80μ.

5. Eastman Super XX film was used for all the tests. The Aero Ektar was provided with an accurately polished 4-mm thick Corning Sextant green filter. This filter has a peak transmission at 5200 A and transmits 12½ percent at 4900 A and 5580 A. All of the picture-evaluation data was taken in green light with this filter.

6. A Wild[7], $\sqrt[6]{2}$ light lines on a dark background, high-contrast resolution chart was used for resolution data. A 1-inch clear area was used as a control area. The films were always developed to a gamma of .8 $\pm$.1. For each focal setting several exposures were made. The resolution was read by at least two observers by viewing the negatives with a 40X microscope.

7. The scene shown in figure 13.6 was photographed with an 8 by 11 camera using a Bausch & Lomb 12-inch Protar. The positive transparency has a range of densities from .3 to 1.4 and a gamma of .8. U. S. A. A. F. charts were placed alongside both sides of the print. These were used as a check on the enlarging step. On the top and bottom of the scene grey scales were used to control the exposure and development. All the negatives were exposed and developed in

[7] Supplied by Alan Murray, Bausch & Lomb Optical Co.

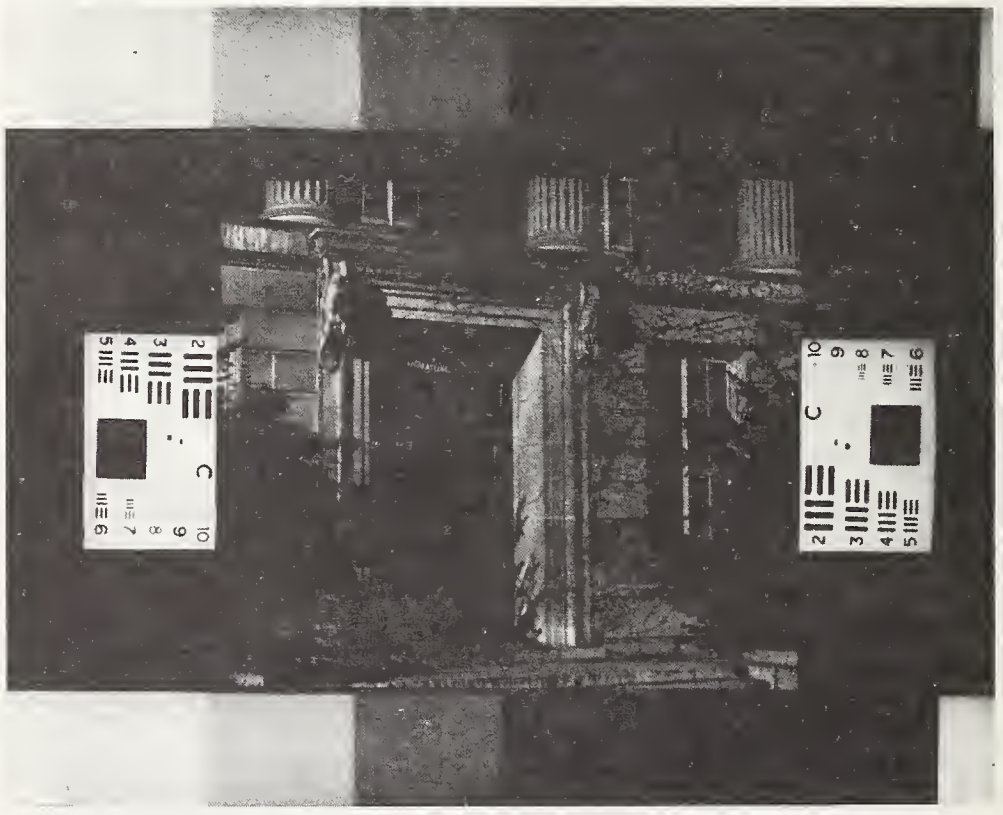


FIGURE 13.6. *Scene photographed for image evaluation.*

microdol to a gamma of .8 ±.03 with a maximum density of 1.8 ±.02.

The negatives were then enlarged eight times in a Leitz condenser enlarger, using a special lens loaned by the Signal Corps. This lens is a Zeiss Biotar 10-cm focal length f/4, stopped down to f/6.3. The lens is nearly perfect at this aperture ratio. The white-light image to the first bright ring measured to be 7μ.

The prints were made on Kodabromide F.3. Exposure and development were controlled carefully to make the reproduced grey scales alike. The grey scales were measured on the Capstaff reflection densitometer. Any picture that had a density step differing by more than ±.1 in the highlights and ±.02 in the shadows was rejected. In the first series of pictures made (f/4 set no. 1) the tolerance was ±.02 but the tolerance was relaxed in order to obtain some results. It was extremely difficult to make prints to fall within this tolerance by using carefully controlled methods. The method finally adopted was to make several prints and then select the ones that fell within tolerance.

An extensive series of tests were performed to determine the loss in resolution from the enlarging process. With a negative resolving power of 37 to 40 lines per mm the prints on the average resolved 30 to 32 lines per mm. With a negative resolving power of 20 lines/mm there was less than 1-line/mm loss in resolution due to the enlarging process.

The above series of experiments provided the following data for each position of focus of the lens: (1) Energy distribution curves; (2) lines/mm resolution for several exposures; (3) glossy prints of a scene.

Energy-distribution curves. Figures 13.7, 13.8, and 13.9 show the energy-distribution curves for aperture ratios of f/2.5, f/4, and f/5.6.

Resolving-power data. The resolving-power data is summarized in figure 13.10. The measured resolving power indicated by the dotted lines is plotted against positions of focus. The first results showed that the resolving power did not change with exposure if the density of the control area remained between 1.4 and 2.0. The resolving

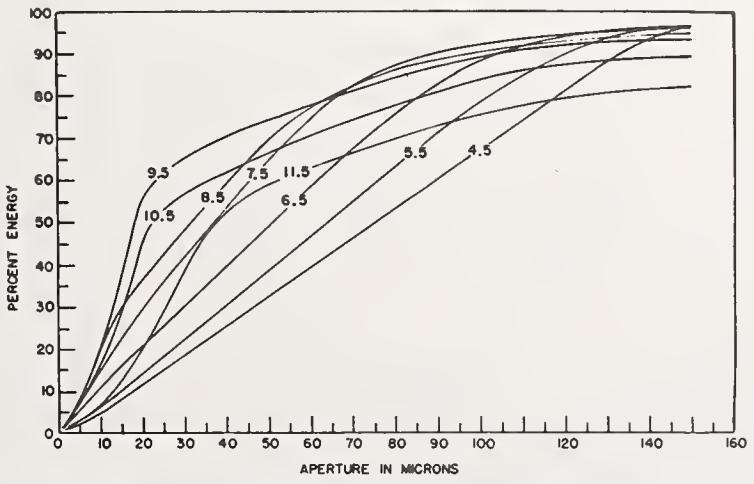


FIGURE 13.7. *Energy-distribution curves for Aero Ektar at f/2.5.*
Lens name, Aero Ektar; focal numbers, 2.5; focal setting, 4.5 to 11.5; filter, Corning Green.

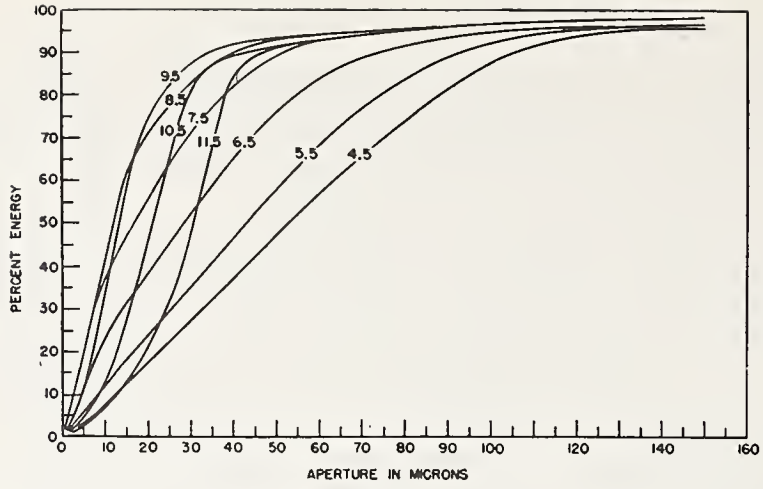


FIGURE 13.8. *Energy-distribution curves for Aero Ektar at f/4.*
Lens name, Aero Ektar; focal number, 4; focal setting, 4.5 to 11.5; filter, Sextant Green.

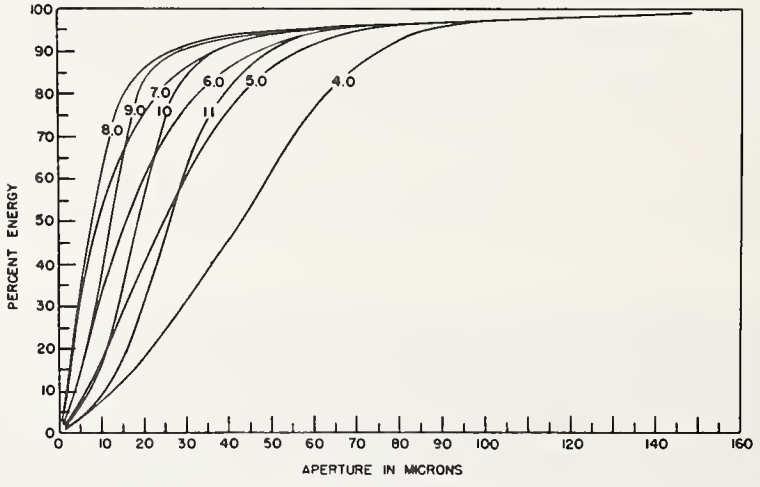


FIGURE 13.9. *Energy-distribution curves for Aero Ektar at f/5.6.*
Lens name, Aero Ektar; focal number, 5.6; focal setting, 4 to 11; filter, Corning Sextant Green.

powers were, therefore, averaged if they fell within this range. At least six independent film strips were used. It is well known that reading resolving power is difficult. The result that one obtains depends on the criterion used. The two observers who read the data agreed on a criterion and then they read resolving power to within an average deviation of not more than 3 lines/mm. It is our opinion that the *R. P.* figure given is subject to a large error, as much as 10 or 15 lines depending on the criterion used. The resolving-power data does, however, offer an interesting relative measure of the image quality.

The solid curves in figure 13.10 are resolving powers calculated from a formula based on the energy-distribution curves. At each focal setting the size of spot that contains 30 percent of the energy was determined. These spot sizes were used in the following formula to predict resolving power.

$$R.P.=605/(d_i^2+d_f^2)^{\frac{1}{2}}, \qquad \text{(A)}$$

where

d_l=spot size for 30-percent energy in microns
d_f=spot size for film (10 microns for Super XX).

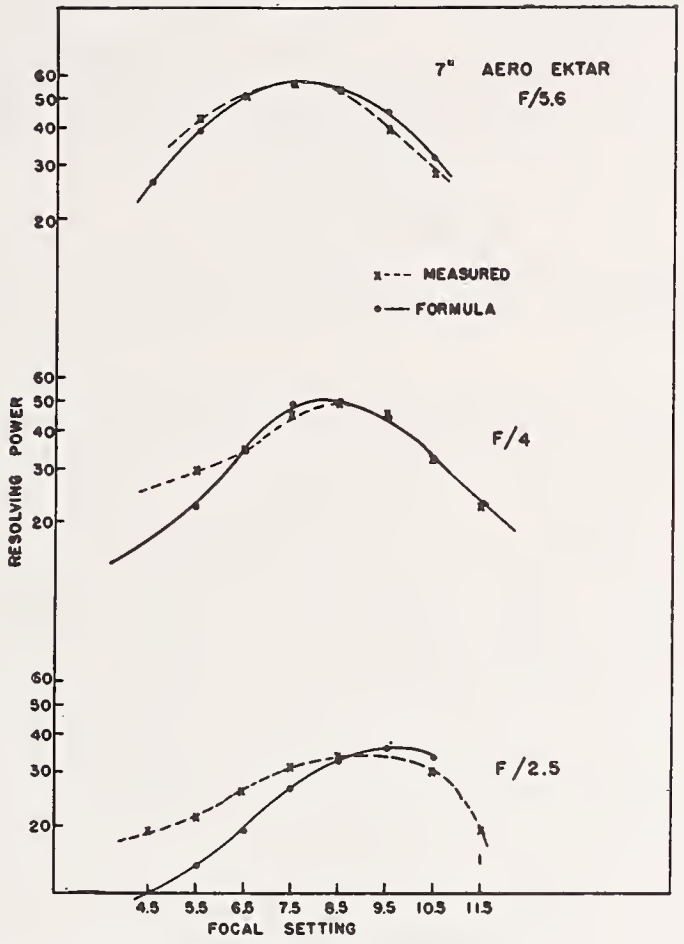


FIGURE 13.10. *Resolving power for several focus settings of the 7 in. Aero Ekton at f/2.5, f/4, f/5.6.*

The formula is empirical and is an oversimplification. It does, however, indicate that the resolving power is determined by a very small percentage of the energy. The agreement is good at the large focal settings but discrepancies show up at the small focal settings. At focal settings 4.5 and 5.5 the images contain small cores of light and a large flare. This small core of light accounts for the high resolution. This type of distribution is due to the spherical aberration in the lens. The effect is most pronounced at f/2.5, where the spherical aberration is considerable. Inspection of the energy-distribution curves indicates that less than 30 percent of the total energy is contained within the small core. If the above formula were to predict the high resolution shown at focal settings 4.5 and 5.5, it would be necessary to use a spot size that contained less than 30 percent of the energy, possibly as low as 10 percent.[8] It would then also be necessary to combine this spot size with one at a higher energy level in order to fit the curve at the higher focal settings. No attempt was made to obtain

[8] We should recall that Huber [6] found that 25 percent of the energy determined the resolving power.

better fittings for several reasons. The curves measured are not accurately determined at such extremely small images. At the 10-percent level, for example, the spot sizes run between 2μ and 14μ in diameter. There is also some question that the limiting resolving power is significant. Finally, much more data would be needed to obtain a reliable empirical formula.

Formula A is clearly incomplete because it limits the resolution to 60.5 lines/mm for Super XX. Resolution higher than this is possible. A much simpler formula using the first power is as follows:

$$R.P.=\frac{700}{d_e+d_f}. \tag{B}$$

A comparison between these two formulas is shown in tables 13.4,a, 13.4,b, and 13.4,c.

TABLE 13.4. *Comparison of calculated resolving powers, using formulas A and B*

d_1	A	B	Measured	d_1	A	B	Measured
(a) *f*/5.6				(b) *f*/4			
μ	*Lines/mm*	*Lines/mm*	*Lines/mm*	μ	*Lines/mm*	*Lines/mm*	*Lines/mm*
21	26	25	--------	33	17.5	17.5	--------
12	38.8	36.8	43.2	25	22.5	21.9	29.4
6	51.7	53.8	51.5	14	35.0	33.0	34.6
4	56.2	63.6	55.9	7	49.6	50.0	44.8
6	51.7	53.8	54.1	7	49.6	50.0	49.2
10.5	42.7	40.0	38.3	9	44.8	43.8	44.8
16	32	30.4	28.1	15.5	32	31.1	32.4
				24.0	23.3	22.6	22.5

d_1	A	B	Measured
(c) *f*/2.5			
μ	*Lines/mm*	*Lines/mm*	*Lines/mm*
47	12	12.2	19
40	14.7	14	21
30	19.1	17.5	26
20	27	23.3	31
15	33.6	28.0	34
13	36.7	30.4	34
15	33.6	28.0	30

It should be pointed out that these formulae do give a lens designer some idea of the relation between resolving power and image size. The examples are limited but it is the type of data a lens designer needs to know, for how else is he to design a lens to meet specifications that are invariably written in lines/mm. The formula gives a pessimistic answer that may be advantageous.

Picture evaluation. The prints of the scene were observed by two of the authors. Several combinations were obviously different in sharpness. The remaining combinations that were difficult to distinguish between were selected for inspection by several observers. The observers were asked to compare the pictures by pairs and to pick the picture that they believed contained the more information. Tables 13.5, 13.6, and 13.7 show the results of this survey. The first

TABLE 13.5. *Image evaluation data at f/2.5*

Picture evaluation F/2.5

Focal Setting	*R. P.*	d_1	V. B. T. Set 1	Comments
	Lines/mm	μ		
9.5	34	13	5	Very difficult to differentiate.
8.5	34	15	2	No significant difference.
9.5	34	13	6	Significant difference.
10.5	30	15	1	
8.5	34	15	6	Significant difference.
7.5	31	20	1	
8.5	34	15	6	Significant difference due to less than 30-percent energy concentration.
10.5	30	15	1	
7.5	31	20	3	No significant difference. Depends on observer's method of viewing.
10.5	30	15	4	
7.5	31	20	7	Significant difference.
11.5	19	26	0	
6.5	26	30	7	Significant difference. Reason not clear.
11.5	19	26	0	

TABLE 13.6. *Image evaluation data at f/4*

Picture evaluation F/4

Focal settings	*R. P.*	d_1	V. B. T.		Comments
			Set 1	Set 2	
	Lines/mm	μ			
8.5	49	7	8	3	
9.5	45	9	1	5	
8.5	49	7	8	4	All these pictures are similar. There is a possibility that 8.5 is better than 7.5.
7.5	45	7	1	3	
9.5	45	9	2	4	
7.5	45	7	7	3	
10.5	32	15.5	2	5	Difference less than errors in printing.
6.5	35	14.0	7	2	
10.5	32	15.5	7	6	10.5 significantly better than 5.5.
5.5	29	25.0	1	1	
11.5	23	24.0	6	3	No preference, however, 5.5 definitely has more detail.
5.5	29	25.0	3	4	

TABLE 13.7. *Image evaluation data at f/5.6*

Picture evaluation F/5.6

Focal settings	*R. P.*	d_1	V. B. T. Set 1	Comments
	Lines/mm	μ		
8	56	4	6	Significant difference.
7	55	5	0	
9	50	8	0	Significant difference.
7	55	5	6	
10	38	14	3	No significant difference indicating influence of low energy concentration.
5	36	15	3	
6	47	9	6	Significant difference.
10	38	14	0	
6	47	9	4	No significant difference.
9	50	8	2	

column is the focal setting. The second column gives the measured resolving power on the negative. The third row contains the 30-percent energy spot size. The fourth row contains the "Votes Better Than" for the final prints. An arbitrary value of 75 percent of the votes was called a significant difference.

The data at f/4 is the most complete. Print set No. 1 was made with the closest tolerance in tone quality. A complete duplicate set of prints No. 2 were made to check the enlarging process.

Summary of Results

The survey may be summarized briefly as follows:

1. There was only one case of a significant difference where there was disagreement between lines/mm and the 30-percent spot-size criterion. This occurred at f/2.5 for focal settings 6.5 and 11.5. The curves for the two focal settings cross below 25 percent of the energy. The lines/mm criterion agreed with the picture valuation. This indicates that less than 25 percent of the energy is determining the resolution and picture quality. This value may be lowered in this case because the curves cross again at 65-percent energy content.

2. In one case (f/2.5) there was a significant difference between focal settings 8.5 and 10.5 in evaluation, no difference in the 30-percent spot size but a difference of 34 and 30 lines per mm. This also indicates that the energy concentration below 30 percent is influencing the results.

3. There are cases, for example, at f/4, focal settings 11.5 and 5.5 where the observers divided their opinion, and the resolving power and the 30-percent spot-size criterion were opposite. Careful observation of the prints reveal definitely more detail in some parts of the picture with the high resolution. If the observers viewed the pictures from a distance they picked picture 11.5. However, if they were asked to view the pictures closely with a low-power magnifier they would agree 5.5 had more detail. This observation agrees with what Hansen [5] pointed out. His observers picked the pictures taken at f/6.3 even though the pictures at f/4.5 resolved more. The observers either did not look for the fine detail, were not able to see it, or there was a small percentage of fine detail in the picture and most of the observers did not notice it.

4. The survey does show that very slight differences in resolving power are significant in the final picture. The pictures also show that as small as a 20-percent difference in spot-size results in a significant difference in performance. No attempt was made to compare the different f-number sets. Table 13.8 shows comparisons between resolving powers for different focal settings and f-numbers when the spot sizes were common. The agreement is fair but disagreement does occur when spot sizes are similar and focal settings are on opposite

TABLE 13.8. *Comparison of resolving powers for equal spot sizes obtained with different conditions*

Spot size	*R. P.*	Condition	
		f/No.	Focal setting
9	47	f/5.6	6
9	45	f/4	9.5
14	38	f/5.6	10
14	35	f/4	6.5
15	36	f/5.6	5.0
15	34	f/2.5	8.5
15	30	f/2.5	10.5
15.5	32	f/4	10.5

sides of focus. The pictures taken at the best setting at f/5.6 were definitely better than those obtained at f/4, which in turn were better than at f/2.5. . The pictures also show very clearly the focus shift between f/2.5 and f/4. There is a focal shift from 9.5 to 8, which is approximately 120μ. This particular lens should be focused for best focus at f/2.5. It is surprising how well the lens performs at f/2.5. The image of a point source viewed visually is very large. The flare extends out to 200μ.

It is clear that the energy-concentration requirements of a lens depend on viewing conditions and interests of the observer. One might generalize in the following manner. If a lens is required to resolve fine detail then between 20 and 30 percent of the energy must be concentrated to within the appropriate spot size. On the other hand, if the prints are to be viewed so that this fine detail will never be resolved by the observers then it would be better to allow a larger spot size and put more energy within the spot size. Two opposite cases are illustrated between the interests of a photointerpreter and an amateur photographer.

The photointerpreter will view the prints with a magnifier looking for every bit of information available. A lens for him then should provide as much resolving power as possible, and the designer should attempt to design a lens to put 30 percent of the energy to within as small a spot as feasible.

The amateur photographer may take his pictures on a $3\frac{1}{4} \times 4\frac{1}{4}$ negative, make a contact print, and then view them from 10 inches. From this distance he will probably not notice any detail less than 14 lines/mm. The lens, therefore, needs to have a spot size 40μ (Formula B) in diameter. The designer has the alternative of placing more than 30 percent of the energy in this spot or simplifying the design and allowing just 30 percent.

It is agreed that one can be carried away by oversimplified formulae. The data obtained so far is scant. More data on more different types of energy distributions should be studied. The data is confined

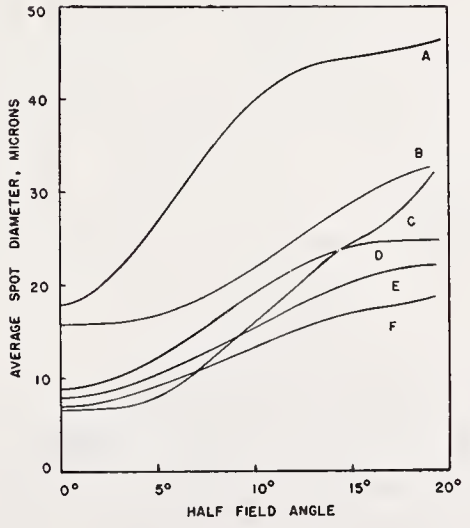


FIGURE 13.11. *Plot of area weight-average spot size for several lenses.*

Weighted-mean spot diameter of 50-mm objectives (50-percent concentration).
A, Ektar f/1.9; B, Sonnar f/2; C, Ektar f/4; D, Elmar f/3.5; E, Sonnar f/4; F, Elmar f/6.3.

to Super XX and green light. Similar studies should be made with red and blue light separately and then added together.

It is our belief, however, that much of the concern about the accuracy of resolution data is splitting hairs. In actual practice a lens must cover an appreciable field of view. Over this field of view the spot size for the 30-percent energy concentration may vary by more than a factor of 10. It would appear, therefore, that a lens can be appraised most profitably by analyzing the energy concentration for several points in the field for several positions of focus, and to place the film in the proper position to obtain the minimum area–weighted-average spot size. Figure 13.11 contains plots of area–weighted-average spot sizes for several lenses. This type of analysis is extremely lengthy and a simple criterion like a 30-percent energy spot size is needed. We believe a simple formula, even though an oversimplification, is what optical designers need now. With it they can make improvements in overall field performance of existing lenses.

Conclusions

The above experiments show that the energy-distribution curve does indicate the performance of a lens. If the requirements on a lens call for maximum resolution, as little as 25 or 30 percent of the energy may be concentrated into as small a spot as possible.

It is recommended that lenses should be tested independent of film by measuring the energy distribution in the image of a point source. It is recommended that the lens should be checked photographically. Line test charts are satisfactory if backed up with energy-distribution curves. Anomalies may show up but they can be explained and advance the knowledge of the subject.

It is also urged that the lens designer's interests should be considered, by specifying lenses by their energy distributions. A single spot size may be sufficient; two spot sizes would probably be even better. Almost all military specifications are written today in terms of lines/mm resolution. There is so little data available connecting lines/mm and lens characteristics that the lens designer can go only on his past experience and theories.

The photoelectric method described in this paper is probably not suitable for production testing of lenses. The task of making all the required measurements is formidable. The method suggested by Schade [2] may offer a more convenient method of measurement, but we urge that specifications should be written in terms of energy concentration. Schade's response factors can be converted to energy concentration.

References

[1] W. Herriott, J. Opt. Soc. Am. **37,** 472 (1947).
[2] O. H. Schade, RCA Publication ST–3BD.
[3] J. D. Hunter, Proc. Roy. Soc. [A] **82,** 307 (1909).
[4] L. A. Jones and R. N. Wolfe, J. Opt. Soc. Am. **35,** 559–569 (1945).
[5] G. Hansen, Zeiss Nach. Heft **6,** 129–137 (Nov. 1942).
[6] S. Huber, Z. Instrumentenk. **63,** 333–341, 369–383 (1943).
[7] F. Twyman, Trans. Opt. Soc. **22,** 174 (1920–21).
[8] H. Coleman, J. Opt. Soc. Am. **37,** 671 (1947)

Discussion

Dr. K. V. Pestrecov, Bausch & Lomb Optical Co., Rochester, N. Y.: Dr. Hopkins, I really appreciate what you said at the very end about too much emphasis put on resolution and perhaps some other test could give you adequate results. But one who gets involved in these data comes to a conclusion because of oversimplification. I would like to challenge this statement. After all, if you have to choose between photographs and one at first looks a little better to you—perhaps because there is a little better contrast—although the resolution is a little lower, the picture with higher resolution still is better when you look for finer detail. I think the interpreter looks for recognition of detail. Resolution is my pet subject although I don't have much time to perform all these excellent experiments.

For example, if you consider a steeple on a church in several photographs, on one of the superficially worse looking photographs the steeple may look a little better. The picture with the highest resolution would have the peak of the steeple clearly defined. The photointerpreter looks for recognition of the steeple. Perhaps in high-resolution photography some corners will be better defined although the overall detail may be hazy; nevertheless, the picture may look a little better and one would have an easier job of recognition. I am not defending my position, I am just questioning. I think there is some work pointing in that direction. It may be recognition of detail but not mathematical detail that is so important. The photograph may actually be better for functional use, aerial interpretation, or something like that. I should like your comments.

Dr. Hopkins: I think there is a great deal in whether you are trying to recognize something or whether you are trying to detect what is there. I think it is splitting hairs. I think basically that if you want to get a lens to satisfy all photointerpreters, you have to put at least 25 percent of the light within as small a spot as is practicable. That, in itself, is theoretical. When you start to design the lens, the minute you get off axis you really have a problem. I am not so sure that anything more refined is necessary at this time. The experience we have had is that if there is high resolution the people can pick up more detail. We would ask them to look at the ivy in this picture and tell whether it had little leaves or little buds on it. They would be able to tell in the high-resolution picture but not in the others.

From the results we had, resolution is very important. Also, I have heard a lot of comments about people saying they can show me a picture that does this or does that, but I haven't seen them. We have a set of pictures you can look at.

Mr. R. V. Shack, National Bureau of Standards, Washington, D. C.: When you get off axis do you recenter your spot?

Dr. Hopkins: Yes, when you have coma it means that you will be pushing the aperture around each time you change it.

Dr. M. Herzberger, Eastman Kodak Co., Rochester, N. Y.: Prof. Hopkins, I think your method and the method that we apply can be very easily coordinated since we have in our calculations a quantity that is equivalent to the light distribution. One can very easily count the number of points within an area, and this determines the area that contains 20 or 25 percent of the light and one

can easily see by inspection where to look for such an area. I think, if you have a lens, it would be quite interesting to try to see whether we could duplicate your results from the data of the calculation.

Dr. Hopkins: It would be very interesting. We actually have the design data on that F/2.5. We have retraced it and checked the energy-distribution curves. There is quite a difference. We don't know whether the lens was not made right or the patent fails to include the latent constructional details.

That actually is a very difficult problem to work out. I would not want to attempt it on anything more than a simple doublet.

Dr. Pestrecov: I would like to state that we have recently found with some of our work that there is a negative correlation between limiting resolving power and what we have been choosing to call picture sharpness. This business of using a resolving-power limit can get you into very serious trouble. It does not tell you about picture sharpness. In a great many cases they do run parallel, but in many significant cases they do not.

Dr. Hopkins: I covered myself by saying that if we had high-resolution data and energy data we would be able to explain this.

Mr. A. H. Katz, Wright Air Development Center, Dayton, Ohio: In connection with the points raised by Dr. Pestrecov and in earlier papers, I notice that a number of people have been gleefully trying to kick the three-line resolution target to death. I want to point out again—and I have done this in other meetings,[1,2] that it has served its purpose well. This purpose, simply stated, is to serially grade lenses in a manner that will correlate with their photograph-making rank. I have yet to be shown that our use of the three-line target in the judging of lenses to be used for aerial photography has led to any error, let alone consistent error.

Now, we have lots of data, most of which is not neat and packaged. The exigencies arising with the working conditions in the Air Force are such as to effectively preclude the careful running of planned experiments. We substitute large numbers of airplane flights and tests, and after a number of years we come to pretty definite conclusions—by statistical osmosis, if you will. We know by now that when we get a lens that performs well in the laboratory (on the much maligned three-line high-contrast target) it will take high-quality photographs in the air on good days as well as bad days. The converse is also true. Laboratory tests enable us to predict the quality of actual aerial photographs. I can't expect much more of a laboratory test. Let us not forget that it is only within the last 10 years that lens performance began to be specified in terms of resolution requirement over the field, and that manufacturers began to use these tests, and it is only within the last couple of years that photointerpreters have begun to hear of lines per millimeter as a measure of performance.

A word about the photointerpreter, whose name has already been taken in vain. He is the consumer of the photographs produced by the lenses, and if he were here today he'd be speechless. Only lately has he really realized that the scale of a photograph (the altitude divided by the focal-length) is not enough to describe what he needs. The fact that some photographs are sharper than others has only com-

---

[1] A. H. Katz, Contributions to the theory and mechanics of photointerpretation from vertical and oblique photographs, Photogrammetric Eng. **16**, 339–386 (June 1950).
[2] Panel Discussion, Cameras, lenses, and calibration, Photogrammetric Eng. **17**, 417–420 (June 1951).

plicated his life, and it is not because of him—as would be logical—that we have specified, procured, tested, and standardized lenses, many of which can hardly be significantly proved. What I mean is that we've done this for the photointerpreter, but not because of any enlightened self-interest on his part.

A few years ago the Air Force issued a statement that we could take photographs of railroad ties from 40,000 feet. It so happens that we had previously resolved railroad ties on photographs from 30,000 feet, but we couldn't find these when asked to produce actual photographs. Well, we had to do it. We did it. It was not extraordinarily difficult, but it was not done the first time we tried. One of the requirements on the photography was that it be of the St. Louis area. We used one of Dr. Baker's 40-inch f/5.0 telephoto lenses, mounted on a standard K–22 camera, and flown in an RF–80 reconnaissance aircraft. No moving film magazine, no special mounts were used. Just straight photographic technique—careful exposure, development, and printing. We found the particular stretch of track later. The contrast—on the ground—between ties and ballast was negligible—the main differences being that of texture between well oiled ties and dirty darkened ballast. Now in this series of photographs was one of the Alcoa aluminum plants in East St. Louis; I identified this as such because on the roof of one of the buildings was painted "ALCOA." However, when I started to show these exceptionally sharp photographs to a photointerpreter here in Washington he identified the aluminum plant before I got the photograph within 2 feet of his eyes. His identification was based on the large residue lake near the plant, a characteristic feature of aluminum production plants, I haven't the vaguest idea how to put this into an equation.

I think many of you here have either cooperated on or seen this standard—Military Standard–150.[3] It represents the best government-industry-designer agreement we have ever had on the subject of definitions and test methods. In this standard we adopted the area–weighted-average resolution of radial and tangential resolution

$$AWAR=\left[\frac{\sum R_i T_i A_i}{A}\right]^{1/2},$$

where R_i and T_i are the radial and tangential resolution in the zone of area A_i, and A is the total area of the field used. Some of us like to ascribe some meaning to this average, although what happened is that we needed a number and we produced the most reasonable and justifiable we could think of.

But how about the difference between two lenses that have the same AWAR? Suppose we have a lens whose resolution varies from 40 lines/mm at the center to 10 lines/mm at the edge of its field, and that the average is 20 lines/mm. Let us say we have another lens that varies from 22 to 18 lines/mm, center to edge, and that also averages 20 lines/mm. Which lens is preferable, assuming uniform probability of occurrence of objects of interest in all parts of the field? I strongly suspect that we really want the second lens.

Other questions dealing with evaluation of two lens systems are even more complicated. Suppose for Air Force purposes, we have,

---

[3] Photographic lenses, Mil-Standard–150 (U. S. Government Printing Office, Oct. 23, 1950) 25¢.

let's say (with our test, which can be disputed the rest of the evening of course) a lens that will average 22 lines/mm with the film used in aerial photography. Let's also suppose it weighs 3 times as much and costs 10 times as much as another lens of the same focal length and angular coverage, but that the second lens averages only 14 lines/mm. (By the way, we never seem to be faced with the converse of this situation. Here's a good problem for the lens designers.) Let's further suppose that the detail we're interested in is caught with the lighter, cheaper, lower resolution lens perhaps but 10 percent of the time it is used, but the better lens records this detail in interpretable form 90 percent of the time it is used. How then compare the value of the lenses?

From this standpoint it is clear that relative figures of 9 to 1, or some function of these figures, is more reasonable than actual cost or resolution figures. I am not throwing these out as answers but am only suggesting that there are large numbers of questions that are unanswered, even after resolution figures or other performance indices are given. Unfortunately, the more we learn of this business, the more questions arise. Progress here seems to lie in the direction of awareness of relevant questions.

I also feel strongly that we are really looking for some measure of the information content of an optical (or photographic) image. This question has not been thoroughly explored, and as far as measuring information content, I feel that all proposals—U. S. Air Force resolution measurements, Howlett's doughnuts, Cobb Charts, Eastman's acutance, etc.—while clearly having some relation to information content, are equally poor measures of information content. This is no criticism of these interesting systems. They weren't supposed to measure information content, and they don't. I think this may well be a fruitful area for careful investigation.

This particular problem is already under investigation by Dr. Macdonald's group at Boston University, where the concepts of communication theory and related techniques are being applied. The main difference between this approach and the ordinary methods of measuring lens performance seems to be in consideration of the overall system—including the ground object, the atmosphere, the camera, its platform, the lens, and its windows, the film, processing system, and the interpreter. But since this is such a large subject it provides me a good place to stop.

CHAIRMAN: Dr. Baker?

DR. J. G. BAKER, Harvard College Observatory and Perkin-Elmer Corp., Norwalk, Conn.: There is one point I meant to emphasize this morning in connection with aerial photography. It is that an increase in focal length appears to be much more effective in improving ground resolution than an increase in the ultimate quality of a lens. In other words, put it this way. Suppose that you are taking aerial photographs under hazy conditions, which is often the case. If you double the focal length, you may realize up to twice the results in ground resolution, whereas if you double the lens quality of the standard lenses, you may realize only a 20-percent gain. On good photographic days, a doubling of the lens quality may realize up to a 60 percent-gain, possibly, which is likely to be a striking difference. The focal length is still the more important parameter.

This does not mean that quality should not be sought after. Where the focal length must be held to a certain value because of space limitations, the high-quality lens will easily out-perform the low-quality lens. Differences of a few lines/mm are discernible and often of importance. But where the focal length can be increased, that is the more certain way to increase ground resolution. Where quality and long focal length are combined, you have the best answer, and this has been the goal of my own work since 1941.

However, at some point in the quality spectrum the law of diminishing returns sets in, and a lens may become too expensive for the percent gain achieved. The designer should go as far as he can in obtaining quality, while employing practicable constructions. But sometimes the demands of a problem are so severe as to require complicated constructions in order to maintain even a fair degree of quality. When I speak of lens quality, I am thinking of microscopic contrast as well as of resolution.

Dr. R. C. Gunter, Clark University, Worcester, Mass.: These remarks are addressed to Dr. Ingelstam and to practically everyone who has spoken before and some who are going to speak subsequently. Some years ago during a rather heated discussion between some medical men, when it didn't seem that we were going to be able to resolve the question as to whether a new theory was or was not applicable to the human brain and various ones were tearing each other apart mentally, I suggested that perhaps the reason for the difficulty and the lack of agreement between the experimentalists and the theorists was similar to that that had existed in objection between the wave theory and the particle theory. The minute I said that Werner turned around like a flash—I was the only physicist I think in the group and that was the only observation I could make—and he said, "The trouble with you is that you neglect the role of the observer."

Now, it has been touched upon several times as I have said, by Dr. Howlett, by Dr. Hopkins, and by many others, that we should be able to separate the observer, the human being, from the equipment. If we cannot at this time define what we mean by the word "energy distribution", perhaps we had better reevaluate some of the fundamental concepts of physics. On the other hand, it is perfectly possible and is indeed present in every day life—as Dr. Hopkins could undoubtedly tell you though I haven't conversed with him on this—if you ask somebody, "Do you see such-and-such in a picture," if there is any possibility of seeing it there invariably people will say, "Yes." There are many other experiments in which the borderline, where we go from psychology, perhaps, into physics, is so nebulous as to influence our decisions. This is a point, as Dr. Howlett mentioned, actually of rather deep philosophical importance, but I believe that the point that was made in Dr. Ingelstam's abstract—it is in quotations where he says so-and-so. I would like to see some discussion at a later date—it is getting late tonight—on some support of experiments that could be conducted to find out just how much of a part the observer plays from the point of view of seeing what he wants to see. I have measured spectral lines and after awhile I knew I was measuring lines that were not there but I was recording wavelengths for them.

I would like to hear Dr. Ingelstam's further delineation of his ideas of putting down this term "standardization."

Dr. E. Ingelstam, Royal Institute of Technology, Stockholm 70, Sweden: I quite agree with you. One should distinguish what is really physically measurable concerning the intensity distribution. How it should be done I can not say at present. I think there could be some discussion tomorrow after we hear Selwyn's paper and some other papers. However, I think that one should make a separation between the physical resolving power and the intensity distribution in a given instrument, and the influence of the receiver, maybe the eye or the photographic plate. I think that there is sufficient work done now to enable one to make decisions of this type. I quite agree with the logic of your question.

14. Optical Calculations at the National Bureau of Standards

By Donald P. Feder [1]

Since the advent of automatic computing machinery in the past few years, these machines have been increasing in numbers very rapidly. There are at present some six, large-scale, general-purpose, automatic, digital computing machines in operation. In the next few years their number will have increased to about 50. In addition there are a larger number of medium-sized machines in use. In view of the increasing availability of this equipment, which can greatly facilitate optical calculations, you may like to hear something about these machines and the optical calculations which have been programmed for them, at the National Bureau of Standards.

Three words have been used to describe these machines, "automatic", "high speed", and "electronic". By "automatic" one means that the initial data and the constants of the system are introduced into the machine, which solves the equations, stores necessary intermediate results, makes logical decisions about the course of the calculations, and finally prints out the answers in a useable form.

The phrase "high speed" can be best illustrated by an example. The SEAC (Standards Eastern Automatic Computer) has been coded to trace a general ray through a system of spherical surfaces. Through a system of 10 such surfaces, the SEAC can trace a skew ray in 8 seconds. In this time it does 440 additions, 320 multiplications, 10 divisions, and 20 square roots, and it does these things to 44 binary places (about 13 decimal places).

Finally "electronic" simply refers to the technology that has made these machines possible.

In addition to the SEAC there is also available at the Bureau a CPC (Card Programmed Electronic Calculator). This is a machine of more moderate speed and capacity. It has a memory for 23 numbers in one type of storage, and, in addition, a small high speed memory. The use of the high speed memory permits the very rapid extraction of square roots.

By comparison the SEAC has a memory of 1,024 words. In the SEAC, however, some of the storage space always contains instructions. This limits the effective memory considerably. In the CPC the instructions are on punched cards and do not occupy any of the memory cells.

The CPC has been used for routine ray tracing since March 1950. This machine required 34 seconds to trace a skew ray through a spherical surface. It traces one ray at a time, which distinguishes it from certain other machines of this type. It has also been coded for non-spherical surfaces so that now rays can be traced through any system of rotationally symmetrical surfaces. The aspheric surfaces are

[1] National Bureau of Standards, Washington, D. C.

handled by adding a power series to the equation of a spherical surface. The power series may contain as many terms as necessary and this introduces no storage problem because the coefficients of the power series are stored on punched cards. The problem is to find the intersection of the ray with the aspheric surface. A simple iteration has been found, which converges and, what is more important, converges in less than five iterations for any aspheric surface likely to be used optically. The entire calculation requires about 2 minutes per surface, but since not many surfaces of a system are apt to be aspheric, this is not very serious.

Something should be said about the type of formulas used in these calculations. They involve only curvatures and never radii of curvature. Each number used by the computer is bounded in a known way by the linear dimensions of the optical system. Hence, by choosing a proper scale factor it is always possible to fit the problem into the machine. Floating decimal operations are thus avoided. All the numbers involved are of the same order of magnitude and the accuracy is very good.

These same statements might also be made in connection with the other optical calculations discussed in this paper. The formulas have always been converted to such forms that no number has the possibility of becoming infinite. In practice this means that no number becomes very large and so accuracy is maintained.

The other type of calculations presently being done at the Bureau are calculations of the image errors of various orders. The CPC was first programmed for the calculation of the first and third orders (that is the Gaussian and Seidel coefficients). For this purpose two paraxial rays are traced through the system, a paraxial marginal and a paraxial principal ray. The machine runs through the calculation and prints the following data: The height of each ray on each surface; the slope angle of each ray in each medium; the coefficient of spherical aberration, of coma, of astigmatism, of field curvature, and of distortion. In addition it prints the first-order coefficient for the two types of chromatic aberration and five checks, which ought to be zero. These checks may be inspected to verify that no mistake has been made. The entire calculation including the printing requires only 40 seconds per surface.

A somewhat more unusual calculation is that to obtain the fifth-order coefficients. The equations are algebraic equations modified from those published by Wachendorf.[2]

It is too early to say how valuable the fifth-order coefficients will be since our experience with them is extremely limited. It is likely that their chief value will be in optical design rather than in the evaluation of image quality.

With the CPC the complete calculation of the fifth-order coefficients requires 7 minutes per surface. While this may seem a long time, the calculation is very laborious and would require almost too much time to be practical for hand computing. This calculation is being programmed for SEAC and on this machine will probably require only a few seconds per surface.

The calculations mentioned above are all being done routinely at present in this laboratory. It can be seen that with the possible

[2] F. Wachendorf, Optik 5, 80 (1949).

exception of the fifth-order coefficients and the tracing of skew rays through aspheric surfaces, these calculations are no more extensive than are commonly performed by lens designers using hand machines. We have seen that there are gaps between the information that the designer gains in this manner and the information that the user of the system would like to have.

In the past the designer has usually obtained curves denoting the aberrations of meridian rays as functions of aperture, and curves showing the position of the sagittal and tangential foci as functions of field angle. These have not been sufficient to answer many questions about the image formed by the system. Such a question, for example, as the calculation of the limiting resolution in lines per millimeter from a particular combination of object, lens, and film could not be answered from such data alone. More generally, one might be interested in the image contrast from an object of given contrast produced by a lens and film of known characteristics. Questions such as these ought to be answered by calculation from the constructional data of the system.

In accomplishing this purpose various approaches might be tried. With a machine such as SEAC, it should be possible to calculate the diffraction disk for a luminous point. On the other hand it may not be necessary to go beyond the geometrical picture, especially for photographic objectives where the aberrations are likely to be many times the Rayleigh limit. In this case it is possible to get an approximate distribution of energy in the image using a method similar to that used by Herzberger.

In this method one divides the entrance pupil into a large number of equal areas. For a particular object point a ray is traced through each of the small elements of area, and its intersection with the image plane is calculated. One supposes that the energy of each element of the wave front arrives at the image plane in the neighborhood of the ray. If the calculation is made for a large number of rays, the distribution of their image points gives an approximate picture of the energy distribution in the image of a luminous point. When the distribution of energy is known, a knowledge of the characteristics of the photoreceptor should enable one to find the appearance of the image.

Another possibility is to calculate the interferometer pattern for various field angles. A large number of rays might be traced by a machine such as SEAC which would retain in its memory the results of the calculations, fit them to some favorable type of function giving the wave surface and then by intersecting this surface with a series of spheres one-fourth of a wavelength apart obtain an interferogram. This could then be interpreted directly to secure an idea of the image quality.

15. Resolving Power of Airplane-Camera Lenses

By F. E. Washer [1]

Shortly after the close of World War I, the National Bureau of Standards began testing photographic objectives for the U. S. Army Air Corps to determine their suitability for use in airplane cameras for mapping projects. The measurements made included determinations of spherical aberration, chromatic aberration, curvature of field, astigmatic difference and distortion.

These measurements were made on a visual optical bench [1][2] and for a time were entirely adequate. However, as the use of airplane photography for both planimetric and topographic mapping increased, it became evident that the volume of work submitted to this Bureau was too great to handle by the visual method. In addition, the evaluation of the quality of definition throughout the image plane together with the location of the plane of best average definition is a laborious and time-consuming process when performed visually. Moreover, legitimate doubt exists as to whether a visual method can be satisfactorily used to determine probable photographic quality of definition.

To meet the increasing demand for a rapid and adequate performance test for photographic objectives, the precision lens-testing camera, shown in figure 15.1, was conceived by I. C. Gardner and designed by F. A. Case [2]. This device in its earliest form permitted the evaluation of the equivalent focal length, distortion, and resolving power from a single negative. Nineteen rows of images were obtained on the negative, each of which showed the quality of imagery at 5° intervals from the center of the field to a point in the image plane 30° from the axis for a specific distance from the lens. Consequently, it became an easy matter to locate the plane of best average definition and to determine the back focal length of the lens associated with the plane of best average definition.

A resolving-power test chart, shown in figure 15.2, was developed for use in this equipment, consisting of a series of patterns of parallel lines with the spacings varying by the square root of two. The contrast between the lines and spaces in the target patterns is high and may be considered as infinite.

With the large number of lenses being tested, it was only natural that a considerable quantity of test negatives would accumulate that could be used to carry out statistical investigations of the quality of imagery inside and outside of focus as well as in the plane of best axial imagery. The first of these studies was reported in 1939 [3] and dealt with three types of airplane camera lenses in most common use at that time. The wide-angle lens had not yet come into extensive use at that time and the investigation was confined to imagery in the angular region between 0° and 30° from the axis for lenses having a focal length

[1] National Bureau of Standards, Washington, D. C.
[2] Figures in brackets indicate the literature references on p. 218.

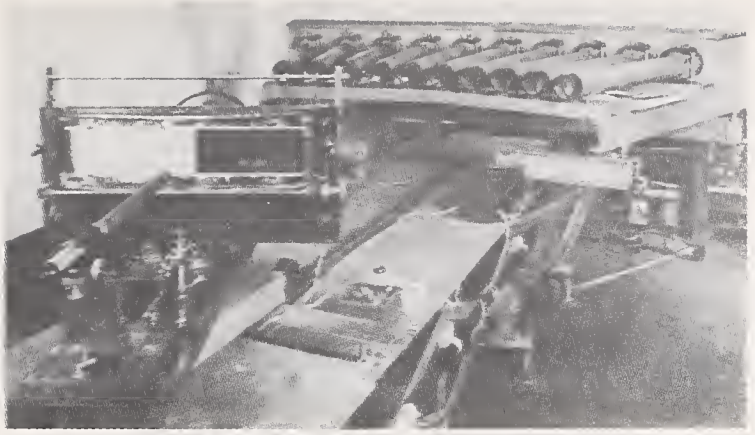


FIGURE 15.1. *Precision lens-testing camera.*

The 10 collimators are in the upper right-hand part of the photograph; the lens testing bench and plate holder are in the lower left-hand part; the lens holder is obscured by the camera back and does not appear in the picture.

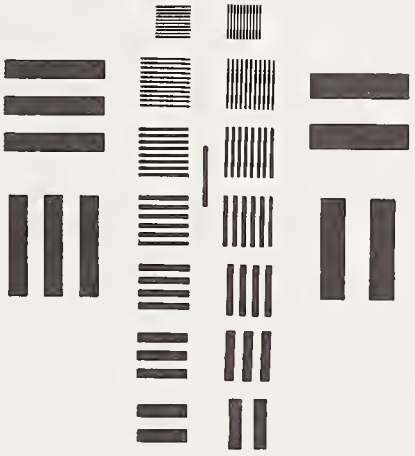


FIGURE 15.2. *Chart for testing resolving power.*

A chart of this type is located in the focal plane of each collimator.

of 8¼ inches. Several items of interest were found in this early investigation. First, the resolving-power characteristics of a given type of lens remain remarkably constant from one lens to another. Second, what appeared to be striking variations from one lens to another so long as a single image plane was considered faded into insignificance with proper choice of the focal plane. Third, by proper choice of a focal plane, the performance of a lens in the outer portions of the field could be considerably improved with only a small diminution of quality in the axial regions. This finding led directly to the concept of plane of best average definition, which is now recognized as the best plane to bring into coincidence with the focal plane of the aerial camera. Fourth, the phenomenon of false resolving power was noted. This led to the concept that a pattern should only be counted as resolved when it and all coarser patterns in the series were resolved.

Each of the collimators of the lens-testing camera is provided with one of these charts with the line patterns so proportioned that when a given pattern is imaged by the lens under test the same value of resolving power is indicated regardless of its position in the field. The values of the resolving power in the image plane of the test lens usually ranges from 3.5 to 56 lines per mm for both tangential and radial lines.

Eastman type V–P spectroscopic plates are generally used. This emulsion, while too slow for airplane photography, has much greater resolving power than the panchromatic films used in aerial photography. Consequently, this makes it probable that an indicated deficiency in the observed resolution arises from a deficiency in the lens rather than in the photographic emulsion. The plates are processed in Eastman D19 developer for 3 minutes at 65° F.

When the lens-testing camera was first put in operation in early 1936, perhaps 50 lenses were tested the first year. The number has increased steadily since, and during one recent year more than 300 lenses were tested. It is safe to say that the performance of well over 1,500 lenses have been evaluated on the lens-testing camera alone since 1936.

By 1940, it was evident that the trend was toward the use of wide-angle aerial camera lenses and the precision lens-testing camera was modified to extend its range to cover the region from 0 to 45 degrees from the axis. The second study on resolving power was reported in 1942 [4] and dealt among other things with the resolving-power characteristics of several types of wide-angle lenses.

Throughout the years of making test negatives on the precision lens-testing camera, it was always evident that the device was admirably well suited to investigating image quality throughout the region of usable imagery and in 1945 [5] a study was reported along these lines. The curves, shown in figure 15.3, illustrate the manner in which the resolving power increases and decreases as the image plane moves steadily from a point well inside focus to a point well outside focus. The results are plotted at each 5° interval from 0 to 45 degrees from the axis for both tangential and radial resolving power. The zero line of abscissae corresponds to the point of best axial focus. The effect of curvature of field is strikingly manifested by the change in relative position of the maximum with respect to the plane of best axial imagery. The varying widths of the principal maximum increase with increasing f-number and in the graphs shown for f/16, the heights of the peaks are beginning to show the expected decrease. From curves of this type, complete information concerning the effect of stop opening on resolving power and depth of focus; the change in maximum resolving power with angular separation from the axis; and the effect of change in stop opening on the resolving power in a given image plane may readily be obtained for a given lens.

The possibilities of the precision lens-testing camera as a research instrument in the study of resolving power are by no means exhausted. However, since this device was first put in operation the study of resolving power of aerial camera lens has become the subject of interest in other laboratories both here and abroad. These other investigations brought fresh points of view and new ideas regarding the manner in which the subject should be approached. A wide variety of test charts was developed by the various groups. Chief among these

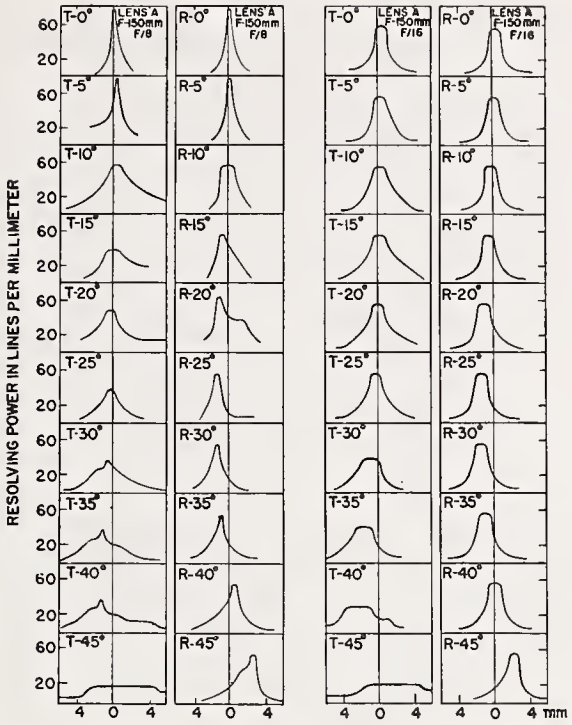


FIGURE 15.3. *Resolving power versus position of the image plane for lens A a apertures f/8 and f/16.*

The resolving power throughout the region of useful imagery is shown for tangential T and radial R lines at 5° intervals from 0° to 45°. The zero of abscissae marks the position of best visual focus on the axis at f/8 and positive values of abscissae indicate positions farther from the lens.

charts are the three-line high-contrast target of the Air Force [6], the low-contrast annulus target of the Canadian group [7] and the low-contrast two-line or Cobb chart of the British group [8].

The principal changes in approach from that followed at this Bureau are the emphasis on low contrast and the use of emulsions having characteristics more nearly approximating the emulsions used in the aerial camera. The low contrast of the test chart was justified on the basis that the difference in brightness between adjacent objects on the ground was small when viewed from an airplane and, consequently, the variation in contrast in an aerial negative is low.

As a result of the findings of these other laboratories, there has been some tendency to question the suitability of the high-contrast chart in the determination of resolving power. It is right and proper for such questions to be raised, but in defense, it ought to be stated that the results of the resolving-power test as used at the National Bureau of Standards have been applied as a go or not-go criterion by the various mapping agencies. A lens to be accepted for use in aerial photography must satisfy certain minimum requirements. These minimum requirements were early established by examining a number of negatives made by different cameras and assessing the negatives as being suitable or unsuitable for use in map making. On checking back to the resolving-power measurements made on these lenses for the high-contrast target, it was found that the unsuitable negatives were obtained with lenses having a resolving power less than 7 lines/mm, whereas for the suitable negatives the resolving power of the lenses exceeded 7 lines/mm in all parts of the useful field. It is

unfortunate that this correlation was not made the subject of a formal report instead of being informally presented as was done at the time.

This work was done by a committee of the American Society of Photogrammetry and the suggested lower limit of 7 lines/mm first appeared in 1935 in a tentative set of standard specifications [9] prepared by this society. In later specifications the figure was changed to 20 lines/mm for central imagery with the figure left at 7 lines/mm throughout the remainder of the useful field. In addition, there are Federal Specifications pertaining to optical performance that require that the quality of imagery be determined with the aid of high-contrast charts. The Federal Specifications are used extensively in connection with government purchases. The requirements contained therein have been developed over a period of years in collaboration with representatives of industry and government agencies. Consequently, any change that is recommended must be in the direction of a definite improvement that is clear to all concerned. A change in requirements frequently necessitates the development of new test methods, the development of which must precede any actual change in the Federal Specifications themselves. Accordingly, changes in these specifications are not quickly made and even though new requirements may appear desirable, the existing requirements continue in use until the new requirements with appropriate test methods are agreed upon.

Because of the uncertainty that had been created in the minds of those people concerned with specifying quality of definition as a result of the many diverse opinions on the subject, it was decided to initiate further research at this Bureau into the factors affecting definition in the photographic image. A new resolving-power test chart was designed that would simultaneously give information on resolving power and the effect of contrast on resolving power. At the same time, the chart was provided with a greater range of frequencies and capable of yielding greater precision in determining the limit of resolution. The design of this chart was reported by I. C. Gardner in March 1950 [10] at the New York meeting of the Optical Society of America.

The actual making of the new chart was reported at the Cleveland meeting of the Optical Society of America in October 1950 [11]. This new resolving-power chart is so made that the transmittance of the lines and spaces along the y-axis vary in such a manner that the contrast is a linear function of y and the transmittance of the chart averaged over an area embracing several pairs of lines is uniform for the entire plate. Along the x-axis, the widths of the lines and spaces progressively decrease so that the "instantaneous" value of the line frequency is a linear function of the distance measured normally to a "zero" line of abscissal of the chart.

The chart was made in two steps, the first step being the making of the master high-contrast chart, shown in figure 15.4. The range of frequencies on the master target plate is from 0.2 to 2 line/mm, which means that the actual widths of lines and spaces vary from 2.5 mm to 0.25 mm. To cover this range in a continuously varying linear manner there are 200 lines and 200 spaces contained in a distance of 180 mm. The zero line is 20 mm from the origin so that the over-all length of the resolving-power portion of the chart is 200

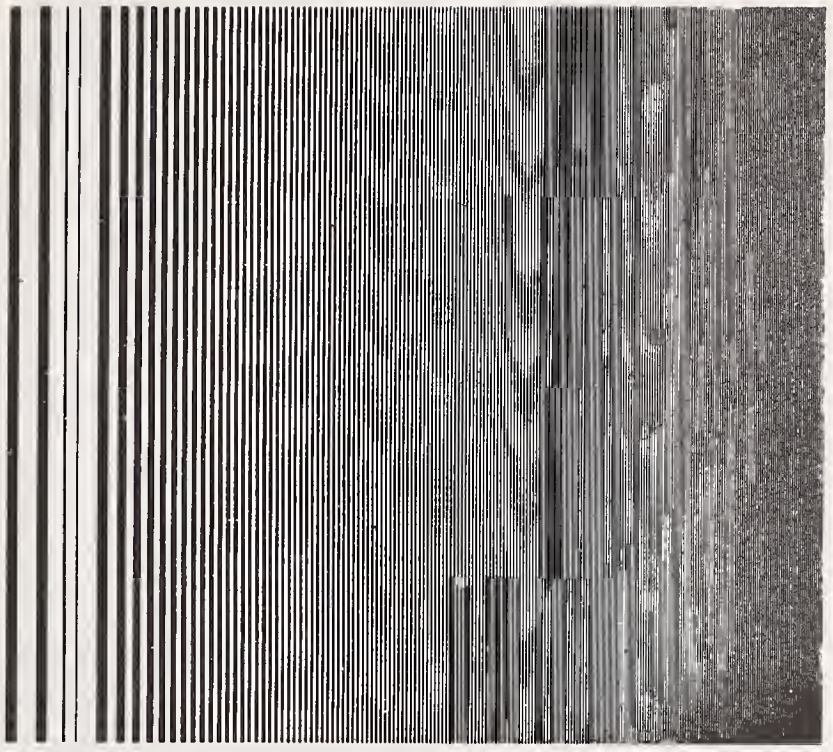


FIGURE 15.4. *Master high-contrast test chart.*

The line frequency is a linear function of the distance measured across the lines. The zero of the chart is at the inner edge of the first broad line.

mm. It is necessary that the lines be as long as possible to provide a contrast scale of approximately the same length as the frequency scale and, moreover, to simplify the process of making microdensitometric studies of final test images. The length of line used is approximately 185 mm, enabling the entire chart to be registered on a standard 8- by 10-inch plate.

The second step consists in contact printing the high-contrast target on a photographic plate under conditions such that the exposure time varies over the plate in a predetermined manner [11]. The high-contrast target is then removed and a second exposure is given to the photographic plate in such a manner that the transmittance averaged over several lines of the finished negative is a constant. Figure 15.5 shows a typical variable-contrast target produced by this process. For the particular variable-contrast target illustrated in figure 15.5, the scale at the left marked "Contrast" shows that in the original transparency the contrast varies in a linear manner from 0 to 1.5. Contrast as used here is simply the difference in photographic density between dark and light areas as measured on a densitometer. The lower scale gives the line frequency or resolving power in lines per millimeter at a 25x reduction. Since the present print is somewhat reduced in size, the marked scale is correct only when the distance from 0 to 50 is equal to 8.0 mm in the final image. The chart was photographed for use in this illustration and it is unlikely that the contrasts shown on the scale are still correct. However, in the original, the contrast remains constant at a given y-height, whereas the line frequency varies linearly from 0.2 to 2.0 lines/mm.

518-743 O - 73 - 28

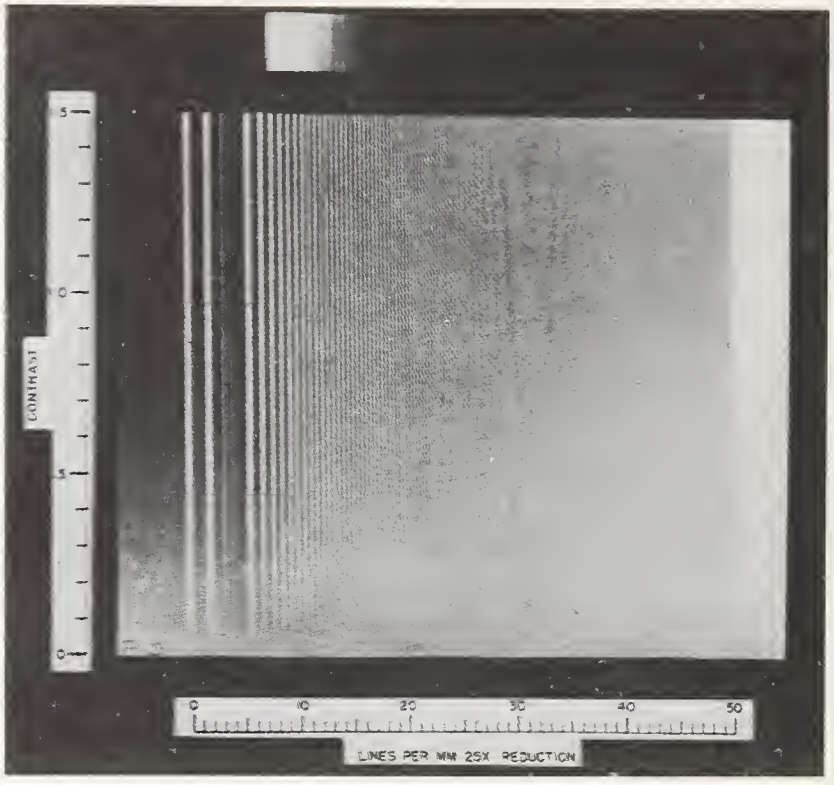


FIGURE 15.5. *Variable-contrast master test chart.*

In addition to the features contained in the master high-contrast chart, the contrast varies in a linear manner from 0 to 1.5. The "moiré" effect appearing in the right-hand portion of the illustration is produced by the interaction of the periodic pattern of the halftone screen and the pattern of continuously varying line spacing in the chart. Consequently, the above figure is not a true reproduction of the chart, although it is a fairly close approximation in the left-hand portion.

Some trials have been made with this new target and the results are shown in figure 15.6. The target was set up at a distance of 51 focal lengths in front of the test lens. It was lighted from the rear by a broad uniform light source and negatives were made using three different emulsions with the lens set at an aperture of f/8. Figure 15.6 shows the results obtained with Eastman Spectroscopic Plate, type V–G. The reduction ratio is 50 to 1 so that the frequency that is shown at the bottom of the figure ranges from 10 to 100 lines per millimeter. Values of the contrast obtained from microphotometric recordings at three different heights across the negative image are shown as a function of the line frequency. The contrast in the target at a given height is approximately constant, so that the decrease in contrast in the image results from action by the lens-emulsion combination. Curve 1 shows the results for high-contrast target, curve 2 for a medium-contrast target, and curve 3 for a low-contrast target. The results appear to be in agreement with what one might anticipate on the basis of the literature dealing with the subject of target contrast and resolving power. However, the chief point to emphasize here is that the data for drawing these curves were obtained from a single negative, with consequent marked reduction in the number of factors that might conceivably make difficult the correlation of resolving power with contrast in the target.

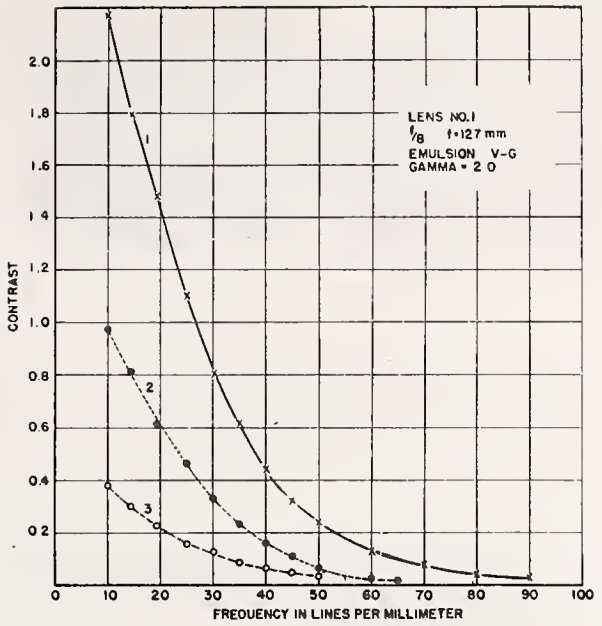


FIGURE 15.6. *The variation in contrast for increasing line frequency for Eastman spectroscopic type V–G emulsion.*

Target contrast is 1.3 for curve 1; 0.5 for curve 2; and 0.2 for curve 3. (The average density of the negative is 0.45).

At the present stage of the investigation, it is clear that it would be difficult to determine just what value to assign to the upper limit of resolving power of the lens for a given target contrast. The resolving power reaches its maximum when the contrast in the image reaches zero. However, because of local irregularities in the image it is not possible in the present stage of the investigation to state with certainty just when the contrast in the image reaches zero. As a practical expedient it may be preferable in reporting the performance of a lens under given conditions to state what the value of the resolving power is for a selected value of contrast in the image for a specified value of contrast in the target.

For example, we may wish to report what the resolving power is for an image contrast of 0.1 for several values of target contrast. This information is readily obtained from these curves. Table 15.1 shows the values of the resolving power for an image contrast of 0.1 for three emulsions for the three values of target contrast that were shown in figure 15.6.

TABLE 15.1 *Resolving power in lines per millimeter for image contrast of 0.1 for three emulsions for three target contrasts*

Target	Contrast	V-G	Pan Process	Panatomic X
1	1.3	64	36	30
2	0.6	46	25	20
3	0.2	33	13	13

The development and use of the multiline variable-contrast target have been of primary interest in our work on resolving power during the past year. Consideration has been given to 3-line, 2-line, and

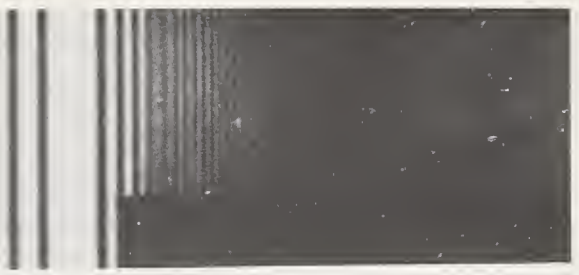


FIGURE 15.7. *Typical out-of-focus exposure showing spurious resolution.*

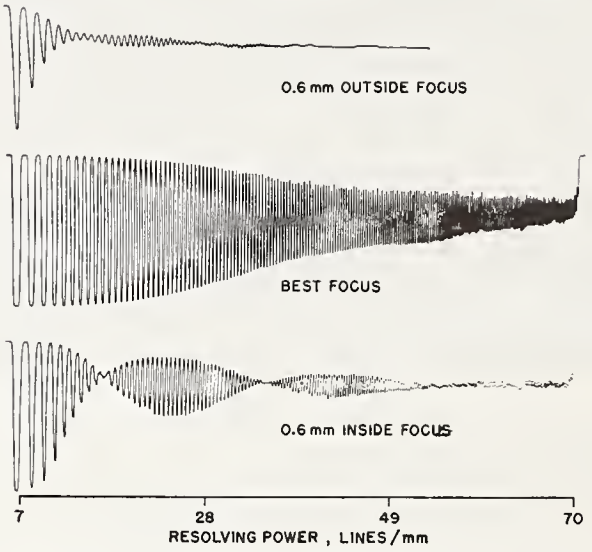


FIGURE 15.8. *Microphotometer traces of images formed by a lens at f/5.6 on Eastman spectroscopic emulsion type 548 GII.*

1-line targets. Some of the results obtained with these targets, together with their theoretical interpretation, are reported in paper 20.

Work with high-contrast targets has not, however, been completely neglected. The high-contrast target with continuously varying frequency has been found to be exceptionally well suited to the study of some phases of spurious resolution. A report on this phase of the investigation was made by R. N. Hotchkiss [12] at the Cleveland meeting of the Optical Society of America.

The axial imagery was studied throughout a region extending from a point well inside best focus to a point well outside best focus. A typical out-of-focus image is shown in figure 15.7. It is readily seen that the contrast between adjacent lines and spaces quickly falls to approximately zero as one proceeds toward the finer patterns. However, instead of remaining zero, the contrast rises and falls repeatedly as one continues in the direction of the finer patterns. This waxing and waning effect is even more strikingly illustrated in the microdensitometric traces shown in figure 15.8. For the cases shown, the phenomena can be explained fairly simply on a geometric basis. This analysis is given in the September 1951 issue of the Journal of the Optical Society [13].

This work was performed in connection with a research project sponsored by the Office of Air Research of the United States Air Force. The success of this project to date is a result of the cooperation of a number of people working on this project. These include Mr. Rosberry and Mr. Shack who devote full time thereto; Mr. Magill, Mr. Tayman, and Mr. Darling who give a part of their time to the project; and to Mr. Sine and Mr. Watts who assist in the necessary design of instruments and the preparation of graphs.

---

[1] A. H. Bennett, J. Opt. Soc. Am. **14,** 235 (1927).
[2] I. C. Gardner and F. A. Case, J. Research NBS **18,** 449 (1937) RP984.
[3] F. E. Washer, J. Research NBS **22,** 729 (1939) RP1216.
[4] F. E. Washer, J. Research NBS **29,** 233 (1942) RP1498.
[5] F. E. Washer, J. Research NBS **34,** 175 (1945) RP1636.
[6] C. W. Kendall and B. A. Schumacher, Photo Tech. **3,** No. 4, 51 (1941).
[7] L. E. Howlett, Can. J. Research **24,** Sec. A. No. 4, 15 (1946).
[8] E. W. H. Selwyn and J. L. Tearle, Proc. Phys. Soc. **58,** 493 (1946).
[9] Standard Specifications, 19 and 20, published by the American Society of Photogrammetry Nov. 15, 1935.
[10] I. C. Gardner, J. Opt. Soc. Am. **40,** 457 (1950).
[11] F. E. Washer and F. W. Rosberry, J. Opt. Soc. Am. **40,** 801 (1950); **41,** 597 (1951).
[12] R. N. Hotchkiss, F. E. Washer, and F. W. Rosberry, J. Opt. Soc. Am. **40,** 802 (1950).
[13] R. N. Hotchkiss, F. E. Washer, and F. W. Rosberry, J. Opt. Soc. Am. **41,** 600 (1951).

16. Theory of Resolving Power

By E. W. H. Selwyn [1]

Fundamental Theory

The test object for a measurement of resolving power is usually one in which the demarcation between areas of different brightness is a hard edge. A common and typical example consists of dark and bright bars. When an image of this is produced, either an aerial image or one in a sensitive photographic layer, light spreads from the bright bars into the dark. The diagram representing the brightness distribution in the test object is like the castellations on old fortresses, wherein the light bars are represented by short horizontal lines interposed between other short lines, representing the dark bars, spaced lower and connected by vertical lines. If the degree of spreading of the light is small the corresponding diagram for the image is somewhat similar but the corners are rounded and the vertical lines are replaced by sloping ones. As the spread of light is increased, relative to the spacing of the lines, the corners become more rounded, and the slope of the connecting lines becomes greater until the spread of light is about equal to the space between two light lines. A still greater spread of light then causes the hollows to start filling up and the tops of the diagram to drop. If we wish to be strict in terminology we may say that the first stages represent a diminution in sharpness and the last stage a diminution in contrast. When the last stage has been reached the top of the diagram representing the distribution of brightness approximates in shape to a curve of sines. If we start with a test object in which the graduation from light to dark lines is smooth and represented by a sine curve, only the contrast change can take place, for the effect of spread of light must then always be to raise the hollows and bring down the tops of the curve. Also the curve cannot change its character. It always remains a sine curve, with the same period. That this is necessarily so can be seen from the consideration that two sine curves of the same period always sum to another sine curve whatever the amplitudes of the original curves and whatever the phase difference between them. We may consider the spreading of the light as producing from the sine curve of the original test object a whole set of sine curves of equal period, with a total phase difference equal to the total spread of light and amplitudes varying according to the character of the spread of light. Thus the final curve for the image is also a sine curve with a lower amplitude (but of the same period). It may also be shifted in phase from that which would be appropriate if the image were perfectly sharp.

---

[1] Research Laboratories, Kodak Limited, Harrow, England.

Apart from any difference of phase, which is related to the optical distortion produced by the image forming system and is not of interest in the pure resolving power problem, the only physical variables involved are the period of the sine curve and the contrast. By eliminating the gradual change in shape of the light-distribution curve (by choosing a sine-curve test object in the first place) it becomes possible to proceed with theoretical arguments. This is mainly because the necessity of knowing how the changes of shape affect visibility of the image of the test object is avoided. The contrast in the image depends upon the interval between the bright bands, for as the test object is made smaller and smaller (the image following suit) the smudging-out effect of any unsharpness in image formation becomes more and more pronounced. In theoretical discussion the relation between the contrast in the image and the period of the sine curve is of fundamental importance. We shall find it convenient to consider first an optical system forming an aerial image.

The diagram representing the brightness of the test object or its image may be split into two parts, one representing a uniform brightness and the other the sine-curve portion, consisting of alternate positive and negative sections. The uniform brightness is unaffected by any lack of sharpness, but lack of sharpness diminishes the positive and negative sections of the sine-curve portion equally since they are of identical shapes. In other words, what happens to the amplitude of the sine-curve is independent of what happens to the mean brightness and if we change the mean brightness of the test object without changing the amplitude (in brightness) of the sine-curve component, the mean brightness in the image will be changed but the amplitude of the sine-curve component in the image will remain unchanged, and vice-versa. If a neutral filter is introduced between the test object and the lamp illuminating it the amplitude will be reduced (so also will the mean brightness but this is immaterial) and the amplitude in the image will also be reduced in the same proportion. Thus the ratio of the amplitudes in image and test object is constant for any given test object and optical arrangement. In fact, however, the amplitudes and mean brightnesses change in the same ratio, so that for both test object and image, the important quantity is the ratio of amplitude of brightness to mean brightness. If we call this ratio for the test object a and for the image a^*, then a^*/a is a constant for any one test object and optical system. Now the amplitude in the image is dependent upon the amplitude in the test object, upon the optical system and the spacing between bright stripes in the image. It will be convenient to use the reciprocal r of the last quantity. The above result therefore shows that

$$a^*=af(r), \tag{1}$$

$f(r)$ being a function characteristic of the optical system, which may be calculated, at least in theory, from the specification of the optical system. The process of calculation starts from the brightness distribution in the image of a long line source. Each linear element of this image is regarded as producing a geometrically perfect image of the test object, a sine wave, that is, with phase appropriate to the distance of the linear element from the center of the image and amplitude proportional to the brightness of the element. All the sine

curves of this type are added. The above is the fundamental formula of the theory of resolving power when the test object is of the sine-curve form.

Visual Resolving Power

The value of r that makes $af(r)$ equal to the least value of a^* that can be appreciated by the eye, under the conditions of observation, is the visual resolving power of the system. The threshold value of a^* is not a quantity that can be calculated, a priori, but must be determined experimentally. It may depend on the effective aperture of the pupil of the eye, on the angular separation between the bright stripes in the image on the retina, and upon the mean brightness of the image, apart from other conditions such as the presence or absence of a bright surround to the image. Let us take a simple system consisting of a test object, a perfectly corrected lens forming an aerial image of it, a perfectly corrected eyepiece to make the emergent wave fronts flat, and the eye. The aerial image may be replaced by a luminous test object exactly similar to the original but with spacing between the bright stripes changed according to the magnification and with lower contrast (an amplitude in fact of a^*), owing to loss of definition as a result of diffraction. That this is possible must be so because the wave front arising from any point is spherical, exactly as with the aerial image. But we have to assume that there is no diffraction by the eye, although there may be distortions of the wave front by refraction. The effective aperture of the eye is determined by the exit pupil of the lens and eyepiece system. With such an arrangement, using a test object with variable contrast, and placing the lens at different distances it is possible to determine a wide variety of conditions under which the image just ceases to be visible, as an array of bright and dark stripes. The value of $f(r)$ can be calculated, since the image-forming properties of a perfectly corrected lens are known from Airy's discussion of diffraction by such a lens. Thus it is possible to calculate $af(r)$ at the limit of resolution, the angular separation between bright bands, as seen by the eye, the effective diameter of the pupil of the eye (the exit pupil of the telescopic system), and the relative brightnesses of the various images. The results that were obtained by such an experiment are exhibited diagrammatically in figure 16.1. There appeared to be no influence of

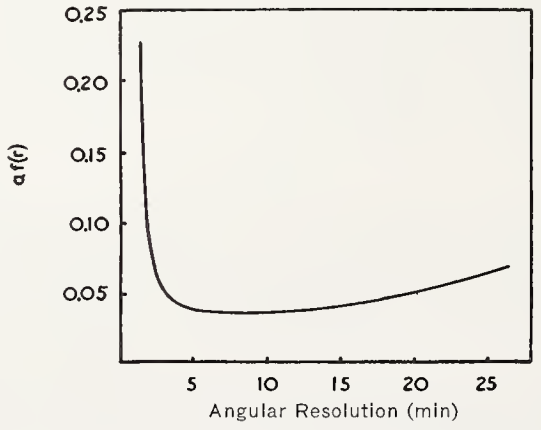


FIGURE 16.1. *Visual threshold as a function of size of detail.*

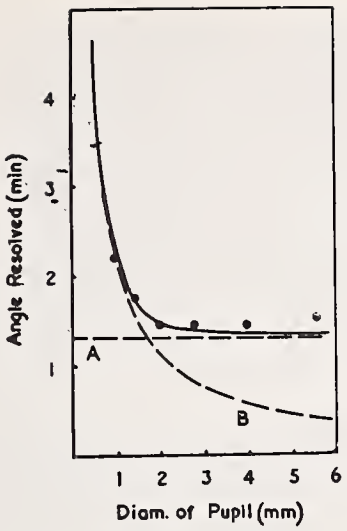


FIGURE 16.2. *Comparison of Cobb's experimental results with theory.*

A, Limit set by retina and/or nystagmus; B, limit set by diffraction at pupil.

brightness (over a range of 10,000:1 !) outside the experimental error. The striking feature about figure 16.1 is that the curve approaches an asymptote at around $1\frac{1}{4}$ minutes of arc, but that apart from this the threshold value of $af(r)$ is remarkably nearly constant. The simplest way of accounting for the asymptotic rise is to suppose that the eye averages out detail over a range of $1\frac{1}{4}$ minutes or so, in consequence of coarseness of structure of the retina, or involuntary movements (nystagmus) or both.

There is no evidence in the results that the value of $af(r)$ at the threshold varies with pupil aperture over a range of 0.1 to 2.0 mm and it is probably legitimate to assume that below pupil apertures of 2 mm the eye is perfectly corrected. The light used in the experiments was monochromatic and it may be that this is the reason why the eye appears to be perfectly corrected to an aperture as large as 2 mm.

If it is assumed that the eye is perfectly corrected and that the retina averages out the illumination over 1¼ minutes of arc, the value of a^* in the image transmitted by the retina to the brain may be calculated, in terms of brightness, from the formula

$$a.f_1(r).f_2(r),$$

where $f_1(r)$ is a factor calculated for diffraction by the eye and $f_2(r)$ a factor calculated for the averaging effect of the retina. If now a^* is taken as 0.03, and a unity, corresponding to a high-contrast test object, the results of Cobb on the resolving power of the eye at different pupil apertures may be well duplicated, except that at apertures above say 3 mm the observed resolving power is rather less than that calculated, presumably owing to the greater effect of aberrations at higher apertures (fig. 16.2).

Photographic Resolving Power

For some purposes the theory of photographic resolving power can be very simple. If sine curve distributions of light, of different contrasts, were projected on a photographic material it would be possible to determine the relation between the resolution of the developed

image and the contrast of the light distribution. Suppose that the results are exhibited in the form of a curve relating the minimum contrast of incident distribution necessary for the light and dark stripes to be just visible in the developed imaged and r the reciprocal of the distance between neighboring light stripes. Now draw on the same diagram the curve showing the relation between the contrast in the aerial image produced by an optical system and r. At the point where the two curves cross, the optical system is producing an image with spaces between stripes equal to those at which resolution just occurs with an image on the material of the same contrast as that produced by the optical system. The value of r at the common point is therefore the resolving power of the combined optical system and sensitive material.

By this means we have accomplished a synthesis between the purely optical and the purely photographic properties of the combined system. In principle, if the photographic curves for N photographic materials and the optical curves for N^* optical systems were known we could immediately find the resolving power for all the NN^* combinations, for any and every contrast of test object.

There is, however, a difficulty. The above argument is sound if the light is monochromatic. But if it is not monochromatic we may be faced with a situation of the following type. Suppose we have a lens giving an aerial image of a point source consisting of a bright central dot of red light surrounded by a thin blue halo. This image is first supposed to be projected on a sensitive layer that diffuses red light more than blue. The image in the emulsion layer will then consist of a diffuse red central area with a moderately diffused blue halo. Now suppose the aerial image is projected on a sensitive layer, with the same diffusing properties, for white light as a whole, as the preceding one, but which diffuses blue light more than red. The image in the emulsion layer will then be a fairly small central red spot with a diffuse blue halo. When converted to black and white (by the process of development) these two images will be quite different. In other words we must take account of the distribution of color in the aerial image and of the variation with color of the diffusion of light in the sensitive layer. In practice this means that one needs to calculate the amplitude of the sine distribution in the sensitive layer for every wavelength of which the light is composed and then average according to the color sensitivity of the material. Alternatively one may calculate the distribution of actinic energy in the sensitive layer of the image of a long thin line of light and use this for calculating the amplitude of the sine-curve image.

Resolving Power of Sensitive Material

The amplitude of a sine-curve distribution of light projected on the sensitive layer is reduced by diffusion of light in the layer, according to the formula,

$$a^{**}=a^{*}\varphi(r), \tag{2}$$

analogous to that for an optically formed image. The process of development converts small percentage variations in exposure into the same percentage variations in transparency, multiplied by γ, the

slope of the usual density versus log exposure curve of the material at the density at which the mean exposure is reproduced. The amplitude of the image after development is therefore $\gamma a^* \varphi(r)$. This is true only if the amplitude is small, otherwise there is some distortion of the sine curve. If the photograph is magnified M times the apparent separation between bright stripes in the image is increased in the ratio M and the reciprocal of the separation is decreased, so far as the eye is concerned, to r/M. Thus the amplitude of the signal to the brain, in terms of light intensity, is reduced to

$$\gamma a^* \varphi(r) F(r/M),$$

where $F(r/M)$ is a factor (equivalent to $f_1(r) f_2(r)$, used in discussing Cobb's results on variation of acuity with pupil aperture) to take account of the unsharpness in image formation in the eye and the averaging effect of the retina.

So far, everything has proceeded on the assumption that the image is "smooth" that is to say, without taking any account of the irregular distribution of grains in the photographic material. In order to deal with the effects of this irregular distribution we imagine that the surface of a uniformly exposed and processed material is divided up by a grid into small equal and contiguous areas and that the densities of all these areas are measured. These densities will be different one from the other owing to the irregularity of distribution of grains. We ask, "What is the statistical distribution of density and in particular what is the root mean square fluctuation of density?" We suppose that the variation from the mean of the density of any elementary area of the grid is small enough for the density of a pair of areas taken together to be equal to the average of the densities of the pair.

If the fluctuations of density from the mean for a long set of N alternate areas of the grid be denoted by symbols like δ_1 the mean fluctuation $\Sigma\delta_1/N$ is zero and the mean square fluctuation Δ^2 is equal to $\Sigma\delta_1^2/N$. For the set of areas interposed between the preceding areas, with fluctuations denoted by symbols like δ_2 the mean fluctuation $\Sigma\delta_2/N$ is again zero and the mean square fluctuation is again equal to Δ^2 and to $\Sigma\delta_2^2/N$. The mean fluctuation for the pairs of areas taken together is $\Sigma(\delta_1+\delta_2)/2N$ and is zero and the mean-square fluctuation is $\Sigma(\delta_1+\delta_2)^2/4N$, which is equal to $\Sigma\delta_1^2/4N+\Sigma\delta_2^2/4N+\Sigma\delta_1\delta_2/2N$. If the distribution of fluctuations is symmetrical $\delta_1\delta_2$ is as likely to be positive as negative, so that in a long sequence $\Sigma\delta_1\delta_2/4N$ is zero. Thus the mean square fluctuation for doubled areas is $\Delta^2/2$. From this it follows that the root mean-square fluctuation is inversely as the square root of the elementary area of the grid, or

$$\Delta = G/\sqrt{\alpha},$$

where G is a constant, for the given sample of material, called the "granularity" and α the area of an element of the grid.

There is a peculiarity about density, since it may legitimately be regarded as having a "dimension" in the same way as mass. For in just the same way as a mass may be measured by equilibrating it against a number of arbitrary units of mass, a density may be measured by equilibrating it photometrically against a stack of units of

density, such as pieces of neutral glass all of the same density. The choice of unit is arbitrary. In terms of the definition of density, namely

$$D=\log (1/T),$$

(where D is the density and T the transmission) all this means is that we have chosen the base of the logarithm arbitrarily. A much more elaborate argument than the above, based on assuming a dimension for density and assuming that the fluctuation of density is controlled by a constant G and that the densities of neighboring areas are uncorrelated, yields both the statistical distribution and the formula

$$\Delta=G/\sqrt{\alpha}.$$

The question now arises whether it is possible to estimate the nature of the signal to the brain obtained from grainy photographic material. We suppose that the retina constitutes a grid with elements having areas determined by the disk of confusion (mainly the Airy disk) of the optical apparatus of the eye and the averaging effect of the retina, already discussed. The corresponding areas on the material itself may be taken as α^*/M^2, where α^* is the area of an element of the retinal grid projected on the material at unit magnification and M the magnification. The signal to the brain, in terms of density fluctuation, is therefore

$$MG/\sqrt{\alpha^*}.$$

The signal from the sine-curve image to the brain is, in terms of light intensity,

$$B\{1+\gamma a^*\varphi(r)F(r/M)\sin 2\pi rx\},$$

where B is the mean light intensity and x distance across the bright and dark stripes. Now the appearance of almost any object, a picture, for example, of the above sine-curve distribution of light, is practically independent of the light intensity (except for very low and very high intensities). We may therefore legitimately take the logarithm of the above expression and drop the term log B, thus giving the equivalent total density fluctuation

$$2\log_{10} e\gamma a^*\varphi(r)F(r/M),$$

provided the amplitude of the sine curve is small.

Now it is easy to show, for instance by superimposing a sample of grainy photographic material upon a grainless transparency, that the irregularity in density reduces the visibility of detail, and we therefore have to enquire what the relation is between the sine-curve density fluctuation, and the density fluctuation due to granularity at which the sine-curve density fluctuation is just visible. Since the two terms in the relation have the same dimension, namely of density, it follows immediately by the very simplest of dimensional arguments that [2]

[2] If the limiting density difference, δ, that the eye can perceive be introduced into the dimensional argument, the right-hand term must be multiplied by an arbitrary function of $\delta\sqrt{\alpha^*}/GM$. Other calculations show that this function is constant when the parameter is small.

$$2 \log_{10} e\gamma a^* \varphi(r) F(r/M) = Const.\ GM/\sqrt{\alpha^*}.$$

The resolving power is to be calculated from this by solving for r. The value obtained is dependent on the magnification. If $F(r/M)$ were proportional to M/r, the observed resolving power would be independent of the magnification, since M would then disappear from the equation. But if the magnification be chosen to obtain the maximum value of r, the value of r will remain the same for small changes in magnification. Thus the maximum value of r will be secured at the value of M/r at which the curve of $F(r/M)$ against r/M contacts a rectangular hyperbola (i. e., a curve of $Const.M/r$ against r/M). This occurs at the point at which the slope of log $F(r/M)$ against log r/M is $-45°$. This point obviously specifies a constant ratio between r and M. That is to say maximum visibility and maximum resolving power are attained when the magnification is chosen so that the apparent separation between stripes is always the same and of a certain magnitude.

In these circumstances the equation becomes

$$\gamma a^* \varphi(r) = Const.\ Gr. \tag{3}$$

The value of the constant has been estimated, from what is known about the eye, at 0.003. The value of a^* is determined by the contrast of the original test object, and the optical apparatus forming the image, according to eq 1.

Thus the complete formula for the photographic resolving power is

$$\gamma a f(r) \varphi(r) = Const.\ Gr. \tag{4}$$

A graphical illustration of the use of this formula is shown in figure 16.3, for a high contrast test object, for which a is unity. The visual resolving power in the aerial image produced by the optical apparatus is the value of r at the point of intersection A of $f(r)$ with the visual

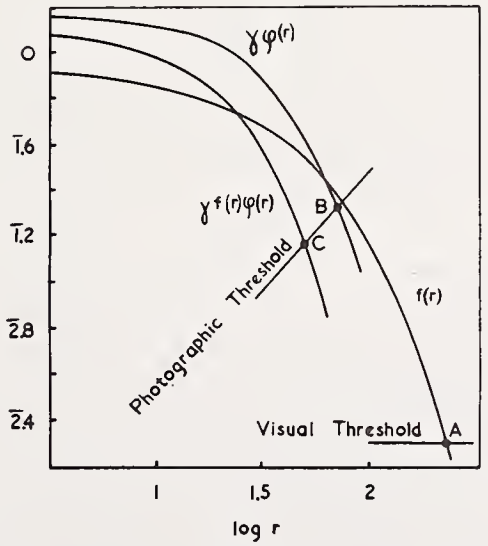


FIGURE 16.3. *Graphical illustration of theory of resolving power.*

A, Visual resolving power of optical system; B, resolving power of photographic material; C, resolving power of photographic material and optical system combined.

threshold, the resolving power of the photographic material is determined by the intersection B, at which eq 3 holds, and the resolving power in the negative produced by the optical system by the intersection C, at which eq 4 holds. It is obvious that there is no simple connection between the visual resolving power and the photographic resolving power.

Results

The value of a theory is either that it enables results to be calculated, or that it enables one to understand a phenomenon or group of phenomena. Obviously, the function of a theory of resolving power is to permit one to understand what is happening, for it is much easier to measure the resolving power in any given case than to calculate it. In spite of its defects (which will be mentioned shortly) the theory given here gives an account, of certain features found in resolving-power measurements, sufficiently nearly quantitative to justify the belief that any physical properties other than those discussed have relatively unimportant effects.

The first result to be discussed concerns the variation with aperture of the resolving power in the center of the field of a photographic objective. The aberrations of such a lens are relatively unimportant at small apertures such as F/45. The function $f(r)$ may thus be taken as identical with that for a perfectly corrected lens. When the aperture is increased a little the central core in the image of a point remains about the same in size as the Airy disk appropriate to the aperture under consideration, but there is a greater amount of light in the outer parts of the image than when the lens is perfectly corrected. The main effect of this is to decrease the value of $f(r)$ by a roughly constant factor for moderate and large values of r. As the aperture is still further increased the image becomes more extended and the central core may disappear or almost disappear. Then $f(r)$ is diminished still more, its value at large values of r diminishing greatly. An estimate of its value at various values of r may be obtained from estimates of the visual resolving power of the lens with test objects of different contrasts. It is not difficult, if this be done for the maximum aperture, to estimate graphically a set of curves of $f(r)$ for different apertures. The form of $\varphi(r)$ for photographic materials may be found if it is assumed (as is commonly done in photographic theory) that the sideways scatter of light in the sensitive layer follows a negative exponential law. The curve for any given material may then be obtained from a knowledge of the granularity G and the resolving power, using eq 3. Sufficient information is thus provided to determine, by the method illustrated in figure 16.3, the resolving power at any aperture. Results of such a calculation are shown in figure 16.4. The visual resolving power for high-contrast test objects at moderate and small apertures approximates closely to that expected from Rayleigh's formula, while the decrease caused by diminishing the contrast of the test object is not great. The resolving power at high apertures is markedly below the theoretical. With a material of low granularity the resolving power is much decreased and the maximum occurs at a smaller aperture than that at which maximum visual resolving power is found. With materials of greater granularity the resolving power is still more decreased and the curve against aperture becomes flatter. These results are of a general character

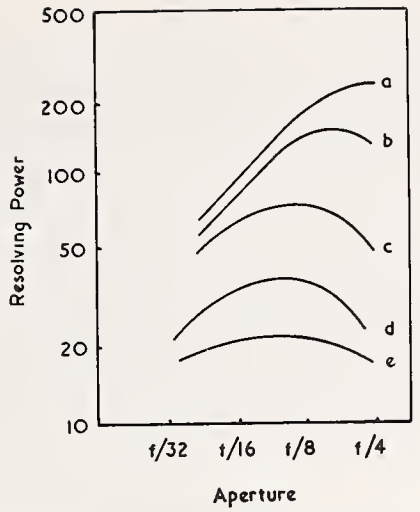


FIGURE 16.4. *Resolving power of camera objective as a function of aperture.*

a, Visual resolving power, high-contrast test object; b, visual resolving power, low-contrast test object; c, photographic resolving power, fine-grain film, low-contrast test object; d, photographic resolving power, medum-grain film, low-contrast test object; e, photographic resolving power, coarse-grain film, low-contrast test object.

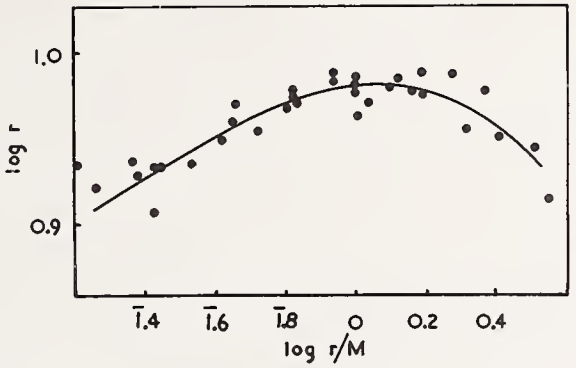


FIGURE 16.5. *Effect of magnification on resolving power.*

and have been observed with all lenses that have been tested in these laboratories. The resolving power observed varies with the density of the processed material. The above results refer to a density of 1.0, at about which density the resolving power is a maximum.

The results obtained by experiments designed to check the theory of photographic resolving power also provided evidence that the observed resolving power depended upon the magnification used. Typical results are shown in figure 16.5. The points shown consist of six sets in which the resolving power varies over a range of 5:1, but in the diagram as shown the six sets of points have been adjusted along the log r direction so as to have the same mean value. It is clear from the diagram that maximum visibility is attained when the apparent separation between bright stripes is 1 mm, at the standard viewing distance of 250 mm. As expected from theory this is a universal condition independent of the actual separation and the granularity, but calculations from $F(r/M)$ suggest a value four times less for the magnification. The reason for the discrepancy is not known, but it is known that different observers prefer different magnifications. The above condition, namely making the apparent size of detail it is

wished to see about 1 mm, is not far from correct for maximum visibility on most photographic records.

If one could provide a diffusing screen with turbidity identical with that of a photographic material, and introduce granularity into the image, the image seen should be identical with that seen on a negative developed so that $\gamma=1$. A difference of γ from unity can be allowed for by changing the contrast of the test object. Apparatus of this kind has, in fact, been devised, and although the test object consisted of two short bars of the type used by Cobb, and not the sine-wave form, the resolving powers so observed were the same as the photographic resolving powers within a standard deviation of 15 percent. This is not much more than can be accounted for by experimental and observational error. The gain in speed over photographic measurement is substantial.

Evaluation

There are certain defects in the preceding theory. The most outstanding are the disagreement between the calculated optimum magnification and that found by experiment as shown in figures 16.5, and the disagreement between the theoretical estimate of the photographic threshold, involving granularity, and experimental determinations. The latter are shown in figure 16.6. Curve B is the theoretical result (eq. 3). Curve A refers to a few estimates made by determining corresponding values of G and r at which images of Cobb test objects projected optically on samples of uniformly exposed and processed photographic materials ceased to be visible. Curve C refers to exposures made with sine-curve test objects of different contrast analysed in such a way as to give the threshold almost directly. Full details are given in the original publication. There is agreement in order of magnitude and in the general result that the photographic threshold increases with Gr approximately according to a power law with an exponent of the order of unity. Perhaps this is as much as can be expected from photographic experiments of a type that is notoriously liable to both systematic and accidental errors. There are, however, stronger reasons than these for supposing that

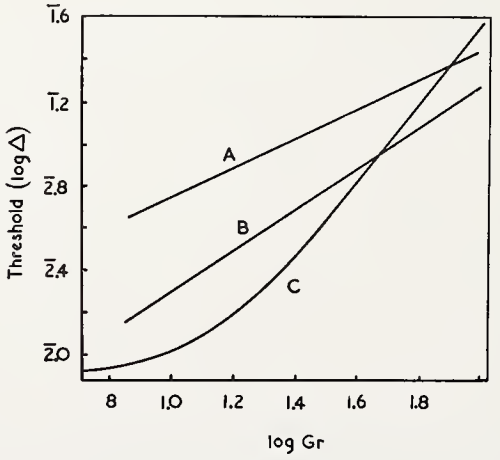


FIGURE 16.6. *Photographic threshold (log Δ) as a function of Gr.*

the theory may be defective. Jones and Higgins have demonstrated that granularity figures, obtained by the method used (theoretically described in principle in this paper) do not correlate well with figures for the graininess, the corresponding quantity estimated by purely visual observation, and have also shown that the theoretical description is not followed when α is small. It seems likely that the density of any given small area on the photographic material is not independent of that of neighboring areas. In other words there is correlation, which may formally be described by a quantity with the dimension of length, corresponding with the distance over which the correlation extends. The theory as given takes no account of such a quantity, nor indeed does any theoretical treatment of granularity yet proposed. Correlation might give the grainy structure a character akin to periodicity and this could well affect both the threshold and the optimum magnification.

References

E. W. H. Selwyn, Phot. J. [B] **88,** 6, 46 (1948).
P. W. Cobb, Am. J. Physiol. **36,** 335 (1915).
E. W. H. Selwyn, Phot. J. **75,** 571 (1935).
E. W. H. Selwyn and J. L. Tearle, Proc. Phys. Soc. **58,** 493 (1946).
W. Romer and E. W. H. Selwyn, Phot. J. **83,** 17 (1943).
L. A. Jones and G. C. Higgins, J. Opt. Soc. Am. **35,** 435 (1945); **36,** 203 (1946).

518-743 O - 73 - 29

17. A New System of Measuring and Specifying Image Definition

By O. H. Schade [1]

Introduction

An objective analysis of image definition leads fundamentally to methods for determining the geometrical properties or an equivalent measure of the point or "star" image of the imaging device. When the intensity distribution of the point image is invariable, it is obvious that image quality and apparent sharpness vary inversely with a significant diameter of the point image. When point images having different intensity distributions are compared, however, this relation is no longer true. It then becomes necessary to specify the relative intensity distribution as well as a diameter. The problem is further complicated by the fact that the point image of practical lenses departs considerably from circular symmetry. Figure 17.1 illustrates the changes in the geometry of the axial-point image in successive stages of a motion-picture process. From a study of resolution and detail contract characteristics of photographic and television images it has long been apparent that the sharpness of an image has no fixed relation to the limit of resolution of the system but depends rather on the steepness and form of the intensity or luminance curve representing a unit function transition, i. e., a sharp edge. It is not difficult to see that the total length of the transition is equal to the diameter of the point image (fig. 17.1), while shape and gradient depend on the *energy distribution* in the point image. It must be appreciated that the energy in a point image is a function of its volume (considering intensity of the third dimension), and that a low-intensity disk of light surrounding a small high-intensity center may contain as much or more light flux as the bright center of the light spot and thus produce a gradual transition with rounded corners. The transition curve that is readily generated by displacing the point image over a brightness step is, therefore, a much more sensitive measure of quality than the intensity distribution in the point image. The maximum gradient of a single transition, however, is not a sufficient specification of image quality. Furthermore, the shape of the transition is affected when several edges (or repetitive contours) occur within a distance less than the maximum spot diameter. It is known from communication theory that any complex transition or waveform can be expressed by its Fourier components. The properties of the point image can thus be specified accurately by a response characteristic to optical sine waves ranging in wavelength from infinity to zero.

---

[1] Radio Corp. of America, Tube Department, Harrison, N. J.

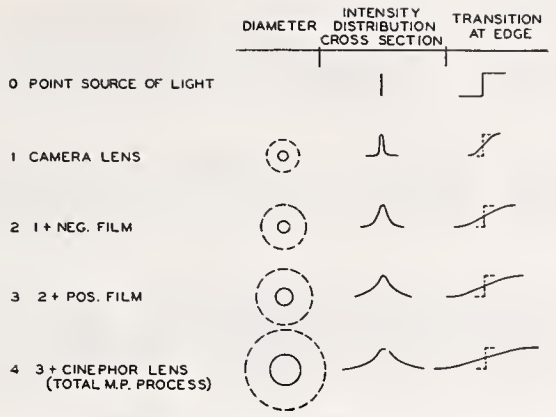


FIGURE 17.1. *Point images and edge transitions in a photographic process.*

Response Characteristics of the Point Image

Normal optical-bar test patterns are "square-wave" flux patterns. Accurate sine-wave test patterns can be obtained from variable-density recordings of electrical constant-amplitude sine-wave frequencies on the sound track of motion-picture film. A number of sections so obtained are shown in figure 17.2. It is customary to express the wavelength in these patterns in reciprocal units: the *line number* N. The line number N will be defined in this paper as the number of half-waves or "lines" (dark and light) in a length unit. The line number has the dimension length^{-1}. When the geometrical properties of the point image are known, its response to sine-wave or square-wave flux patterns is readily computed by scanning the flux pattern with the point image, considering it as a scanning aperture. The point image is thus defined as the *resolving aperture* or *sampling aperture* of the image-forming device referred to the image plane, the intensity distribution in the point image is represented by the transmittance τ of the aperture.

Response characteristics of a number of aperture types are shown in figures 17.3 to 17.8. The line number is given in relative units N_δ defined by the equation

$$N_\delta = l/\delta, \tag{1}$$

where l is the unit of length (l=1 mm) and δ is a significant diameter of the aperture. The sine-wave or square-wave response is likewise given in relative units. The sine-wave response factor $r_{\tilde{\psi}}$ is defined as the ratio of the sinusoidal aperture flux $\tilde{\psi}_N$ at a line number N to the sinusoidal flux $\tilde{\psi}_0$ at a line number N approaching zero as a limit as expressed by

$$r_{\tilde{\psi}} = \tilde{\psi}_N/\tilde{\psi}_0. \tag{2}$$

The square-wave flux response factor $r_{\Delta\bar{\psi}}$ is defined similarly by

$$r_{\Delta\bar{\psi}} = \Delta\bar{\psi}_N/\Delta\bar{\psi}_0. \tag{3}$$

The symbol $\Delta\bar{\psi}$ indicates that the square-wave response is measured not by the peak amplitude [2] of the waveform but by the differential

[2] The square-wave amplitude response $r_{\Delta\hat{\psi}}$ is shown for comparison in some figures.

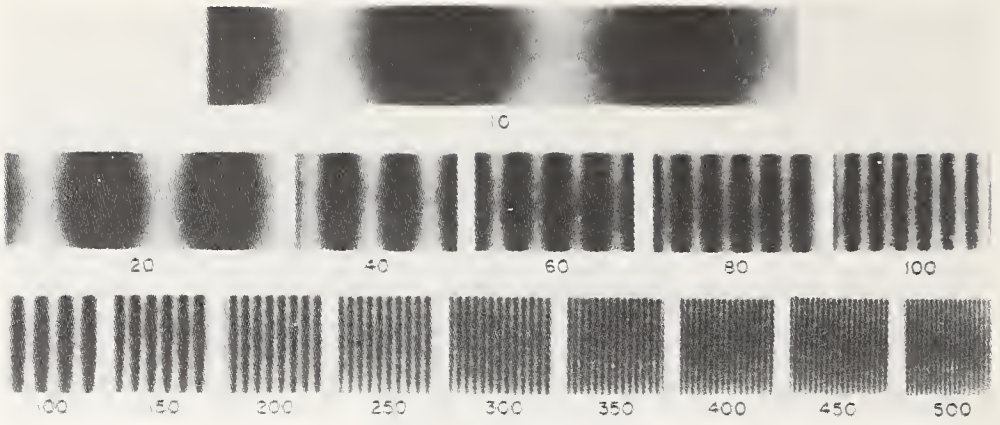


FIGURE 17.2. *Sine-wave test pattern.*

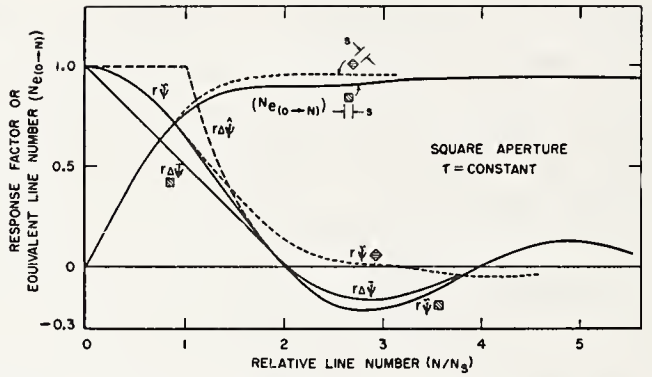


FIGURE 17.3. *Response characteristics of square aperture* ($\tau = 1$).

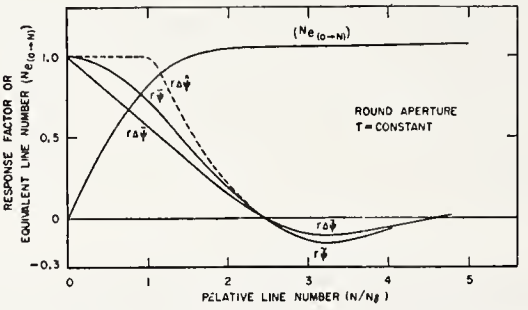


FIGURE 17.4. *Response characteristics of round aperture* ($\tau = 1$).

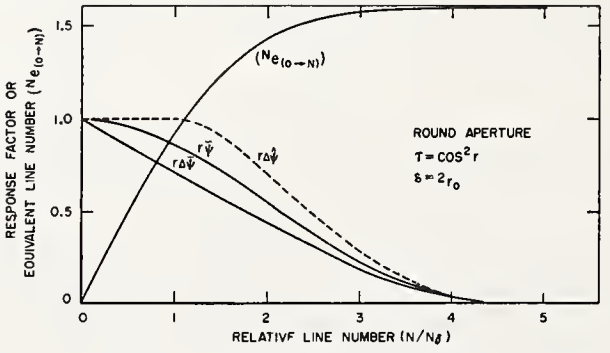


FIGURE 17.5. *Response characteristics of round aperture* ($\tau = \cos^2 r$).

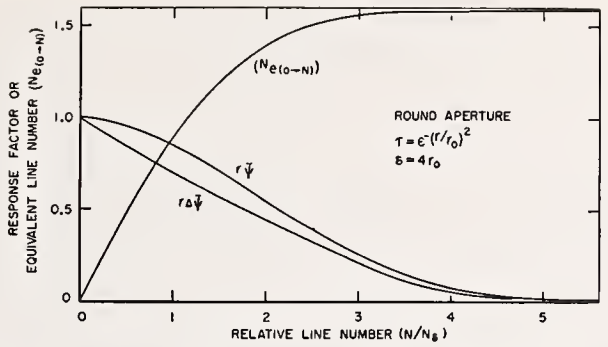


FIGURE 17.6. *Response characteristics of round aperture* ($\tau = \epsilon^{-(r/r_0)^2}$).

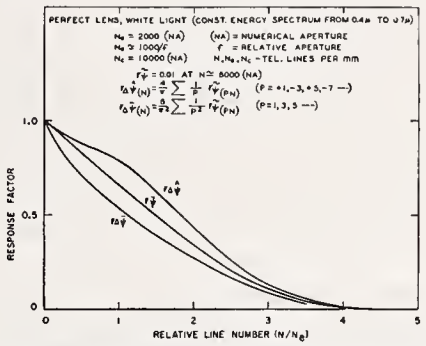


FIGURE 17.7. *Response characteristics of theoretical lens (white light).*

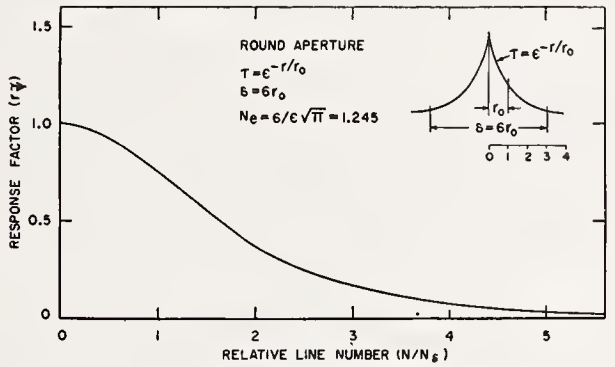


FIGURE 17.8. *Response characteristics of round aperture* ($\tau = \epsilon^{-r/r_0}$).

flux, which is proportional to the mean value or area under the rectified half-waves of the waveform, because the latter changes from a rectangle to trapezoids and triangles or sinusoidal shapes when N is varied from zero to infinity. Both response factors are single valued and independent of contrast, provided waveform distortion is avoided by the use of small "signals" when nonlinear devices such as photographic films are measured. The use of sine-wave patterns has the advantage of permitting a direct and accurate evaluation of the overall response of systems containing a number of imaging stages in cascade. The over-all sine-wave response factors of the system are simply the product of the response factors of the components at corresponding line numbers. This advantage is lost for square-wave response factors, because of the variable harmonic content of the waveforms. Inspection of the computed characteristics,

figures 17.3* to 17.8, however, shows that the ratio $r_{\bar{\psi}}/r_{\Delta\bar{\psi}}$ is substantially independent of the aperture type at given values $r_{\Delta\bar{\psi}}$, permitting a point by point conversion of square-wave to sine-wave response characteristics with satisfactory accuracy.

The response characteristic of a theoretical aberration-free lens limited only by diffraction is of interest as a standard of comparison. The characteristic (fig 17.7) shows the "white"-light response computed for a coaxial superposition of diffraction spots for a constant energy spectrum for 0.4 μ to 0.7 μ.[3] The round aperture with a transmittance $\tau=\epsilon^{-r/r_0}$ (fig. 17.8) is of interest as a mathematical equivalent for the aperture effect of grain structures in which the intensity from an infinitesimal pencil of light decreases exponentially from its center outward because of diffusion, absorption, and diffraction in the grain layer; the magnitude of these optical effects (diameter of the aperture) depends on the layer thickness, transparency, spacing, and size of the particles. The theory is well substantiated by measurements on a variety of film types and kinescope phosphors.[3] Plotted to a normalized scale with reference to the rated resolving power N_{cr} of photographic film[4] at which $r_{\bar{\psi}}\simeq 0.02$, the measured values result in a single curve shown in figure 17.9. A comparison with figure 17.8 shows excellent agreement when N_{cr} is placed at $N/N_\delta=5.17$.

Another characteristic of special interest is that of the round aperture with $\tau=\epsilon^{-(r/r_0)^2}$ (fig. 17.6), which has the form $r_{\bar{\psi}}=\epsilon^{-kN^2}$. It can be seen from this form that a repetition or cascading of such aperture processes will not change the shape of the characteristic, but only the constant k.

The response characteristics of practical lenses, microscopes, television tubes, and other image-forming devices can be measured accurately by photoelectric methods, either by scanning test patterns with their point image or by using a photoelectric or electronic microphotometer [3] to scan the test-pattern image formed by the device. One of these latter methods developed by the author appears to be particularly simple and flexible for testing lenses and was described and demonstrated.

The quantitative information obtained from response characteristics is demonstrated by table 17.1 showing the changes in the square-

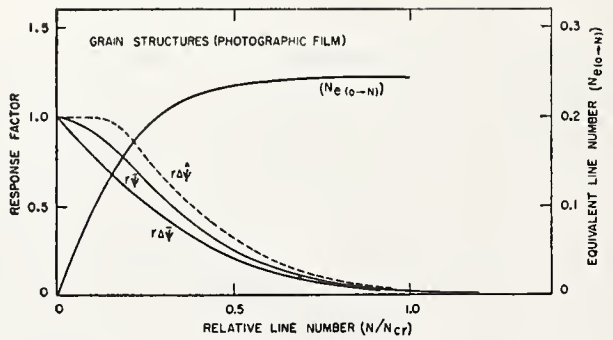


FIGURE 17.9. *Response characteristics of grain structures (measured).*

[3] Compare first reference on page 249.
[4] The photographic value N_{cr} (lines/mm) is to be multiplied by a factor of two according to the definition of N used in this paper.
* The broken line curve for $N/N_\delta > 2.8$ should be positive in figure 17.3.

TABLE 17.1 *Square-wave response ($r_{\Delta\bar{\psi}}$) of 8.3-mm apochromat, NA=0.65 with 10x compensating eye piece*

(White light)

N/mm	Theoretical*		$r_{\Delta\bar{\psi}}$				
	r_{Ψ}	$r_{\Delta\bar{\psi}}$	Cover glass.....0.18 Tube length: (mm) 160	0.127 160	0.127 172	None 160	None 196
100	0.98	0.95	0.925	0.91	0.925	0.76	0.89
250	.94	.88	.81	.78	.81	.49	.76
500	.87	.78	.655	.57	.648	.34	.585
700	.83	.72	.565	.45	.555	.25	.485
1,000	.75	.64	.45	.31	.44	.195	.37
1,500	.64	.53	.30	.17	.30	.13	.225
2,000	.52	.43	.21	.13	.21	.08	.135
2,500	.42	.34	.14	.12	.14	.03	.07
3,000	.32	.26	.095	.11	.095	.01	.033
3,500	.22	.18	.055	.065	.055	.0	.018
4,000	.14	.11	.036	.038	.036	.0	.002
4,500	.07	.06	.018	.018	.018	.0	.0
5,000	.03	.025	.008	.008	.008	.0	.0
5,500	.01	.008	.0	.0	.0	.0	.0

Computed for 2,870 K source and 1P21 multiplier phototube.

wave response ($r\Delta\bar{\psi}$) of an 8.3-mm Bausch & Lomb Apochromat ($N.A.=0.65$) used with a 10x compensating eye piece at a magnification of 300 when cover-glass thickness and tube length are varied. The values have not been corrected for aberrations in the measuring microscope (4-mm Apochromat $N.A.=0.95$ with 25x compensating eye piece). It is of interest to the lens designer to note that a response characteristic (sine wave or square wave) can be considered as the sum of a number of component response characteristics of known shape thereby permitting a reconstruction of the point image. This process is illustrated on the characteristic of a 40-mm Ciné Ektar lens shown in figure 17.10. The measured characteristic (at 5°) can be regarded as the sum of response characteristics from $\cos^2$ apertures. The first component, curve 2, is obtained by continuing the "tail" end of the curve as a $\cos^2$ response curve. Subtraction of this component from the normal characteristic leaves curve 1, which closely matches the response characteristic of a second $\cos^2$ aperture. The flux division (ψ_1 and ψ_2) between the corresponding coaxial spots is indicated by the response at $N=0$ and happens to be equal for both components. The diameters of the corresponding $\cos^2$ spots follow from the respective N_δ values (eq 1); the relative flux intensities

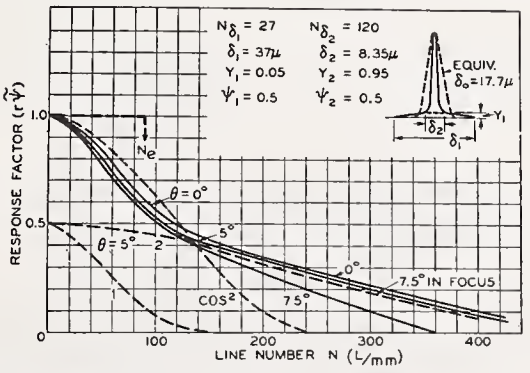


FIGURE 17.10. *Response characteristics of a camera lens for 16 millimeter film.*

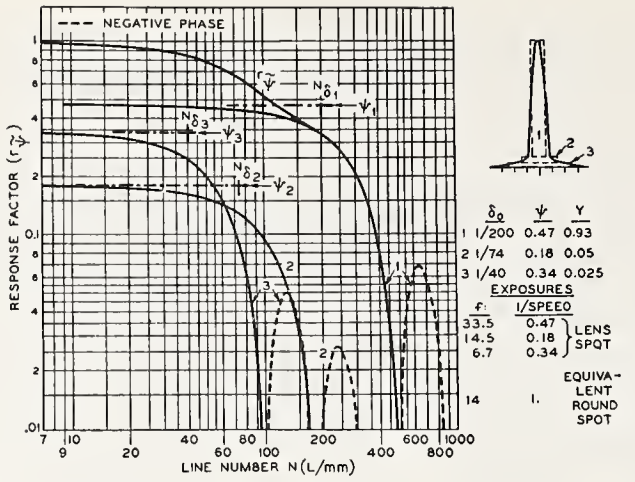


FIGURE 17.11. *Synthesis of response characteristic from three round aperture characteristics* ($\tau = 1$).

(Y_1 and Y_2) at the spot centers are obtained with

$$Y_1/Y_2 = \psi_1\delta_2^2/\psi_2\delta_1^2 \quad \text{and} \quad Y_1 + Y_2 = 1. \tag{4}$$

Addition of the cosine-squared intensity curves of the components furnishes the approximate intensity in the point image from the lens. A practical method for performing this analysis is shown in figure 17.11 illustrating a breakdown into round apertures with uniform transmittance that appear as coaxial cylinders in a topographic representation of the point image. The response characteristic of the lens is drawn in log-log coordinates (curve $r_{\tilde{\psi}}$ in fig. 17.11). The response characteristic of the round aperture (fig. 17.4) is drawn on a separate sheet of log-log paper. The two curves are superimposed and a position matching the tail end of the lens curve is found. This portion (curve 1 in fig. 17.11) is subtracted and the process is repeated. In this particular case, the two additional component curves shown result. Flux values, diameters, and intensities are then found as outlined above and shown in the insert of figure 17.11. It should be noted that the negative-phase portions of the round aperture responsemust be included with negative sign when making the summation. The process is reversible, illustrating a synthesis of the sine-wave response from a complex symmetric aperture as the sum of three component characteristies. It is not difficult to see that the response of an asymmetric aperture can be obtained by the same process of subdivision into cylindrical sections having displaced axes. This displacement is equal to the displacement of all sine waves between respective response characteristics, and causes a phase shift proportional to N that must be duly considered when the summation is made.

Evaluation of a Measure of Equivalence (N_e) from the Response Characteristic

The response characteristics of practical imaging devices depart in general more or less from theoretical characteristics and may differ considerably in the relative response between low and high line numbers. In the case of optical lenses the response may decrease more

or less rapidly at low line numbers because of aberrations, although the theoretical resolving power may still be measurable in some cases.

The differences in response characteristics and the condition of astigmatism indicates that a comparison of image definition by a single figure of merit requires the evaluation of a measure such as an equivalent response characteristic or an *equivalent aperture* with standard transmittance that can be specified by one significant number. It is important that this equivalent measure agree with a visual impression of sharpness. In an objective evaluation the "quality" of an image or, more general, the quality of a device or system transmitting information (code, sound, or pictures) is specified by three fundamental characteristics: Transfer characteristic, fluctuation or "noise" level, and sine-wave response characteristic. The seemingly corresponding visual impressions of tone range, graininess, and sharpness in optical images, however, are not exact equivalents, because they represent combinations of the above objective characteristics. A judgment of contour sharpness, for example, is influenced by graininess and by the amplitude of the brightness step (contrast), whereas the sine-wave response characteristic is a *relative measure* describing shape and length of the brightness transition only. In an objective system of ratings, amplitudes, transmittance, "flare", wave-form distortion, rectification, etc., are specified or determined separately by the transfer characteristic, which, in general, is not controlled by the same parameters determining the frequency response of a system. A low-reflection coating of lens elements, for example, improves the light-transfer characteristic of the lens, but has no effect on the sine-wave response characteristic. In order to agree with an objective measure of definition, a visual comparison must be arranged for judging the *relative sharpness* of images and requires that image content, transfer characteristic (luminance and contrast) and fluctuation level in the images are made substantially identical. The impression of relative sharpness in an image can then be considered as a visual evaluation of the complex waveforms from a large number of arbitrary cross sections of the intensity distribution in an image frame. In an ideally sharp image the sum of the Fourier components of these cross sections fills an infinite sine-wave spectrum with constant amplitudes. Superposition of all components in random phase relation results in a most general and interesting test pattern, appearing optically as a *random grain structure*. Any cross section of this structure taken by a microphotometer with infinitesimal aperture is a complex wave containing a constant-amplitude sine-wave spectrum

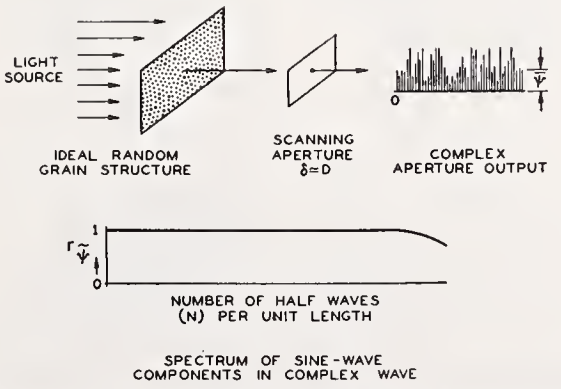


FIGURE 17.12. *Fourier components in ideal random grain structure.*

up to a very high line number, as illustrated by figure 17.12. An imaging device with a finite scanning or sampling aperture integrates high-"frequency" components, as illustrated by figure 17.13, a harmonic analysis furnishing the sine-wave response characteristic of the device. The complex wave contains a variational or modulation component $[\tilde{\psi}]$ that can be measured in total by an rms current meter, and a steady component $\bar{\psi}$ that can be measured by a d-c meter. Their ratio, known electrically as the "noise-to-signal" ratio, is *the relative deviation*

$$\sigma=[\tilde{\psi}]/\bar{\psi}=\left[\int_0^\infty \tilde{\psi}^2_{(N)}\,dN\right]^{\frac{1}{2}}\Big/\bar{\psi}. \tag{5}$$

When the sine-wave response is normalized according to eq 2 so that $\tilde{\psi}_{(N)}=1$ at $N=0$, the mean squared deviation is expressed by

$$[\tilde{\psi}]^2=\tilde{\psi}_0^2\int_0^\infty (r_{\tilde{\psi}})^2_{(N)}dN, \tag{6}$$

where $r_{\tilde{\psi}}$ is the sine-wave response factor and $\tilde{\psi}_0$ is a measure of the magnitude of the a-c flux passing through the aperture at a line number N approaching zero. A hypothetical aperture having a constant response ($r_{\tilde{\psi}}=1$) from $N=0$ to a line number N_e^* [5] where the response drops abruptly to zero, would give a mean-squared deviation

$$[\tilde{\psi}]^2=\psi_0^2 N_e^*.$$

The integral of squared response factors in eq 6 may hence be interpreted as a normalized mean-squared deviation or as an *equivalent passband* of constant amplitude extending to the line number N_e^* as defined by

$$N_e^*=[\tilde{\psi}]^2/\tilde{\psi}_0^2=\int_0^\infty (r_{\tilde{\psi}})^2_{(N)}dN. \tag{7}$$

The measure N_e^* has the dimension length^{-1}. Its reciprocal value expresses an equivalent length or diameter of the aperture *in the scanning direction*. Like the aperture response, N_e^* depends, in general, on the aperture orientation relative to the direction x of aperture displacement. Apertures with circular symmetry have a

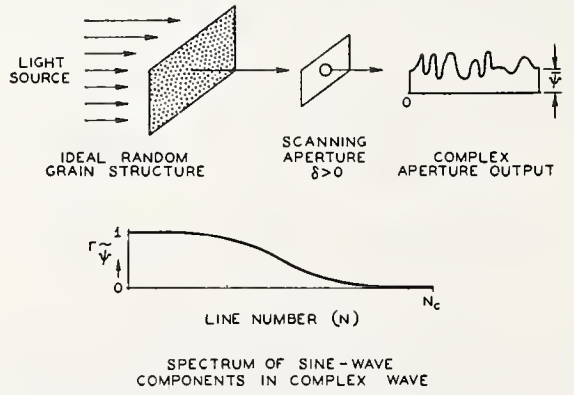


FIGURE 17.13. *Fourier components in grain structure integrated by a finite aperture.*

[5] The asterisk on the value N_e^* is used to indicate that this value is obtained when a random grain structure is scanned. Other values will be introduced subsequently.

single effective length proportional to their diameter δ and a single value N_e^*. Elliptical or rectangular apertures can be specified by two values $N_{e(a)}^*$ and $N_{e(b)}^*$ obtained by orienting their major or minor dimensions (a or b) in the direction of scanning. These two values can be combined into a single value

$$\overline{N}_e^* = (N_{e(a)}^* N_{e(b)}^*)^{1/2} \tag{8}$$

representing an equivalent symmetric aperture.

The direct evaluation of the measure N_e^* for an unknown aperture requires a calibrated random grain pattern that must be tested by a harmonic analysis of the complex aperture output. A practical alternative is a synthesis of the sine-wave characteristic from the aperture response to constant amplitude sine-wave patterns of various wavelengths and an evaluation of N_e by eq 7. Optical sine-wave patterns consisting of parallel "lines", however, do not duplicate exactly the sine-wave components in a random flux pattern, but rather in a pattern that is random only in the direction x of scanning and uniform in the direction y, perpendicular to the scanning direction. Figure 17.14 illustrates the difference in cross sections through a random grain structure and a synthetic structure representing a random addition of sine-wave test patterns (random phase relation). The differences resulting from scanning a random grain structure or sine-wave test patterns and the suitability of N_e-values, in general, for the purpose of indicating an equivalent aperture area can be determined by a comparison with an equivalent $\overline{N}_0$ based on the *sampling* of a normalized random structure. The various equivalents N_e^*, N_e, and $\overline{N}_0$ can be computed without recourse to response characteristics when the geometrical properties of the aperture are known.

The effective sampling area of an aperture (pictured as a three-dimensional body, the aperture transmittance τ representing height) may be determined by subdividing the aperture into differential columns (fig. 17.15) with a base area $\Delta a = \Delta x \Delta y$ and constant or varying height representing the transmittance $\tau = f(x,y)$. The relative deviation obtained by taking a large number of samples from a random

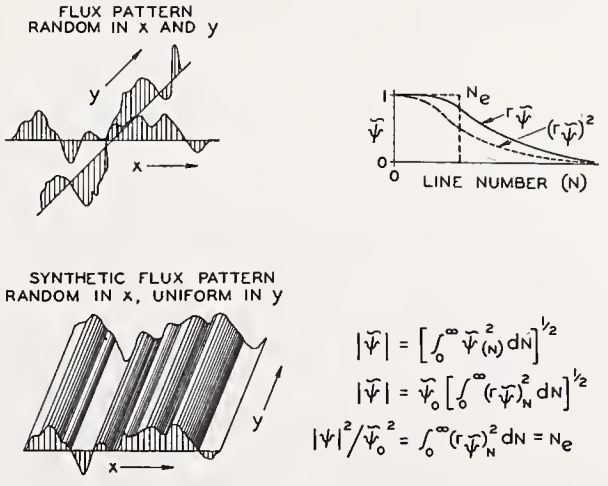


FIGURE 17.14. *Cross sections of random and synthetic grain structures (see text).*

grain pattern with one differential column is $\Delta\sigma=(\tau^2\overline{n}_0\Delta a)^{1/2}/\tau\overline{n}_0\Delta a$, where $\overline{n}_0$ equals average number of grains per unit area. For a normalized grain density $\overline{n}_0=1$, the above relation becomes $\Delta\sigma_0=(\tau^2\Delta a)^{1/2}/\tau\Delta a$. Integration over the aperture yields the normalized relative deviation

$$\sigma_0=\frac{[\lim\sum\tau^2\Delta a]^{1/2}}{\lim\sum\tau\Delta a}=\frac{[\int\int f^2(x,y)dxdy]^{1/2}}{\int\int f(x,y)dxdy}. \tag{9}$$

The normalized relative deviation σ_0 has the dimension length^{-1}. The length may be regarded as the geometric mean of the sides of an equivalent rectangular sampling area a_e having constant transmittance $\tau=1$. According to eq 1 the relative deviation $\sigma_0=1/(a_e)^{1/2}$ can also be interpreted as the line number $\overline{N}_0$ of an equivalent square sampling aperture and eq 9 may, hence, be stated in the form

$$\overline{N}_0=\frac{[\int\int f^2(x,y)dxdy]^{1/2}}{\int\int f(x,y)dxdy}. \tag{10}$$

The measure $\overline{N}_0$ is independent of the aperture orientation for both symmetric and asymmetric apertures and can, hence, be used as a standard for comparison. The equivalent passband N_e^* of an aperture scanning a grain structure random in x and y directions can be computed by subdividing the aperture into incremental sections parallel to the direction of scanning (fig. 17.15). The mean-squared flux obtained is the same as that obtained when the aperture is sampling. The flux $\tilde{\psi}_{0(y)}$ at $N=0$ contributed by each section to $\tilde{\psi}_0$ is represented by the areas $\int\tau dx$ of the sections, and because the flux is random (out of phase) in y, the total flux $\tilde{\psi}_0^2$ is obtained by the sum of the squares $\tilde{\psi}_0^2=\sum[\int\tau dx]^2$. The measure N_e^* obtained when a random grain pattern is scanned is, therefore,

$$N_e^*=[\tilde{\psi}]^2/\tilde{\psi}_0^2=\frac{\int\int f^2(x,y)dxdy}{\int dy[\int f(x,y)dx]^2}. \tag{11}$$

The asterisk is used to distinguish N_e^* from the value N_e that will henceforth be used to indicate a sine-wave synthesis.

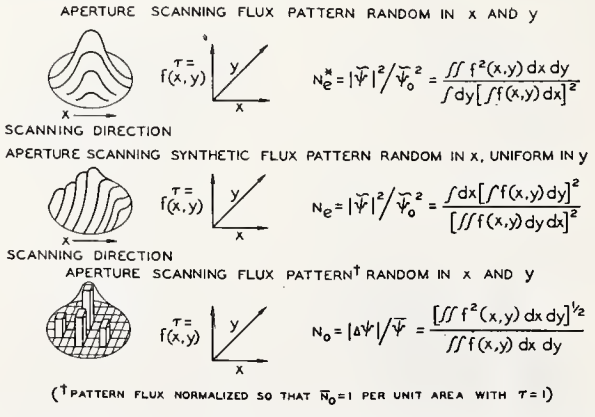


FIGURE 17.15. *Subdivision of apertures for evaluation of the equivalent measures N_e N_e^* and $\overline{N}_o$.*

Evaluation of the equivalent passband N_e from a response characteristic obtained by the method of scanning sine-wave test patterns represents the case in which a synthetic structure random in the x direction but uniform (in phase) in the y direction is scanned. The aperture is subdivided into sections parallel to y. The mean-squared flux $[\tilde{\psi}]^2$ is the sum of the squares of the section flux values $\sum[\int \tau dy]^2$, and the flux $\tilde{\psi}_0^2$ is the squared sum of the section flux values, furnishing the ratio

$$N_e = [\tilde{\psi}]^2/\tilde{\psi}_0^2 = \frac{\int dx[\int f(x,y)\,dy]^2}{[\iint f(x,y)\,dy\,dx]^2}. \tag{12}$$

All measures $\overline{N}_0$, N_e^*, and N_e represent dimensionally a length^{-1}, but the formulations appear to have little resemblance to one another. Because the measures N_e^* and N_e depend on the direction of scanning, asymmetric apertures require evaluation of two N_e-values as stated by eq 8. For apertures having circular symmetry, however, the sampling equivalent $\overline{N}_0$ is seen to equal the geometric mean $(N_e^*\,N_e)^{1/2}$. To evaluate the relative accuracy of the three measures it is of interest to determine how closely the values computed with eq 10, 11, and 12 compare in a number of representative cases. To obtain N_e in relative units N_e/N_δ, the above equations must be multiplied by the ratio of the characteristic lengths δ/u when the length u chosen for computing the measure N_e differs from the length expressing a characteristic diameter of the aperture. In relative units eq 10, 11, and 12 can be written.

$$\overline{N}_0/N_\delta = \frac{\delta}{u}\,\frac{[\iint f^2(x,y)\,dx\,dy]^{1/2}}{\iint f(x,y)\,dx\,dy}, \tag{10}$$

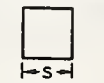
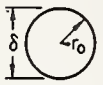
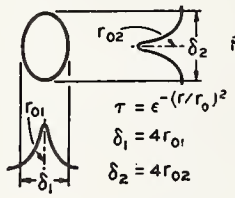
$$N_e^*/N_\delta = \frac{\delta}{u_x}\,\frac{\iint f^2(x,y)\,dx\,dy}{\int dy\,[\int f(x,y)\,dx]^2}, \tag{11}$$

$$N_e/N_\delta = \frac{\delta}{u_x}\,\frac{\int dx\,[\int f(x,y)\,dy]^2}{\iint f(x,y)\,dy\,dx]^2}. \tag{12}$$

It must be kept in mind that the length u in eq 10 is the square root of an area and, therefore, independent of the aperture orientation. The length u_x in eq 11 and 12 however, is always the characteristic aperture length in the direction x of scanning.

The measure N_e for a round aperture with $\tau = \epsilon^{-(r/r_0)^2}$, for example, may be computed in terms of a radius length $r_o = u_x$; the corresponding relative line-number unit N_δ in figure 17.6 represents a length^{-1} measured by the diameter $\delta = 4r_0$. The ratio δ/u_x in this case is, therefore, four. The relative values in table 17.2 show that the sine-wave equivalent N_e is as good an equivalent as the value N_e^* obtained by the scanning of a random grain structure. Both values are somewhat in error for a round aperture with $\tau = 1$ and for a square scanning diagonally. Practical apertures such as lenses, grain structures, or electron beams have nonuniform transmittances similar to the aperture types 4 to 6 in table 17.1, for which the error is negligible or zero. The definition of N_h^* as the integral of squared response factors given by eq 6 applies also to the measure N_e, which is obtained from a

Table 17.2. Relative passband-equivalents of apertures

APERTURE TYPE	$\tau = f(x,y)$	δ	$\bar{N}_0/N\delta$	$N^*_e/N\delta$	$N_e/N\delta$
1 SQUARE 	$\tau = 1$	S	1	1	1
2 SQUARE 	$\tau = 1$	S	1	$\frac{1}{(2/3)\sqrt{2}} = 1.06$	$(2/3)\sqrt{2} = 0.943$
3 ROUND 	$\tau = 1$	$2r_0$	$\frac{2}{\sqrt{\pi}} = 1.13$	$\frac{3\pi}{8} = 1.178$	$\frac{32}{3\pi^2} = 1.08$
4 ROUND 	$\tau = 1$ $\cos^2 r$	$2r_0$	1.575	1.56	1.59
5 ROUND 	$\tau = \epsilon^{-r/r_0}$	$6r_0$	$\frac{6}{\epsilon\sqrt{\pi}} = 1.245$	—	1.245
6 ROUND 	$\tau = \epsilon^{-(r/r_0)^2}$	$4r_0$	$\frac{4}{\sqrt{2\pi}} = 1.596$	1.596	1.596
7 RECTANGLE 	$\tau = 1$		$\bar{N}_0 = \frac{1}{\sqrt{ab}}$	$N_e(a) = \frac{1}{a}$; $N_e(b) = \frac{1}{b}$	$\bar{N}_e = \frac{1}{\sqrt{ab}}$; FOR $a = \delta$; $\frac{\bar{N}_e}{N_\delta} = \sqrt{\frac{a}{b}}$
8 ELLIPSE 	$\tau = \epsilon^{-(r/r_0)^2}$	$\delta_1 = 4r_{01}$; $\delta_2 = 4r_{02}$	$\bar{N}_0 = \frac{1}{\sqrt{r_{01} r_{02} 2\pi}}$	$N_e(1) = \frac{1}{r_{01}\sqrt{2\pi}}$; $N_e(2) = \frac{1}{r_{02}\sqrt{2\pi}}$	$\bar{N}_e = \frac{1}{\sqrt{r_{01} r_{02} 2\pi}}$; $\frac{\bar{N}_e}{N_{\delta 1}} = 4\sqrt{\frac{\delta_1}{\delta_2 2\pi}}$

sine-wave synthesis; i. e.,

$$N_e = \int_0^\infty (r\tilde{\psi})^2_N \, dN. \tag{13}$$

The results obtained by a numerical integration of the squared aperture response according to eq 13 are illustrated by the curves $N_{e(0\to N)}$ in figures 17.3 to 17.9 that show the growth of the partial integral when the limit is increased from $N=0$ towards $N=\infty$. The accurate agreement of the values obtained by this method is a check on the accuracy of the sine-wave response characteristics as well as the formulation of eq 12. The ϵ^{-r/r_0} aperture is of interest as a mathematical equivalent for grain structures with finite thickness. The line-number scale of this aperture is referred to a diameter $\delta = 6r_0$, which for identical values N_e places the rated resolution N_{cr} of film at the value $N_{cr}/N_\delta = 1.245/0.241 = 5.17$ of the theoretical characteristic. A comparison of figure 17.8 and 17.9 shows an almost perfect agreement of the sine-wave response characteristics. The resolving or sampling aperture of grain structures is, therefore, well represented [6]

[6] A finite grain size removes the pointed tip of the aperture transmittance. The effect, however, is negligible because the flux contributed by a transmittance exceeding the value $\tau = 0.65$ ($r = 0.6\, r_0$) is only 2.5 percent of the total flux.

by a round aperture with a transmittance $\tau=\epsilon^{-r/r_0}$. The value $\overline{N}_e$ of an asymmetric aperture of width a and height b can be determined accurately when the deformation of the dimensions a or b from circular symmetry does not alter the relative aperture transmittance in the b or a dimension, respectively. In this case the sine-wave measure $N_{e(a)}$ or $N_{e(b)}$ obtained with eq 12 is determined by the dimension of the aperture (a or b), which is oriented parallel to the direction x of scanning, the measure being independent of the aperture scale factor in the y direction. The aperture is thus simply considered first as an aperture with circular symmetry and a diameter $\delta=a$, furnishing the value $N_{e(a)}$, and second as an aperture with the diameter $\delta=b$, furnishing the value $N_{e(b)}$. The geometric mean of these values (eq 8) furnishes the symmetric equivalent $\overline{N}_e$. The corresponding procedure when the sine-wave response of an astigmatic lens is measured, for example, requires orientation of the sine-wave pattern and scanning direction parallel or perpendicular to the direction of astigmatism. The values $N_{e(a)}$ and $N_{e(b)}$ are then determined by numerical integration from the two corresponding sine-wave response characteristics (eq 13). The evaluation of $\overline{N}_e$ is illustrated by two examples in tables 17.3 and 17.4

The numerical evaluation of the measure N_e from a sine-wave response characteristic by means of eq 13 is illustrated by table 17.3 for a 40-mm f:1.6 Ciné Ektar lens measured at f:1.6 and 5° off axis. The value $\bar{a}$ is the mean response factor within the increment ΔN. The equivalent passband N_e is obtained directly in television lines per millimeter; $N_e=90$ L/mm. Table 17.4 illustrates the evaluation of N_e for grain structures from figure 17.9 in relative units. With reference to the rated resolving power N_{cr} of film, $N_e=0.241\ N_{cr}$. Hence, for fine-grain positive film (type 5302) with $N_{cr}=180$ *television* lines per millimeter, $N_e=43.4$ L/mm.

It may be of interest to the lens designer that the measure N_e can be estimated from the diameter $\delta_{0.5}$ of the physical aperture passing 50 percent of the light flux in the star image. The relation

TABLE 17.3. *Evaluation of N_e for 40-mm Ciné Ektar lens at f/1.6 (5°)*

$N_e=90$ lines per millimeter

N/mm	$r\tilde{\psi}$	$\bar{a}$	$\bar{a}^2\Delta N$	$\Sigma(\bar{a})^2\Delta N$
10	0.98	0.99	9.8	
20	.94	.96	9.2	
30	.90	.92	8.5	
40	.85	.88	7.7	
50	.79	.82	6.75	41.95
60	.74	.765	5.85	
70	.67	.70	4.9	
80	.62	.65	4.22	
90	.57	.59	3.5	
100	.53	.55	3.0	63.42
120	.46	.49	4.8	
140	.42	.44	3.88	
160	.39	.40	3.2	
180	.36	.37	2.76	
200	.33	.345	2.38	80.44
250	.27	.30	4.5	
300	.20	.23	2.65	
350	.14	.17	1.45	
400	.08	.11	.61	
450	.03	.05	.13	89.78

TABLE 17.4. *Evaluation of N_e for grain structures*

$N_e = 0.241\ N_{cr}$

N/N_{cr}	$r_{\tilde{\psi}}$	$\bar{a}$	$\bar{a}^2\Delta N$	$\Sigma(\bar{a})^2\Delta N$
0.05	0.97	0.985	0.049	
.10	.91	.95	.045	0.094
.15	.835	.88	.0385	
.20	.740	.79	.031	.1635
.25	.67	.685	.0235	
.30	.53	.585	.0171	.2041
.35	.44	.50	.0125	
.40	.37	.41	.0084	.225
.45	.30	.335	.0055	
.50	.245	.275	.0038	.2343
.55	.20	.22	.0024	
.60	.16	.18	.0016	.2383
.65	.125	.14	.0010	
.70	.10	.11	.0006	.240
.75	.075	.085	.00035	
.80	.058	.065	.0002	.2405
.85	.04	.045	.0001	
.90	.03	.04	.0001	.2407
.95	.02	.03		
1.0	.018	.02		.241

TABLE 17.5. *Diameter δ and equivalent passband N_e of various aperture types*

Aperture type	Relative transmittance	Diameter (δ)	Relation of δ to N_e
Square	$\tau=1$	s	$s=\ l/N_e$
Round	$\tau=1$	$2r_0$	$\delta_0=1.08\ l/N_e$
Do	$\tau=\cos^2 r$	$2r_0$	$\delta_{\cos}=1.59\ l/N_e$
Do	$\tau=\epsilon-r/r_0$	$6r_0$	$\delta_\epsilon=1.245\ l/N_e$
Do	$\tau=\epsilon-(r/r_0)^2$	$4r_0$	$\delta_{cx}=1.6\ l/N_e$

TABLE 17.6. *Equivalent passband (N_e) and diameter (δ_0) of equivalent round sampling aperture of imaging components*

		N_e	δ_0 (microns)	
Theoretical lens	$f/4$	250 L/mm	4.32	See figure 17.7.
	$f/6.3$	159	6.8	
	$f/8$	125	8.65	
	$f/16$	62.5	17.3	
40 mm Ciné Ektar	at $f/1.6$ (5°)	90 L/mm	12	See figure 17.10 and table 17.2.
	at $f/2.8$ (5°)	180	6	
50 mm Baltar at	$f/2.8$ (5°)	64	17	
4-inch Super Cinephor	$f/2$	27.3	39.5	
Film	N_{cr}*	0.241 N_{cr}	See figure 17.9	*N_{cr}="rated" resolution at $r_{\tilde{\psi}} \cong 2\%$.
Plus X	110 L/mm	26.5 L/mm	40.8	
Fine-grain negative (5203)	220	53	20.4	
Fine-grain positive (5302)	180	43.4	25	
16 mm reversal	150	36	30	
Square spot	$\tau=1$	0.50 N_c**	See figure 17.3	**N_c=limiting resolution at $r_{\tilde{\psi}}=2\%$.
Round spot	$\tau=1$	.45 N_c	See figure 17.4	
Round spot	$\tau=\cos^2 r$	.38 N_c	See figure 17.5	
Exponential spot	$\tau=\epsilon^{-r/r_0}$	.244 N_{cr}*		
Exponential spot	$\tau=\epsilon^{-(r/r_0)^2}$	.222 N_c	See figure 17.6	
Theoretical lens		.20 N_c	See figure 17.7	

$N_e \simeq 0.63/\delta_{0.5}$ is an approximation computed from the flux distribution occurring in practical star images. The factor 0.63 is a compromise value depending on the percentage of flux in the haze surrounding the nucleus of the star image.

The line number for known round apertures is expressed in relative units N/N_δ that refer to the aperture diameter $\delta = l/N_\delta$, where l is the unit of length ($l = 1$ mm, or $l = V$ = vertical picture dimension). Relative to the equivalent passband N_e, the diameter of these apertures is expressed by the relations given in table 17.5.

An equivalent aperture or point image of specified characteristics can thus be obtained for a system element by the insertion of its N_e-value into the relations given in table 17.5.

The equivalent passband N_e (television lines) and the equivalent aperture sizes of a number of system elements used in photographic processes are summarized in table 17.6.

Equivalent Passband and Aperture Diameter of Processes Containing a Number of Elements in Cascade

The sine-wave response characteristic of a number of system elements in cascade, including the eye if desired, can be computed accurately by forming the products of the response factors $r_{\tilde{\psi}_1}$ $r_{\tilde{\psi}_2}$, . . . $r_{\tilde{\psi}_n}$ of actual response characteristics at corresponding line numbers. The equivalent passband $N_{e(p)}$ of the process is thus given accurately by the integral

$$N_{e(p)} = \int_0^\infty (r_{\tilde{\psi}_1} r_{\tilde{\psi}_2} \ . \ . \ . \ r_{\tilde{\psi}_n})^2 dN \tag{14}$$

Because of the nature of the response characteristics of lenses, films, and television tubes it has been found that the equivalent sampling area of a combination of such "apertures" can be evaluated with usually less than 5-percent error by simply adding the equivalent aperture areas of the components or, as expressed in terms of equivalent aperture diameters:

$$\delta_{(p)} \simeq (\delta_1^2 + \delta_2^2 + \ . \ . \ . \ + \delta_n^2)^{\frac{1}{2}} \tag{15a}$$

also

$$1/N_{e(p)} \simeq (1/N_{e_1}^2 + 1/N_2^2 + \ . \ . \ . \ + 1/N_{e_n}^2)^{\frac{1}{2}} \tag{15b}$$

Thus, it becomes a simple matter to compute the equivalent passband $N_{e(p)}$ and the aperture diameter of photographic systems by the use of eq 15 in conjunction with table 17.6. Equation 15 is exact for exponential apertures $\tau = \epsilon^{-(r/r_0)^2}$ because the response characteristic figure 17.6 has the form $r\tilde{\psi}_{(N)} = \epsilon\text{-}KN^2$. The response characteristic of a system of two-dimensional apertures tends to approach this form (fig. 17.6), which may therefore be used as an equivalent response characteristic with a line-number scale $N_\delta = N_e/1.6$.

The simplified method will lead to larger errors and should not be used when electrical components of a television system such as amplifiers or filters with sharp cutoff or a rising frequency characteristic

518-743 O - 73 - 30

are included. Although equivalent pass-bands (N_e) for such components have a significance, they cannot be treated as normal optical apertures.

Tests for Visual Equivalence

Now that an objective measure for the equivalence of imaging devices or systems has been established, it remains to be shown that the equivalence holds under visual examination. It is obvious that an equivalence indicated by equal measures N_e will be satisfactory when obtained from sine-wave response characteristics that are very similar in shape. A repetition of aperture processes always tends to approach an ϵ^{-kN^2} shape. This observation can be proven when apertures of widely different form such as the combination of lenses and film shown in figure 17.16 are cascaded. In combination with the eye (fig. 17.17), the curve shape becomes even more normalized, indicating that the measure N_e agrees with visual observations. A most critical test for visual equivalence is a comparison of single imaging processes with widely different characteristics but equal measures N_e. This test can be made by comparing pictures made with a complex aperture (lens) such as shown in figure 17.10 with pictures made by an equivalent round aperture of constant transmittance or with a round aperture with $\cos^2$ transmittance. The response characteristic of the latter is shown by broken lines in figure 17.10, differing materially in resolving power from the actual lens. From this point of view the equivalence may appear rather inade-

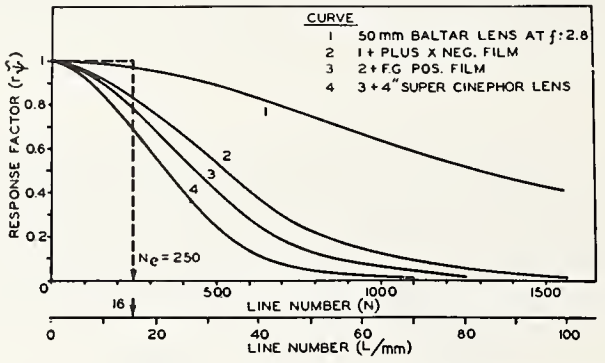


FIGURE 17.16. *Response characteristics in a 35-millimeter motion-picture system.*

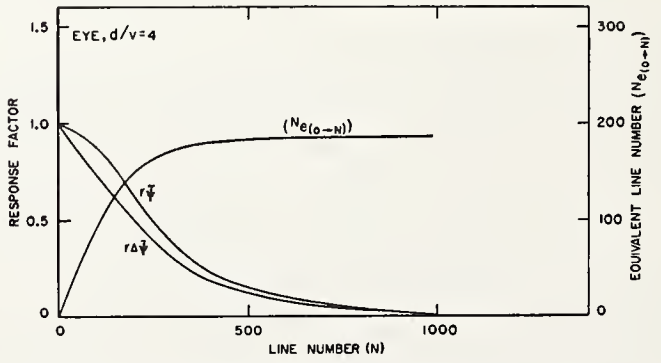


FIGURE 17.17. *Response characteristics of the eye at a viewing distance of four times the vertical picture dimension.*

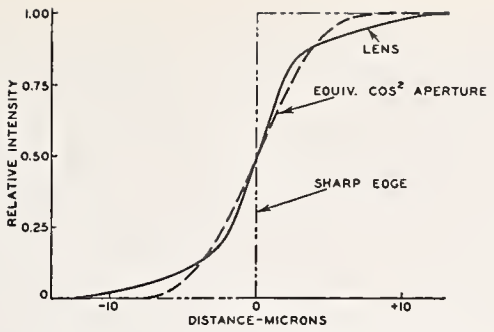


FIGURE 17.18. *Edge transitions of equivalent apertures shown in figure 17.10.*

quate; but when pictures are scanned or produced with these equivalent spots, the transitions at sharp edges do not differ nearly as much as shown by figure 17.18. To permit a visual comparison of images, different aperture shapes and intensity distributions have been synthesized accurately by out-of-focus projections with a precision enlarger. When a pinhole is imaged out of focus it assumes the shape of the enlarger lens diaphragm. In this position, variation of the lens-stop diameter causes a proportional variation in diameter of the out-of-focus point image. The f numbers corresponding to the lens aperture components in figure 17.11 are indicated in the lower portion of the insert as well as the relative exposures (1/speed) for obtaining respective flux values.

Photographs were produced in this manner by multiple exposures, the flux values being measured by a photoelectric device. To be able to show any difference more pronouncedly, the pictures were made with aperture sizes three times larger than that of the lens (respective to the picture frame for which the lens is designed) and thus represent a three times magnification of the image. The N_e-value in the originals was N_e=210, whereas that of the lens is N_e=630 in a frame 7 mm high. At a normal viewing distance these magnifications are found to have substantially equal sharpness, the equivalent round aperture with constant transmittance appearing slightly sharper. Upon close study the actual lens image exhibits higher resolution but slightly softer edges. Similar observations and tests made by comparing television images and motion pictures of equal N_e ratings have shown that a visual evaluation of sharpness is in good agreement with the objective equivalent N_e.

Conclusions

A study of the factors determining the sharpness in images produced by optical, photographic, and electrical image systems has shown that the relative performance of an imaging device with respect to detail rendition and edge sharpness is accurately specified by its sine-wave response characteristic. Given the characteristics of the components, the performance of any combination forming a multi-element imaging system can be computed accurately. It has thus been possible to solve many problems arising from a combination of optical, photographic, and television processes. Measurement of a

considerable number of lenses and television components has shown clearly that the resolving power has little significance as a measure of quality being perhaps the least important point on a response characteristic. The objective measure N_e specifying the equivalent optical passband of the imaging device or system is ideally suited for a universal system of rating image quality, its use permitting a great simplification of many problems. The sine-wave response factors and the measure N_e state performance in relative units. Absolute values of contrast require one additional specification—the large area contrast in the image plane, which can be determined by well known methods. A specification of lens quality must, therefore, include data on its large area contrast as well as a family of response characteristics (or N_e-numbers) as a function of f:number and angle. In view of the inadequacy of the present method of rating image quality by the resolving power of the device, and because of the lack of precise or useful information on detail rendition by lenses intended for pictorial purposes, the author recommends that the above system of rating be studied for adoption as an objective standard. A lens bench for photoelectric measurement of the response characteristic of lenses (including microscopes) has been developed by RCA for the Office of Naval Research under a study contract requiring an accurate evaluation of image quality. It employs moving sine-wave or square-wave test patterns photographed on the sound track of motion-picture film and analyzes the image formed by the lens with a slit aperture, which, referred to the image plane may be made as small as 0.01 micron. Direct readings of the measure N_e can be obtained with a calibrated "noise" film permitting, for example, an experimental and rapid determination of optimum spacings of lens elements. The proposed system of rating image-forming devices has been applied successfully to solve many problems in television systems that require a combination of optical, electrical, and photographic elements.

The author acknowledges the helpful criticism and contribution of W. A. Harris of the RCA Tube Dept., Harrison, N. J., and Dr. D. O. North of the RCA Laboratories, Princeton, N. J., in the analytical evaluation of equivalent-aperture passbands (N_e).

O. H. Schade, Electro-optical characteristics of television systems, RCA Rev. **9** (1948).

O. H. Schade, Image gradation, graininess and sharpness in television motion-picture systems, J. Soc. Mot. Pict. Televis. Engrs. **56**, No. 2 (Feb. 1951) and **58** No. 3 (March 1952).

O. H. Schade, A new system of measuring and specifying image definition, (unpublished paper given at the 69th Semiannual Convention of the Society of Motion-Picture and Television Engineers in New York City on May 3, 1951).

Discussion

Dr. R. C. Spencer,[1] Air Force Cambridge Research Laboratories, Cambridge, Mass.: The repeated accounts over the past 2 days of the subtle differences between resolving power on close-packed parallel lines and overall picture quality, coupled with Otto Shade's excellent treatment of the subject using communication theory, prompt the following remarks.

[1] Presented in written form following the Symposium as a summary of Dr. Spencer's oral discussion.

Most optical tests involve the high-frequency cut-off of the optical instrument, i. e., its ability to resolve two points (such as stars), two parallel lines, or to follow an abrupt change in illumination (step function). · Let us consider now a more general target or input function $G_0(x,y)$ containing lower frequencies that can be represented in the neighborhood of any point by a Taylor's series. Assume that if G_0 were a point (impulse) the optical instrument would spread it out into a system function $F(x,y)$. In general the output function $G_1(x,y)$, when $G_0(x,y)$ is the input function, is the convolution (Faltung) of G_0 and F, expressed as $G_1=G_0\times F$. I have shown[2] that for a one-dimensional case

$$G_1(x)=\left(\sum \frac{i^n/y_n}{n!}\,D^n\right)G_0(x)=PG_0$$

where P is an operator. Thus, if the nth moments of $F(x)$, defined by $\mu_n=\int x^nF(x)dx$ exist, the output function is a series of products of moments of $F(x)$ and derivatives of $G_0(x)$. Thus, for an even function with μ_0 normalized to unity

$$G_1(x)=G_0(x)-\tfrac{1}{2}\mu_2D^2G_0(x)+\ .\ .\ .\ .$$

This form is ideal for the correction or estimation of error of, say 10 percent, introduced into an input function by the limited resolving power of the instrument. Note that for positive moments the output curve G_1 lies always on the concave side of the $G_0(x)$ curve, thus cutting the corners. The case of the second moment for estimating distortion due to television scanning spots was substantiated by Mertz and Gran[3] in 1934. I later generalized the term "instrument" to include circuits, recording galvanometers, etc., in which cases the μ_2 term can be made zero or negative by proper underdamping.

It is noteworthy that the series of corrections is a differential operator P; that a series of instruments can be used in tandem and that the overall operator P is the product $P_1P_2P_3\ .\ .\ .$; that if each $F(x)$ is shifted to its own center of gravity, thus making each $\mu_1=0$, and all μ_0's are normalized to 1, the second moments μ_2 are additive; also the μ_3's.

These results are in agreement with the diffusion processes, and laws of combination of random errors; in particular, the *root-mean-square width* of the apparatus function is the *square root* of the *sum* of the *second moments* of the individual system functions.

The simple theory breaks down in the case of the second moment of the slit diffraction pattern $[(\sin\phi)/\phi]^2$, which is infinite. This was solved by Norbert Wiener who had independently realized the importance of the second moment and by 1941 had derived the best filter for an aperture, such that the second moment of its intensity diffraction pattern, and hence its rms width, would be a minimum. According to Wiener the amplitude transmission function over a slit aperture should be cos x with the edges of the aperture at the first zeros; also for a circular aperture the amplitude transmission

---

[2] R. C. Spencer, Phys. Rev. **38**, 618–629 (1931); [A]**46**, 337 (1934); [A] **48**, 4 3 (1935); **52**, 761 (1937), of which eq 4 should read $\mu_n=K^n\int_{-\infty}^{\infty}\psi^{2n}f(\psi)d\psi$; [A] **55**, 239 (1939); [A] **60**, 172 (1941). J. Appl. Phys. **20**, 413–414 (1949).

[3] P. Mertz and F. Gray, Bell Systems Tech. J. **8**, 464–515 (1934).

should be $J_0(r)$ with the edge coincident with the first null circle.

In conclusion, let me say that with most of the effort going into the study of the limit of resolution using sharp targets, we should begin to pay more attention to faithfulness of reproduction. The discussion above presents the low-frequency extreme. It does reiterate Shade's contention that the moment of inertia is important, especially for faithfulness of reproduction.

A complete solution, if successful, would have to incorporate considerable portions of the theory of information and in particular Wiener's theory of filtering which not only encompasses the complete spectrum of frequencies but enables the engineer to specify the optimum filter for maximum faithfulness of response of a typical class of input functions in the presence of unwanted noise. During the present sessions we have seen repeated indications of variety of: (a) *input functions* such as star points, lines, sine waves, step functions, square or round areas and typical shapes for letters, trucks, ships, etc.; (b) *noise* from atmospheric haze and temperature variations, photographic haze and particle size and resolution of the human eye. There is still lack of agreement on what is considered to be maximum faithfulness of reproduction for any one set of conditions but there is agreement on the fact that defocusing and other aberrations change the filtering characteristics of the instrument in a measurable manner, these characteristics being resolution and contrast.

Dr. D. S. Grey, Polaroid Corp., Cambridge, Mass.: I would like to comment on the papers by Selwyn and Schade. It has been some 3 years now since Selwyn and Schade have answered for us two questions that have received considerable discussion at this symposium. One question is, what sort of resolution target should we use? The second question is, how, from the energy distribution of the image of a point source, can the lens designer determine what results his lens will achieve under any particular test system with particular targets?

Now, the answer that Selwyn and Schade have given to the first question is that it is entirely immaterial what type is used. It is a question of how you interpret your results in connection with the type of target you use. If you use a line target it does not matter what the contrast in your lens is. If you interpret your results properly you can get the factors that Selwyn and Schade have shown to be pertinent and apply their analysis, and get the test results you would receive with any other type of test target.

The other question that is of interest particularly to the lens designer, is simply that the lens designer may obtain a spectrum of his point image and then he can predict just what his lens will do under a particular specified resolution of test.

I would like to point out one limitation to certain general aspects of the Selwyn-Schade method of analysis, which is the dependence on the object being illuminated by incoherent radiation. If the radiation is coherent, many of the linear properties do not hold, and since I have been working mainly in microscopy since Selwyn and Schade's paper appeared, I am, therefore, excused for not having climbed on the bandwagon sooner.

Now, we can go even further than Selwyn and Schade have indicated directly this morning. I am sure they had it in mind, however. If we analyze the image by their method we can predict what the results will be under test by any particular method. In particular, we can

predict what we are going to get in the photographing of a real object, if we know what the object is.

We can go even further than that by using certain very general theorems that have been developed in communication theory. We need not know exactly what our object is before we specify how our lens is going to perform, that is, to permit us to distinguish certain detail and say, "Is it there?" or "Isn't it there?" We can use the communication theorems and not concern ourselves with precisely what objects we are trying to image, but with the certain general properties of certain general classes of objects that we would like to examine. All the theorems of communication theory can go right over into optics just by a trivial change from one-dimensional notation to two-dimensional notation. We just draw a wiggly line under certain variables indicating that they are vectors and that certain multiplications are dot products of certain vectors.

I have been quite surprised that there has been so much discussion about just what type of target we should use so long after Schade and Selwyn first published their work.

Chairman B. O'Brien, University of Rochester, Rochester, N. Y.: Thank you, Mr. Grey. Is there any further discussion?

Dr. G. Toraldo di Francia, Instituto Nazionale di Ottica, Florence, Italy: I should like to remark upon Dr. Selwyn's paper and Dr. Schade's paper.

Dr. Selwyn measured the increasing contrast that shows up in the curve representing the contrast versus the spacing. He attributed it to the defects of the eye. I think that one can very well explain the phenomenon if one keeps in mind the fact that the resolution by the eye is not a static thing but a kinetic thing. When the spacing between the lines you have to resolve is larger, you need a larger contrast because the eye needs a longer time to scan it.

In this connection, may I remind you of a very fine experiment which I think was known to Helmholtz; looking through a pinhole in a cardboard at the sky and moving the pinhole before one's eye, one may see the blood vessels of the eye very clearly. But as soon as the cardboard is held stationary the blood vessels can no longer be seen. This indicates that the veins in the eye cannot be seen because they are not moving with respect to the retina.

Dr. Schade mentioned that the resolving power depends not on the amplitude of the illumination but on the flux, and I agree with him. We have made many experiments at the Optical Institute in Florence that confirm that the resolving power does not depend on the amplitude but on the energy. We have called it the energetic theory of resolving power. Our experiments are in agreement.

Chairman O'Brien: Thank you, Dr. Toraldo. Now, at the risk of being informal, and with Dr. Gardner's permission, I would like to add a comment to the discussion. What I have to say is that one can resolve quite successfully complex patterns where the total time of illumination is a microsecond or so and no possible scanning of the eye can take place. In order for the visual mechanism to sense the detail, it is necessary that the scanning be in space or in time. It is not essential that it be a scanning in space. Either is adequate. This does not mean that Dr. Selwyn's results are to be disagreed with. As a matter of fact, the agreement of published data is even better than he said. This is, no doubt, modesty on his part. Even

in Cobb's figures you will find perfect agreement with the curve sketched by Dr. Selwyn. Moreover, the fact that the angular separation of lines as it broadens requires a higher, not a lower, contrast of threshold, is in keeping not only with the physiology but the anatomy of the structures.

I shall now call time on myself and ask for more discussion.

Dr. R. E. Stephens, Optical Instruments Section, National Bureau of Standards, Washington, D. C.: Dr. O'Brien, there is another thing to add to that discussion, I think. The eye seems to be responsive to gradients and also gradients of gradients. There is an experiment I remember in which one produces a split field, one side of which has uniform brightness, the other having a constant gradient where the brightness at the boundary is the same as that of the uniform field. On opposite sides of the boundary there is an infinitesimal difference in brightness nevertheless, a definite line is visible. For sinewaves with the constant contrast, not only the gradient but the gradient of the gradient is smaller for large spacing than for short spacing.

Chairman O'Brien: I know you will agree that this is perhaps a more elegant way of stating the same quantity. Data have been published as a function of the breadth of that. All of these tie neatly together. Is there any further discussion?

Mr. J. M. Naish, Royal Aircraft Establishment, South Farnborough, Hants, England: I would like to make a remark upon the question of standardization of resolving power by measurements. This may not be a very appropriate moment to mention it in the course of the symposium but I feel that sooner or later some remark should be made upon this question. We believe that progress will be accelerated by standardization of resolving-power measurements.

From the symposium so far the importance of resolving power as a means of specifying the performance of the lens in terms of picture quality must be amply illustrated. In spite of the many defects in resolving power as a means of measurement, I think it seems that there remains a thing that will be measured and Dr. Washer's description of the measurement he makes, makes it very clear that it is not a simple matter to carry out a large number of measurements on a given day. The importance of other criteria for picture quality cannot be called in question and, indeed, we have done a certain amount of work on the relation between knife-edge test-object gradients and interferometer quality values with the corresponding resolving-power measurements in the same focal plane. I would think that is an important correlation, especially if the results may be weighted for the variation of wavelength, because of the importance of particularly transverse chromatic aberrations, but nevertheless we feel that we shall all be called back to the question of making these resolving-power measurements in the standard fashion.

I would like to invite attention to the importance of standardizing this procedure. Difficult though it is, it remains very important. We have a British standard. I am not sure whether there is an American standard for measuring resolving power, but I think that eventually we must arrive at an international standard.

The difficulties envolved in accomplishing this are enormous but may I briefly invite attention to the necessity for at least keeping

this goal in mind in the future?

CHAIRMAN O'BRIEN: Thank you. That is a noble goal. I hope some day we do achieve it. It certainly is worth keeping in mind.

Is there any further discussion of these papers?

PROF. F. ZERNIKE, Natuurkundig Laboratorium, Groningen, Netherlands: One question to Mr. Feder. In considering the calculation of intensity in the aberration image, he has said that he is troubled somewhat by the infinities that may occur there and he tries to get away from them by looking at the energy on a certain small surface.

I would like to know how he determines the size of that surface because everything will depend on that. Of course, in reality the finite wavelength of light takes care of this infinity and it will never be infinity at all.

CHAIRMAN O'BRIEN: Is Mr. Feder here?

Mr. D. P. FEDER, National Bureau of Standards, Washington, D. C.: I think my point was that with respect to geometrical optics alone, in which the question of energy density has no meaning. Do you agree with this?

PROF. ZERNIKE: Yes.

MR. FEDER: And that one might try something of this nature if the image is poor with respect to the Airy disk? One could, for instance, divide the entrance pupil of the lens into a series of apertures that cover the entire entrance pupil—say you divide it up in squares—and consider where the rays from the corners of the squares go. Assume that the energy then goes into the spot formed by these rays.

PROF. ZERNIKE: How many squares do you need?

MR. FEDER: I don't know the answer to that, but I would like to hear from Dr. Herzberger. Perhaps he will have an answer.

DR. M. HERZBERGER, Eastman Kodak Co., Rochester, N. Y.: Thank you very much. I would like to give an answer to that. In the case of very well corrected optical systems, microscopes or telescopes, it is , of course, necessary to calculate the diffraction image. This is the only way to get the light intensity; but you can get a kind of geometric optical pattern for a photographic lens in the following way. Make a grating in the entrance pupil by dividing the entrance pupil into equal parts. One can do that by a large number of equilateral triangles. Then each ray represents an equal amount of light. If one has a large number of rays and intersects them with the image plane, one gets a distribution that is equivalent, or practically equivalent, to the light distribution in the image. In that manner one can get, by calculating the mumber of points within a given area, the amount of light falling within that area.

PROF. ZERNIKE: Well, my impression is that in order to make, as I would say, an improvised method of taking into account the wavelengths of light—

DR. HERZBERGER: No, the wavelengths of light do not come into this pattern because it would not be very different for different wavelengths—at least as long as the aberrations of the system are large with respect to the wavelengths of light.

PROF. ZERNIKE: I do not agree because Mr. Feder said expressly that the intensity would be infinite.

DR. HERZBERGER: Yes. You see, if one looks at these pictures one gets an immediate impression of the light distribution because the points are so dense. In geometrical optics there would be an

infinite density at the caustic. W. R. Hamilton already has investigated this problem and has tried to compare the intensity at different parts of the caustic. One compares equal areas of the caustic and thus finds a comparative measurement of the intensity at the caustic. The spot diagrams, on the other hand, do not give a correct but an approximate picture of the intensity distribution. They do, however, seem to provide a more graphic method of showing the intensity differences in the image than drawing lines of equal intensity.

Chairman O'Brien: Thank you, Dr. Herzberger. This subject could easily be the topic of another symposium as extensive as this one, and Dr. Gardner has asked that we try to undertake the afternoon program very soon. However, before I spoke, at least Dr. Lucy had waved and perhaps others, so rather than cut off too abruptly, may we have your remarks?

Dr. F. A. Lucy, University of California, Los Alamos Scientific Laboratory, Los Alamos, N. Mex.: I have a brief comment. I don't think we have to worry too much about the product of an infinite density over a zero area. The exact limit of the caustic here being really zero in area, we have an indeterminate form that can have a finite product.

Dr. G. H. Conant, Jr., Harvard College Observatory, Cambridge, Mass.: May I say in defense of Mr. Feder, that the problem that he was discussing was one of computing with automatic computing machinery and your machine cannot evaluate these indeterminate things. The machine, itself, can only handle a certain finite range of magnitude.

Chairman O'Brien: Mr. Feder, would you like to cap your climax, so to speak?

Mr. Feder: No, I was thinking about Dr. Zernike's question. It seems to me that it is essentially a matter of how closely you examine the image. In other words, if you are going to look at it with a very high powered microscope, then I don't think this method will work. But if you are going to examine it with something that has farily low sensitivity, that is if you take a fairly large area of the image, and if the image is bad with respect to the Rayleigh limit, it is a question of selecting the triangles or squares or what have you in the entrance pupil, so that the energy from one triangle or one square gets into the portion of the area you examine in the image.

Chairman O'Brien: Dr. Hopkins?

Dr. R. E. Hopkins, University of Rochester, Rochester, N. Y.: It seems to me that it is easy to see what the problem is if you think of the interference patterns that I showed. When you have a round, circular area free from fringe, you have to tell from physical optics where the rays within that area are going.

Chairman O'Brien: It is refreshing to see you in such complete agreement.

Prof. A. Maréchal, Institut d'Optique, Paris, France: I have to mention that Prof. Durand in France obtained a better approximation in taking account of the different phases of the vibrations corresponding to the traced rays. It is simpler than complete calculation of diffraction patterns and nevertheless gives a good approximation.

In order to know the contrast of images, or the resolving power of an instrument, information is needed on the diffraction pattern. Those quantities determining the optical quality of the instrument

can be computed by various integrations over the diffraction image. Fortunately, Prof. Duffieux [4] and his co-workers have shown that those quantities can also be computed by convenient integrations performed on the exit pupil. The example of the contrast of images of lines has been studied in a previous communication.

Mr. W. A. Allen,[5] U. S. Naval Ordnance Test Station, Inyokern, China Lake, Calif.: In reference to the paper presented by Mr. Feder, it might be mentioned that many other laboratories have been using automatic computing machinery for the purpose of ray tracing. Specifically, the Los Alamos Scientific Laboratory [6] and the U. S. Naval Ordnance Test Station [7] have investigated and designed many lenses by automatic computing-machine methods. Recent work at the Naval Ordnance Test Station has been based on generalized equations used to trace rays through uncentered-spherical interfaces, prisms, and aspheric surfaces. All surfaces may be considered either refracting or reflecting. Our equations for an aspheric surface appear to converge more rapidly than those used by the National Bureau of Standards.

We have heard much discussion here about the desirability of considering the problem of lens design from the standpoint of physical optics. It should be mentioned that ray tracing in the notation used by the National Bureau of Standards and the Naval Ordnance Test Station provides a calculation of the optical pathlength from object to image. Machine performance is such that this distance is accurate to the wavelength of light. Thus, the data from geometrical ray tracing can be translated by straightforward methods into the language of physical optics. Aspheric surfaces have been designed by the methods of equalizing all optical pathlengths from object to image.[5]

In conclusion, I would like to suggest that there exists a great need for a program involving the systematic evaluation of existing lens prescriptions in terms of resolving power. Such a program would entail the adoption of some reasonable arbitrary standard, such as the criterion suggested by Dr. Hopkins, relating the results of ray tracing to resolving power. Some agency should undertake the analysis and publication of results pertaining to all lenses described in the patent literature. At the present time, any lens design can be investigated exhaustively and rapidly with respect to available glasses, contemplated magnification, and configuration of the image surface, by means of automatic computing machinery. Some resolving-power data, however, brief and inadequate, should be known in advance of such a thorough analysis in order to screen the large group of all known lenses down to a few that might meet prescribed minimum-performance standards. As Dr. Hopkins has remarked, many electronic calculators have been built at the public expense; ray-tracing data obtained by their use should become public property. In my opinion, publication of such information would eliminate much present duplication of effort.

---

[4] P. M. Duffieux, L'intégrale de Fourier et ses applications à l'optique, (Faculte des Sciences de Besançon, 1947).
[5] Received after the close of the symposium.
[6] W. A. Allen and R. H. Stark, J. Opt. Soc. Am. *41*, 636 (1951).
[7] W. A. Allen and J. R. Snyder, J. Opt. Soc. Am. (in press).

EDITOR'S NOTE: Dr. Allen's reference to the early use of automatic computing machinery for ray tracing at the Los Alamos Scientific Laboratory and at the U. S. Naval Ordnance Test Station is most welcome. At the same time this suggests the advisability of listing other early applications of a similar nature.

In 1944, Dr. James G. Baker used the IBM Automatic Sequence Controlled Calculator (Mark I) at Harvard for tracing a series of skew rays through an optical system. The equations used and an example of recorded data will be found in one of the National Defense Research publications [1]. This is believed to be the earliest use of automatic programmed computing machines for ray tracing.

In 1945 Grosch [2] reported on the tracing of meridional and skew rays by means of International Business Machines Corporation equipment. The equipment available at that time did not have large storage capacity or provision for sequence programing. Consequently the tracing of a ray through a single surface required the application and removal of 28 plug-boards. Such a procedure really constituted a *tour de force* and could be only used when a large number of rays, say 100 or more, were to be traced through a given optical system. With such a requirement it is possible and desirable to perform a given operation on all the rays before the plug-board is changed for the next operation. In this method the rays are traced *in parallel* instead of in series.

At the fall meeting of the Optical Society of America, October 1949, Grosch [3] reported on the tracing of skew rays by the Selective Sequence Electronic Calculator (SSEC) of the International Business Machines Corporation. This is a programmed machine with sufficient storage capacity to permit ray tracing *in series* through a series of surfaces. In other words, a given ray is traced through the entire optical system before a computation of a second ray is begun. On March 9, 1950, at the time of the meeting of the Optical Society of America in New York there was a demonstration of ray tracing by this machine for the members of the society. At this same meeting Epstein [4] described a method for calculating the third-order aberrations of an optical system by means of some of the less complex IBM equipment.

In March 1950 at a meeting of the Association for Computing Machinery, Donald P. Feder and Benjamin F. Handy, Jr., of the National Bureau of Standards presented a paper titled "Optical Ray Tracing Problems and The Card Programmed Calculator" (CPC). This paper described the use of the CPC in tracing skew rays through a system of spherical surfaces. Also in March of 1950 the Standards Eastern Automatic Computer (SEAC) was first used to trace rays. The use of both of these machines was further discussed by Feder [5] in a talk given a year later at the March 1951 meeting of the Optical Society. At this same meeting Wooters [6] described a method of ray tracing by means of the IBM 604 unit.

In September 1951 Feder [7] gave the detailed formulas that had been used with automatic computing machinery for tracing skew rays through any rotationally symmetric optical system and also formulas for calculating third-order image errors for such systems. At the October 1951 meeting of the Optical Society papers on the use of automatic digital computers for use in geometrical optics were given by Cox and Ledda [8]; Jacobs, May, and Scholnick [9]; Wooters [10]; and Woodson [11].

The foregoing is believed to be a complete summary of the papers appearing in the Journal of the Optical Society of America as presented at its meetings during the past 7 or 8 years. Other applications of programmed computing machinery to optical computations have perhaps been made by other firms and institutions in this country but the information is not available for a complete listing. It may be mentioned, however, that at the National Bureau of Standards, fifth-order aberration equations derived from the equations of Wachendorf [12] are being used for computations on the SEAC machine.

---

[1] Design and development of an automatically focusing 40-inch f/5.0 distortionless telephoto and related lenses for high-altitude aerial reconnaissance, NDRC, Section 16.1, Optical Instruments.

[2] H. R. J. Grosch, Ray tracing with punched-card equipment, Abstract 27, J. Opt. Soc. Am. **35,** 803 (1945).

[3] H. R. J. Grosch, Ray tracing with the selective sequence electronic calculator, Abstract 42, J. Opt. Soc. Am. **39,** 1059 (1949).

[4] L. Ivan Epstein, Calculation of third-order aberrations with the aid of IBM machine, Abstract 9, J. Opt. Soc. Am. **40,** 255 (1950).

[5] Donald P. Feder, Ray tracing with automatic computing machinery, Abstract 45, J. Opt. Soc. Am. **41,** 289 (1951).
[6] Glenn Wooters, Ray tracing with the IBM Electronic Calculating Punch Type 604, Abstract 44, J. Opt. Soc. Am. **41,** 289 (1951).
[7] Donald P. Feder, Optical calculations with automatic computing Machinery, J. Opt. Soc. Am. **41,** 630 (1951).
[8] Arthur Cox and Catherine E. Ledda, IBM automatic equipment in optical design, Abstract 48, J. Opt. Soc. Am. **41,** 874 (1951).
[9] Donald H. Jacobs, Michael May, and Seymour Scholnick, A compact ultra-high Speed digital ray-tracing machine, Abstract 49, J. Opt. Soc. Am. **41,** 874 (1951).
[10] Glenn Wooters, Computing effects of lens variations with electronic calculator, Abstract 50, J. Opt. Soc. Am. **41,** 874 (1951).
[11] Robert A. Woodson, An analysis of H. D. Taylor's f:2 photo lens using a card programmed electronic calculator, Abstract 51, J. Opt. Soc. Am. **41,** 874 (1951).
[12] F. Wachendorf, Bestimmung der Bildfehler 5. Ordnung in zentrierten optischen Systemen, Optik. **5,** 80 (1949).

18. Position of Best Focus of a Lens in the Presence of Spherical Aberration

By R. Kingslake [1]

Introduction

The relation between the image formed by a lens and the amount and kind of spherical aberration present, is of perennial interest to optical designers. Since some residual spherical aberration is usually unavoidable, the designer is concerned with knowing how best to subdivide this residual between marginal overcorrection and zonal undercorrection. The properties that must be considered in this connection are (a) resolving power; (b) depth of focus; (c) the position of the best image, and hence the "Focus shift" (i. e., a longitudinal displacement of the plane of best definition as the lens is stopped down to a smaller aperture); (d) contrast (aberrational haze causes a loss of contrast in the image; this is, of course, entirely distinct from the loss of contrast due to flare light from the lens).

No attempt was made in the present work to measure image contrast, although that would be necessary for a complete study of the problem.

Historical Outline

For very small amounts of primary spherical aberration, in which the sum of the maximum positive and negative departures of the light wave from a perfect sphere does not exceed about a quarter of a wavelength (the Rayleigh limit), it is well known [1][2] that the best-focus plane falls midway between the marginal and paraxial images, the light distribution in the elementary star image being then very similar to that of the ideal Airy disk. From integrations made by Lommel and others, Picht [2] was able to plot a longitudinal section of the light distribution in that simple case.

In 1925, H. G. Conrady [3, 4] studied the properties of a lens having up to 20 times the Rayleigh limit of primary spherical aberration, and found that neither the position of the best focus, the resolving power, nor the depth of focus then agreed with simple theoretical predictions. Her graphs indicate that as the amount of aberration is gradually increased from zero, the plane of best resolution first falls midway between the paraxial and marginal foci, up to the Rayleigh limit, but after that it remains approximately stationary relative to the paraxial focus while the marginal focus moves progressively further away. Thus, neither the simple optical-path-difference theory nor the circle-of-least-confusion theory derived from ray paths represents the facts when large amounts of primary aberration are present.

The more general case of a lens having a zonal residual of undercorrection with some marginal overcorrection has been attacked by

[1] Eastman Kodak Co., Rochester, N. Y.

[2] Figures in brackets indicate the literature references on p. 266.

various workers, an outstanding paper being that by Flügge [5], but without any general results being established. In view of the importance of the problem, it seemed worth while to make some direct determinations of the properties of a lens having a variable amount of zonal aberration, in the hope of finding out some definite laws.

Experimental Procedure

A convenient source of a variable amount of spherical aberration was found to be an f/4.5 photographic lens of the Tessar type, of 4 inches focal length, focused by moving the front element. At each position of the front element, the spherical aberration curve was plotted by a modified form of the well-known Hartmann Test, taking five zones of the lens in succession, using "Super-X" panchromatic 16-mm film, D–76 developer, and the band of green light transmitted by a No. 57 filter. The lens was mounted on a nodal-slide lens testing bench, with a 25-mm photographic lens used as a microscope objective to magnify the aerial image 14 times before it was photographed. The film was housed in a Ciné-Special 16-mm camera, without lens, this camera being particularly suitable for the purpose as it is equipped with a direct-view reflex finder and a single-frame crank. The longitudinal position of the viewing microscope was indicated by a dial gauge on the bench, graduated in thousandths of an inch (.025 mm). The test object was either a point source or a resolution chart about 70 feet distant, the chart lines being spaced at intervals of $\sqrt[3]{2}$ from 24 to 300 lines per mm in the aerial image.

For each lens position, a strip of film was obtained, showing on successive 16-mm frames: five pairs of dots from the five lens zones at the inner Hartmann plane just within the best focus; then five pairs of dots from the same zones of the lens at the outer Hartmann plane 0.080 inch (2 mm) beyond the inner Hartmann plane; and finally a succession of photographs of the resolution test object at one-thousandth-inch steps, passing through the best focus region. The exposure time for the Hartmann dot patterns was about 40 seconds, and for the resolution charts an 8-second exposure was satisfactory. In some cases, the lens was subsequently stopped down to a known reduced aperture by an iris diaphragm mounted behind the lens in the same plane as the Hartmann diaphragm, the exposure time being increased to compensate for the reduced lens aperture.

The spacing of the successive pairs of dots in the Hartmann images was measured by projection on a distant screen, the exact projector magnification being immaterial since the position of each zonal focus between the two Hartmann planes was found by simple proportion. Four of the zonal foci were used to determine the four coefficients a, b, c, and d in a power series of the type

$$F=a+bY^2+cY^4+dY^6, \tag{1}$$

the coefficient a representing, of course, the position of the paraxial image-point. The fifth zonal focus was then used as a check, the observed and calculated positions never differing by more than 0.001 inch (.025 mm).

The series of resolution-chart images on the film were first examined macroscopically with the unaided eye, to locate the cleanest image where the contrast is a maximum. They were then further studied

under a 10X loupe to read the actual resolution achieved at each focus position. The resolution figures were plotted on the same graph as the measured spherical aberration curve for easy comparison.

Experimental Results

The effects of unscrewing the front element of the lens used for these experiments, are shown in figures 18.1 to 18.6. The front element of the lens was mounted on a triple-40 thread, hence it moved 1.905 mm/turn.

In these diagrams, each aberration curve is plotted against semi-aperture, and above it is added a curve of the resolving-power data read directly from the film. The shaded circle on each chart represents the position of the image of greatest contrast and clearness, determined from the film by direct view without a magnifier. It is probable that this would represent the preferred image position for

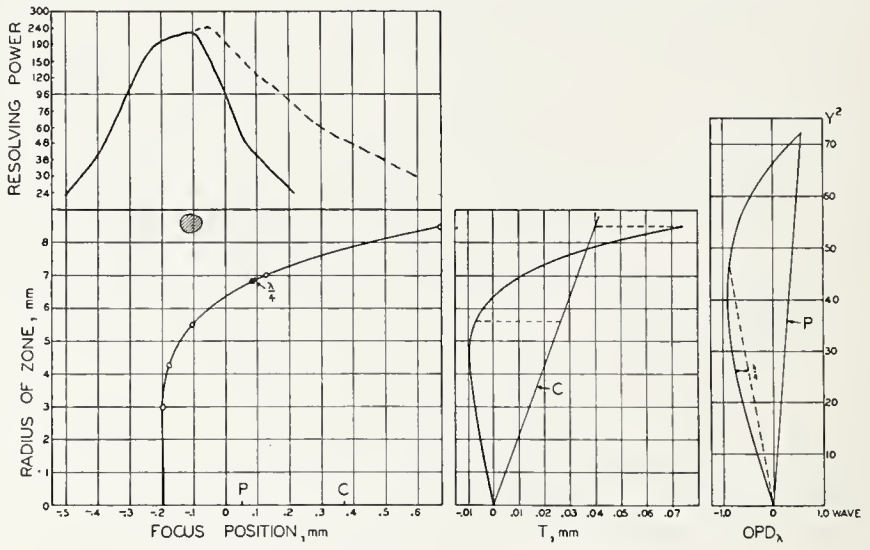


FIGURE 18.1 *Front element screwed in fully.*

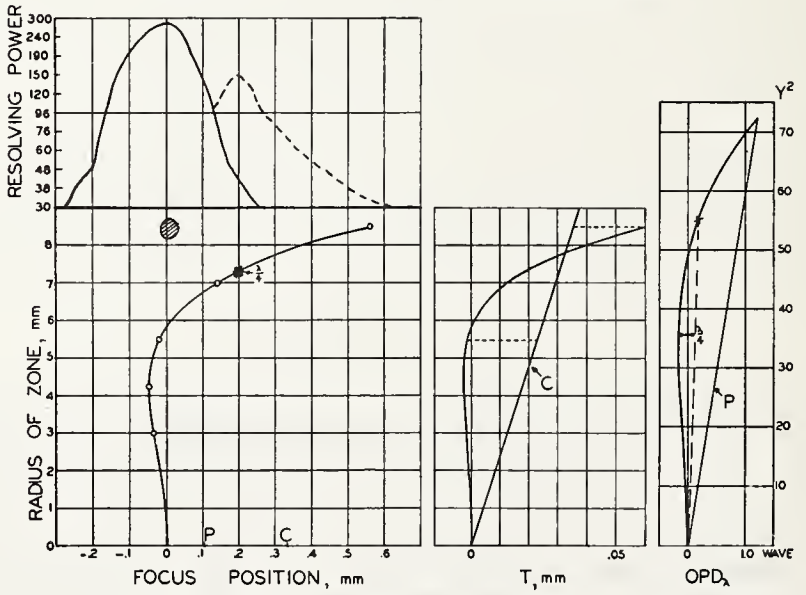


FIGURE 18.2. *Front element unscrewed 1/8 turn (.24 mm).*

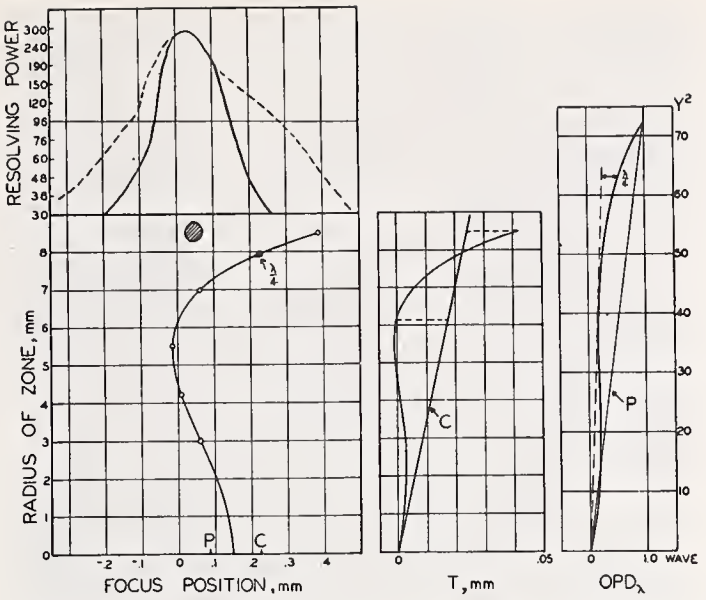


FIGURE 18.3. *Front element unscrewed 1/4 turn (.48 mm).*

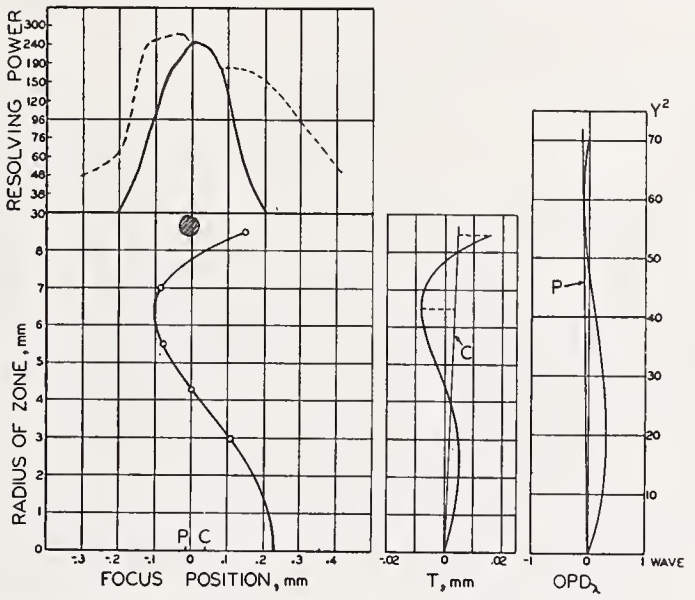


FIGURE 18.4. *Front element unscrewed 3/8 turn (.71 mm).*

ordinary photography, although the focal position corresponding to the peak of the resolution curve might be preferred if fine detail of high contrast is being studied. An excellent illustration of this phenomenon is given in [6].

The resolving-power graphs consist of two branches. The heavy curve represents genuine resolution in which the separate lines in the target are clearly imaged with sharp edges. However, in many cases after genuine resolution had ceased or almost ceased, the *finer* patterns would become clear again down to a very small pattern, after which resolution would vanish a second time (fig. 18.7). The limits of this extra-fine pseudo resolution are marked on the various graphs by dotted lines. It must be regarded as a form of spurious resolution,

518-743 O - 73 - 31

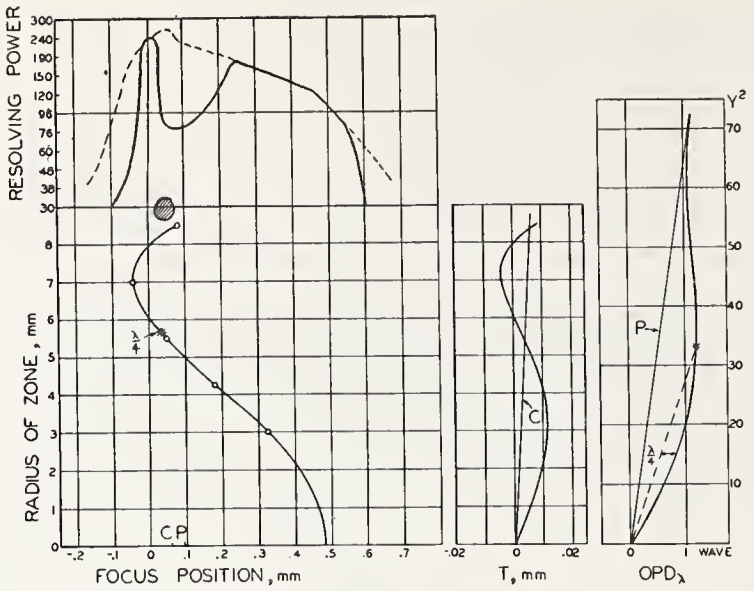


FIGURE 18.5. *Front element unscrewed 1/2 turn (.95 mm).*

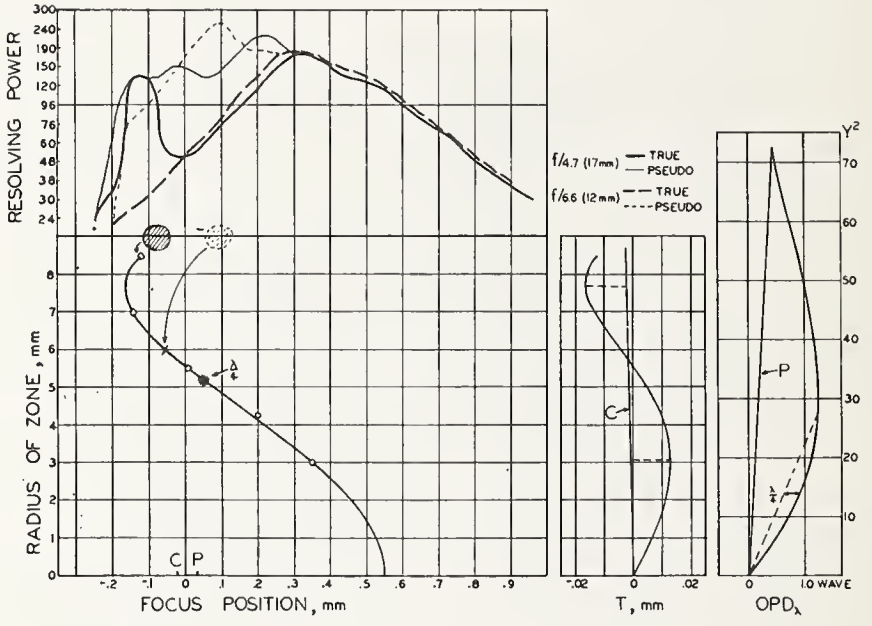


FIGURE 18.6. *Front element unscrewed 5/8 turn (1.19 mm).*

even though each pattern shows the correct number of lines (in this case four) and there is no reversal of black and white. This case has been recently recognized by workers at the National Bureau of Standards [7].

The curves for extreme undercorrection, figures 18.5 and 18.6, show two distinct peaks of genuine resolution, one a broad peak representing good resolution but a rather hazy image, and the other an abrupt peak at the point of maximum contrast where the image is very clear and clean. The latter resolution peak is evidently caused by the bend in the aberration curve, for it vanishes when the lens is stopped down (fig. 18.6). It should be noted that the peak resolution in lines per millimeter is not greatly different in any of the charts, showing

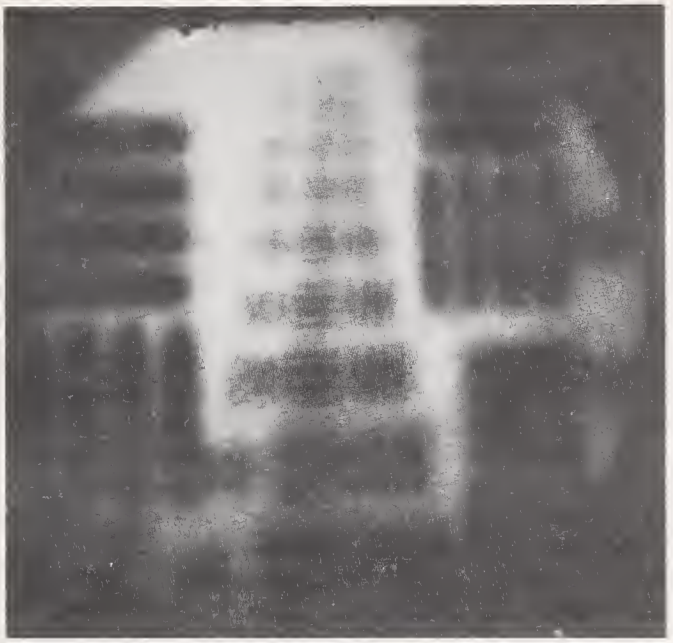


FIGURE 18.7. *An enlargement from a typical 16-mm frame showing pseudo resolutions.*

that even 1 mm of spherical aberration is not serious in this respect. The depth of focus increases greatly in the presence of undercorrection, but it is very little effected by overcorrection.

The resolution figures obtained in this test were very high, being substantially the visual resolution of the lens for monochromatic light. It is probable that the pseudo resolution observed in these experiments would not have been found had the aerial image been photographed directly without magnification.

By far the worst practical objection to spherical undercorrection is the shift of focus with aperture. The small shaded circle on each chart represents the position of the "best" image for ordinary photography, and it is clear that if the lens were stopped down to a very small aperture, the plane of "best" focus would move to the paraxial image position. To illustrate this for the case shown in figure 18.6 the series of test-chart photographs was repeated after stopping the lenses down from 17- to 12-mm aperture, and the "best" focus was found to have moved to the position marked by the dotted shaded circle on that chart. However, the peak of the *resolution* curve (ignoring the psuedo resolution) was unchanged in both position and height by stopping the lens down in this way. This bears out the conclusion of Conrady [3, 4] for lenses having pure primary aberration, that the best-resolution plane remains approximately fixed relative to the paraxial focus, even though the marginal focus is moved. This phenomenon of focus shift did not appear when the lens was overcorrected.

Theoretical Attempts to Explain the Observed Phenomena

Two standard modes of approach have been used in the past in attempting to predict the position of the best focus when the form of a spherical aberration curve is known: (a) to determine the position of

the geometrical circle of least confusion, and (b) to locate the center of the sphere that best fits the emerging wave-front.

These procedures are well known, and they have been applied to each of the six cases considered in this investigation. The geometrical approach consists in plotting the curve of transverse aberration, T, against semiaperture, Y, computed by $T=F(Y/l')$. Here F is the—longitudinal focus position relative to an arbitrary zero, and l' is the distance from the Hartmann diaphragm to the image, in this case about 80 mm. The position of the geometrical circle of least confusion marked C on the graphs, is found by drawing such a sloping straight line through the origin that the maximum positive and negative departures of the T curve from the sloping line are equal to each other. Taking any point on this line, the coordinates of which are Y and T, the longitudinal position of the circle of least confusion on the F scale is found at $\delta F=T(l'/Y)$. It is seen from the charts that only in the presence of an undercorrected zone did the geometrical circle of least confusion agree closely with the observed position of best contrast.

The second approach, namely, to plot the shape of the emerging wavefront and try to fit a sphere to it as closely as possible [8, 9], is more elaborate as it requires some additonal calculation. The optical path difference, OPD, is equal to the integral of the angular spherical aberration with respect to the aperture-height Y. Hence,

$$OPD_Y=\frac{1}{l'}\int_0^Y T\,dY=\frac{1}{l'^2}\int_0^Y FY\cdot dY=\frac{1}{l'^2}\int_0^Y (aY+bY^3+cY^5+dY^7)dY.$$

$$\therefore OPD_Y=\frac{aY^2}{2l'^2}+\frac{bY^4}{4l'^2}+\frac{cY^6}{6l'^2}+\frac{dY^8}{8l'^2}. \qquad (2)$$

(The most convenient unit of OPD is the wavelength.) After substituting in this formula the values of a, b, c, and d found for each case, the OPD curve was plotted against Y^2 as ordinate, and a straight line added such that the sum of the greatest plus and minus departures of the curve from the line is a minimum. In most of the cases considered here, such a line merely joined the two ends of the curve. By reading the Y^2 and OPD values of a point on this line, and substituting them in the formula

$$\delta F=2l'^2\cdot OPD/Y^2,$$

the position of the center of curvature, P, of the best-fitting sphere could be found. The maximum residual OPD between the wavefront and the straight line was also read off the graph, giving a measure of the aberration in multiples of the Rayleigh limit ($\lambda/4$). By adding a second straight line to the wavefront graph, at a maximum distance $\lambda/4$ from it, it was possible to ascertain by how much the lens must be stopped down to reduce the aberration to the Rayleigh limit; this point is shown on the aberration curves by an asterisk.

The position of physical best focus, P, agreed with the focus for maximum contrast remarkably well, except in the case of extreme overcorrection.

The numerical results of this investigation may be summarizd in tabular form as follows. The "depth of focus", in column 7, represents the longitudinal distance between the planes of 30-line genuine resolution on the two sides of the peak. The lens was, in all cases, used at its full f/4.5 aperture.

Movement of front lens	Marginal spherical aberration	Spherical aberration in multiples of Rayleigh limit	Focus shift, i. e., distance from paraxial focus to:		Peak resolution	Depth of focus (30 to 30 line)
			Maximum-contrast image	Maximum-resolution image		
mm	*mm*		*mm*	*mm*	*Lines/mm*	*mm*
0	+0.85	5.0	+0.10	+0.10	230	0.62
0.24	+0.57	3.4	0	−0.02	280	0.51
0.48	+0.24	1.5	−0.10	−0.10	270	0.46
0.71	−0.08	1.2	−0.24	−0.20	240	0.40
0.95	−0.40	2.8	−0.45	−0.25 −0.45	180 240	0.70
1.19	−0.68	4.2	−0.63	−0.23 −0.67	180 130	1.17

Conclusions

The best distribution of undercorrection and overcorrection is probably that shown in either figure 18.2 or figure 18.3, in which the marginal overcorrection is several times as great as the zonal undercorrection. In these cases the peak resolution is very high, while the best-contrast image and the peak of resolution fall close to the paraxial focus. Such a lens will show negligible focus shift when it is stopped down. The actual position of the best focus can be calculated by finding the center of the sphere that best fits the calculated wave front, but in making this determination the extreme overcorrected margin of the lens must be ignored as the overcorrected aberration tends to move the calculated point, P, more than it moves the observed best focus. It must always be remembered that any significant zonal undercorrection will lead at once to a focus shift.

The experiments described in this paper refer, of course, to a lens of one particular focal length and aperture, and to one amount of spherical aberration. It may well happen that in a much larger or much smaller lens, different conclusions as to the best distribution of marginal overcorrection and zonal undercorrection would be reached. Also, if accurate measurements on image contrast had been included, the advantages of a large amount of overcorrection might well be offset by the excessive haze accompanying it.

References

[1] A. E. Conrady, Optics of the microscope, Glazebrook's Dictionary of App. Phys. **4,** 217–218 (Macmillan, London, 1923).

[2] J. Picht, Uber den Schwingungsvorgang, Ann. Physik [4] **77,** 217–218 (Macmillan, London, 1923).

[3] H. G. Conrady, An experimental study of the effects of varying amounts of primary spherical aberration on the location and quality of optical images, Phot. J. **66,** 9–25 (1926).

[4] H. G. Conrady, Effects of primary spherical aberration on optical images, Proc. Opt. Conv. p. 830–838 (1926).

[5] J. Flügge, Uber die Verundeutlichung des Bildes photographischer Systeme durch die Sphärische Aberration, Z. Instrumentenk. **46,** 333–354, 389–415 (1926).

[6] M. Koomen, R. Scolnik, and R. Tousey, Night Myopia, J. Opt. Soc. Am. **41,** 89 (1951).

[7] R. N. Hotchkiss, F. E. Washer, and F. W. Rosberry, spurious resolution of photographic lenses, J. Opt. Soc. Am. **41,** 600–60; (1951).

[8] R. Richter, Zur beugungstheoretischen Untersuchung optischer Systeme, Z. Instrumentenk. **45,** 1–15 (1925).

[9] Wang Ta-Hang, Note on the best focus in the presence of spherical aberration, Proc. Phys. Soc. **53,** 157–169 (1941).

19. Image Evaluation by Edge Gradients

By Arthur Cox [1]

The elementary fact to be faced in connection with any lens is that it yields a light patch of finite size in the focal plane, and that the quality of image it yields depends on the shape and size of this light patch. A practical problem of importance is to bridge the gap between knowing that the image quality is correlated to the form of light patch, and establishing the correlation in a simple and quantitative way.

No correlation can be attempted until an adequate definition or series of definitions of image quality, are established. Setting up a suitable and useful criterion of image quality is a major problem, and a number have been proposed. Two criteria of quality will be considered in this paper, namely the resolving power of a lens as measured with a target having equal lines and spaces, and the ability of a lens to detect a fine dark strip against a bright background. The way in which these may be correlated with the form of light patch, either measured or computed, will be examined. The correlation between other criteria of quality and the form of light patch may be established, in some cases, by an extension of the method proposed.

Suppose that the object field in front of a lens comprises two semi-infinite areas, one bright, the other perfectly dark, with an abrupt brightness transition across the boundary between them. If the light patch were infinitesimally small the intensity distribution in the focal plane would be represented by the line $ABB'C$ in figure 19.1. Because of the finite size of the light patch the actual intensity is represented by the continuous curve of figure 19.1.

Suppose next that the object field comprises a bright strip of finite width on a dark background. In the ideal case the boundaries of the image of this strip would lie along BB' and DD'. With a semi-infinite area terminating on BB' the intensity at a point P would be given by PX in figure 19.1. With a semifinite area terminating on DD' the intensity at P would be given by PX'. The difference between PX and PX' is the intensity at P due to the strip of finite width. The intensity curves in figure 19.1, both the continuous curve corresponding to BB' and the broken curve corresponding to DD', are identical and parallel to one another. Hence PX' is equal to QY where $DB=PQ$, and the intensity at P is the difference between PX and QY. This procedure for obtaining the intensity due to a bright strip may be extended to the case where the object field contains lines and spaces of equal width, the spaces being completely dark. Thus with the edge intensity distribution and the object field pattern shown in figure 19.2 the intensity at P is given by $PX-P_1X_1+P_2X_2 \ldots P_nX_n-P_{-1}X_{-1}+P_{-2}X_{-2}-P_{-3}X_{-3} \ldots +P_{-m}X_{-m}$. The summation of the terms with negative subscripts is over an even

[1] Farrand Optical Co., New York, N. Y., (now Chief Optical Designer, Bell & Howell Co., Chicago, Ill.

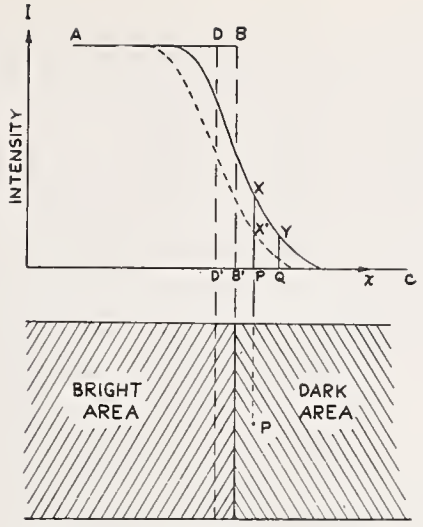


FIGURE 19.1. *Variation of light intensity across a single boundary.*

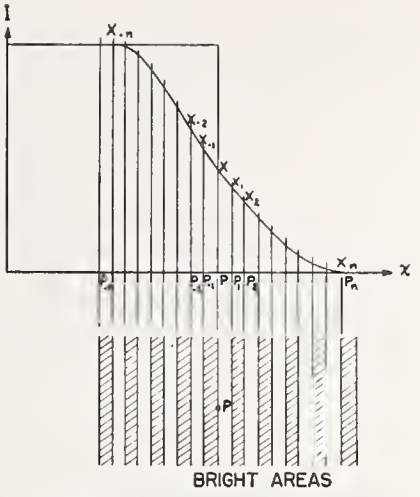


FIGURE 19.2. *Variation of light intensity due to a repeated pattern.*

number of terms ending on the flat part of the intensity curve.

In this way the intensity at any point, due to a repeated pattern of bright lines and dark spaces, may be readily determined. In particular, the intensity at the center of a dark space may be compared with the intensity at the center of a bright line. With a coarse line pattern there is a considerable difference between these intensities. As the pattern becomes finer the difference becomes less, goes through zero and reverses its sign. By setting a suitable ratio of these intensities, usually less than unity, as a criterion of resolution we can readily determine the resolving power of the lens.

The above treatment is appropriate to the case where there is infinite contrast between lines and spaces. When the contrast is finite we can consider a pattern of infinite contrast superimposed on a background intensity, and so derive the resolving power for finite contrast.

Thus if we have as light patch a uniform disk of diameter D, and if X is the relative width of line or space, the intensities I_1 and I_2 at the centers of the lines and spaces are given by table 19.1. Assuming a criterion of resolution as $I_2 \geq .8I_1$ we have $x=.45$. For a finite contrast we can take a base intensity I_0 and $(I_0+I)/I_0=\alpha$ (α is the contrast ratio). Then we have

$$I_1'=I_0+II_1=I_0(1+\overline{\alpha-1}I_1)$$

$$I_2'=I_0+II_2=I_0(1+\overline{\alpha-1}I_2)$$

For $\alpha=2$ and $I_2' \geq .8I_2'$ we have $I_2=-.2+.8I_1$, and $X=.55$. The reduction in contrast has reduced the resolving power by about 20 percent.

The same basic technique used to determine resolving power can be used to determine the visibility of a dark object on a bright background. Thus if the object field consists of a strip of width d separating two semi-infinite areas, as shown in figure 19.3, the intensity dip at the center of the strip is given by the difference between PX

TABLE 19.1.

x	I_1	I_2	x	I_1	I_2
1.0	1.000	0.000	0.35	0.440	0.560
0.9	0.963	.037	.30	.414	.586
.8	.896	.104	.25	.462	.538
.7	.812	.188	.20	.538	.462
.6	.715	.285	.15	.486	.514
.5	.609	.391	.10	.511	.489
.4	.495	.505	.05	.508	.492

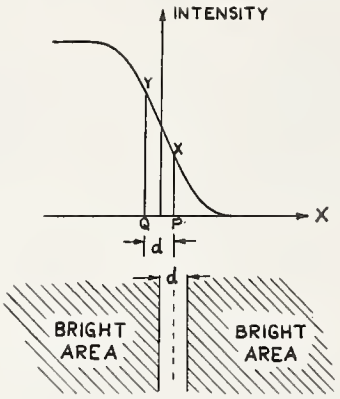


FIGURE 19.3. *Light intensity in a narrow gap.*

and QY. With a suitable criterion for a perceptible dip in intensity it is a simple matter to determine the least angular width of a dark strip that may be detected against a bright background.

The intensity gradient across an image boundary may be found quite readily from the computed lens data, or determined experimentally. As far as the computed intensity gradient is concerned, it may be evaluated purely on the basis of geometrical optics, or on the basis of diffraction theory and physical optics. These methods of evaluation will be considered in turn.

At any point in the focal plane of the lens the deviations from the principal rays, of rays through other points in the entrance pupil, may be evaluated by standard ray tracing. A considerable economy in ray tracing is effected if an aberration function of suitable form is adopted to represent these deviations. A small number of rays then furnish aberration constants from which the points of intersection with the focal plane of a large number of rays may be readily determined. A convenient way in which to represent the deviation components X and Y is the following:

$$\Delta X = \xi a_0 + a_1\xi\eta + \xi(a_2\eta^2 + b_0\xi^2) + \xi\eta(a_3\eta^2 + b_1\xi^2) + \xi(a_4\eta^4 + c\xi^3) + b_2\xi^2\eta^2$$

$$\Delta - 1 = \eta\alpha_1 + (\alpha_2\eta^2 + \beta_0\xi^2) + \eta(\alpha_3\eta^2 + \beta_1\xi^2) + (\alpha_4\eta^4 + \beta_2\eta^2\xi^2 + \gamma_0\xi^4)$$
$$+ \eta(a_5\eta^4 + \beta_3\eta^3\xi^3 + \eta a_5\eta^4 + \beta_3\eta^2\xi^2 + \gamma_1\xi^4)$$

This is the most general formula taking in terms up to the fifth order of the aperture, and is in effect a determination of the characteristic function relative to a principal ray. The coefficients are functions of

the position of the image in the field of the lens. To determine them at any field position 16 rays have to be traced. This is no great problem with automatic computing machinery. Using these coefficients we can determine the points of intersection with the focal plane of rays that come through a large number of pupil points. We have prepared tables for a less general aberration function, and calculated the intersection points of 64 rays in each half of the exit pupil, for a total of 128 intersection points, the calculation being carried out with a Marchant desk calculator. We estimate that with the more general aberration function, using an IBM 604 Electronic Calculating Punch, we can calculate the intersection points of 200 rays from each half of the pupil, giving a total of 400 intersection points, in about 15 minutes. Prepunched decks of cards are used on which the aberration coefficients are subsequently punched.

Given these intersection points we can prepare spot diagrams of the type used by Herzberger and by Linfoot and Hawkins.

For present purposes they can be used to determine edge gradients and then resolving power, using the technique already described. Thus if we have the spot diagram shown in figure 19.4 for the points of intersection of rays from an object point, and if we have a semi-infinite bright area terminating on AA', then the intensity at P is given by the fraction of image points lying to the right of the line $P'P''$ through P parallel to AA'. By varying the position of P, and so of $P'P''$, and counting the number of points to the right of $P'P''$ we get the intensity gradient across AA'. The counting can be carried out quickly on an IBM Sorter.

To summarize the method of evaluating resolving power we have the following steps:

1. Trace sufficient rays to determine the coefficients in an aberration function.
2. Determine a sufficient number of points in the focal plane, using the aberration function.
3. Determine the edge gradient across the boundary between a light and dark area by counting the points in a focal plane to the right of a line parallel to the boundary.
4. Determine the intensity, due to a repeated pattern of lines and spaces, by erecting ordinates to the edge gradient curve with the spacing of the test pattern and summing these alternatively positive and negative.
5. Carry out the process described in step 4 for points corresponding to the centers of lines and spaces. The resolving limit is achieved when the ratio of these intensities surpasses a prescribed value. The procedure described in step 4 must be modified, of course, to take into account finite contrast, in the way previously described.

The above treatment has been described for the case where the ray theory of geometrical optics is applicable and gives a sufficiently close account of image formation. It may readily be extended, in theory, to the case where the image intensity is determined by diffraction theory. Thus if the intensity is $I(x, y)$ at the point (x, y) in the neighborhood of the Gaussian image X of an object point,—then the intensity of $I(P)$, at a point P near the boundary of a bright area and a dark area, as shown in figure 19.4 is given by

$$I = \int I d\sigma,$$

where $d\sigma$ is an element of area, as shown, and the integration is carried out over the shaded area shown in figure 19.4. This integration may be carried out numerically if $I(x, y)$ is known at a sufficient number of points within this area. Evaluating $I(x, y)$ at a large number of points is a tedious process even with automatic computing machinery commercially available. With the IBM 604 Electronic Calculating Punch, for example, we estimate about 4 minutes per point in the image plane, the integration over the pupil being replaced by summation over 400 points in the pupil. We have no estimates of how long the process would take on SEAC or any like computer. The time taken by Maréchal's analogue computer is 8 minutes per point in the image plane, and is restricted to a narrow class of aberrations.

The equation for $I(P)$ given above leads immediately to an experimental method of determining the edge gradient, namely by using the knife-edge test. If a knife edge is traversed from right to left across the image of a fine object point, then the intensity at a point such as P due to the area bounded by AA', is the light transmitted past the knife edge when it has moved from $S'S''$ to $P'P''$.

Under suitable circumstances the same basic technique can be applied to determine the resolving power of a film-lens combination.

Thus if we have an evenly illuminated area of film, sharply bounded by the line AA', the film density across the line AA' will not show an abrupt change, but will be represented by the curve shown in figure 19.5, a. Suppose that we can consider this curve as resulting from the operation of two factors, without specifying in detail the physical basis underlying them. As a result of the first factor, which may be connected for example with scattering in the emulsion, points near the boundary AA' receive an effective exposure represented by the curve shown in figure 19.5, b. Exposure here is measured in terms of incident light energy, not in terms of its logarithm. The second factor

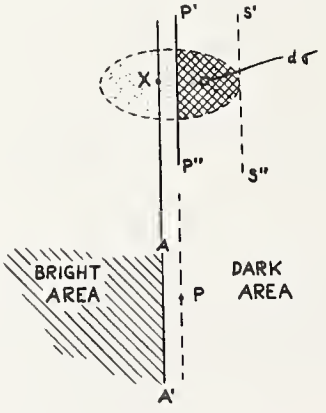


FIGURE 19.4. *Relation of light intensity to image patch.*

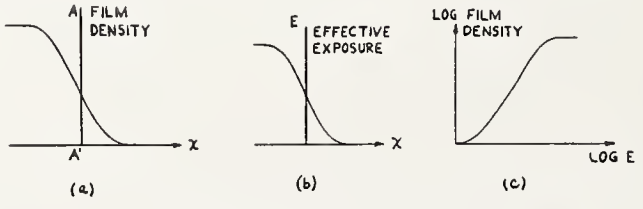


FIGURE 19.5. *Combination of factors into film-density curve.*

conditions the response of the emulsion to the effective exposure, and may be of the general form of the H and D curve of the emulsion, showing an initial toe and a region of saturation, as shown in figure 19.5, c.

We have to determine the effective exposure at a point due a given pattern of radiation incident on the emulsion. Again we may proceed by determining first the effective exposure at a point due to a semi-infinite bright area bounded by a line AA', taking into account the intensity curve $I(x)$ due to the finite size of the light patch, and the effective exposure curve $E(x)$ due to the scattering another factor previously defined. The effective exposure $E(y)$ at a point distant y from the boundary AA' is given by

$$E(y) = \int I(x) \cdot \frac{dE(y-x)}{dx} \cdot dx$$

taken over all values of x for which the integrand does not vanish.

The geometrical interpretation of this equation is that if we graph $I(x)$ as ordinate against $E(y-x)$ as abscissa, then $E(y)$ is the area under the graph. To evaluate this geometrically for a range of values of y we proceed as follows. With the same abscissa x draw the graphs of $I(x)$ and $E(-x)$. Erect ordinates at a suitable number of points x_1, x_2 . . . x_w and read off values of $I(x)$ and $E(-x)$ to serve as ordinates and abscissae in the $I(x)$ versus $E(-x)$ graph. By erecting ordinates at a distance y, positive or negative, from the original ordinates, and by reading $I(x)$ from the original ordinates and $E(-x)$ from the new ordinates we get the appropriate values of $I(x)$ and $E(y-x)$ to insert in a graph such as that shown in figure 19.6. In one extreme position the graph becomes the straight line MN; in intermediate positions it takes the form shown by the dotted lines; and in the other extreme position the graph degenerates to the origin O. The parameter y defines a family of curves, and the area under the curve defined by a particular value of y represents the effective exposure at a point distant y from the boundary of an illuminated area.

Given the effective exposure due to a semi-infinite area we can proceed as before to find the effective exposure due to a repeated pattern of lines and spaces. Using the response curve, which relates the image density to the effective exposure, we can determine the image densities due to the repeated pattern of lines and spaces. By adopting a criterion of resolution in terms of image density we can proceed in this way to determine the resolving power of the lens-film combination.

It is readily seen that the effective exposure curve obtained in this way extends over a distance x_1+x_2, where $I(x)$ and E(x) extend over distances x_1 and x_2. This corresponds to the rough rule given by Katz that the reciprocal of the resolving power of a combination is the sum

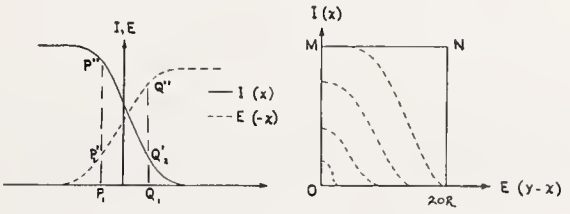


FIGURE 19.6. *Density distribution on film of finite resolving power.*

of the reciprocals of the components. The exact relation between component and combination resolving powers depends on the detailed forms of $I(x)$ and $E(x)$. Under the conditions stated it is not a matter of particular difficulty to establish this in any given case when the forms of $I(x)$, $E(x)$ and the response curve are known. Numerical and graphical methods have to be used, since the determination of the area under the $I(x)$, $E(y-x)$ curve does not as a rule lend itself to simple analytical treatment.

The technique described enables us to predict the resolving power of a lens when its aberrational characteristics are known. When the lens performance is described adequately by geometric optics the determination of resolving power may be carried out quickly using an IBM 604 Electronic Calculator. When the lens performance is governed by diffraction theory, and if the diffraction pattern is known, a simple adaptation of the method enables us to predict resolving power in this case also. If we make reasonable assumptions about the photographic emulsion we can obtain the resolving power of the lens-film combination.

20. A Proposed Approach to Image Evaluation

By R. V. Shack [1]

Introduction

The basic element in the formation of any general image is the image of a point. The description of a point image in terms of the aberrations affecting it evaluates the imaging properties of the system. This is the information with which the lens designer is most concerned, but the user obtains from it only a rough qualitative idea of the imaging properties as far as general usage is concerned. That is, he knows that if the aberrations are small, the complex image he obtains is better than if they are large. However, he finds it difficult to establish a good psychophysical correlation between this type of evaluation and the quality of the complex image he generally observes.

For quite some time the user has had his own way of evaluating an image forming system. He determines its ability to resolve the components of a repetitive object, the most common type now in use consisting of alternate black and white lines of equal width. A resolving-power test has several virtues. It is simpler and more rapid to perform than any other existing test, and it gives the user an indication of the size of the finest image detail he can obtain.

However, such a test has its disadvantages also. It is very difficult for a designer to evaluate his design in terms of resolving power. The resolving power obtained in a given test depends not only on the imaging properties of the system, but also on the contrast of the object and the amount of superimposed, nonimage-forming light falling on the image plane. The fact that it is a threshold phenomenon results in a low degree of precision in its determination and variability from observer to observer. It is in a sense an artificial test, even to the user, because he is in general not interested in the ability to separate similar objects but in the ability to image random detail. It has been found recently that the correlation between resolving power and the subjective feeling of sharpness in the image is not as good as it had been assumed to be.

Consequently, considerable work has been done in analyzing the imaging process in order to improve testing procedures. The following approach is based on the image of a straight edge, the basic method of image derivation being suggested by Cox.[2] This was chosen because its properties are two-dimensional, thereby simplifying the analysis, and because its use makes the presentation of the concepts involved much simpler than they otherwise would be. The information obtained from a straight-edge image, or edge gradient, also has the advantage of lying midway between the type of information contained

---

[1] National Bureau of Standards, Washington, D. C.
[2] A. Cox, Technical Report No. 102, Farrand Optical Co., Inc.

in a point image, of interest to the designer, and the type of information that the user is interested in. The light distribution in the edge gradient can be obtained directly from the light distribution in a point image, and yet the nature of the edge gradient is of direct interest to the user because much of the structure of his general image consists of edges.

The Edge Gradient Concept

In figure 20.1, a, the upper drawing represents the light distribution in a straight-edge object, where the light passing the straight edge is assumed to be superimposed on a general background light. The lower drawing represents the light distribution in the image. This curve is the edge gradient. In figure 20.1, b, the straight edge is assumed to block out light from the general background. The broken curve in the lower drawing might be called an edge gradient attenuation curve, because it represents the amount of light to be subtracted from the background to yield the edge gradient, which is shown as a solid curve.

Single Line Image

1. *The formation of a line image.* A line can be considered to be a combination of two straight edges. In figure 20.2, a, the solid curve in the center drawing represents the edge gradient for the left edge of the line, as if the right edge were indefinitely far away. This light

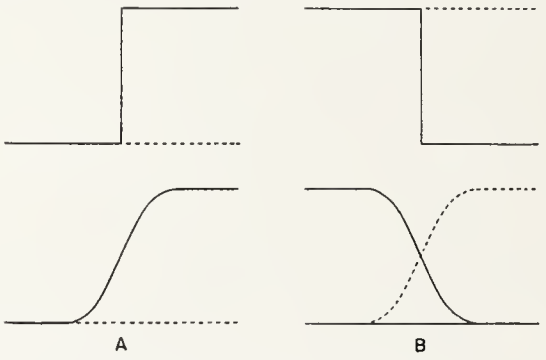


FIGURE 20.1. *The edge gradient.*

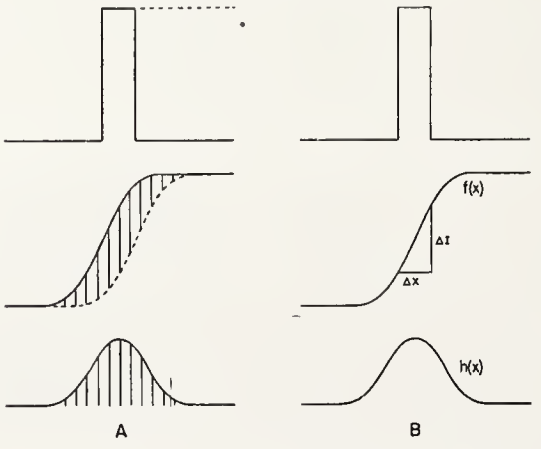


FIGURE 20.2. *Derivation of a line image.*

distribution is attenuated according to the dashed edge gradient attenuation curve for the right edge of the line. The remaining light is the light actually existent in the image of the line. The point by point ordinate difference between the two curves yields the image curve for the line, as shown in the bottom drawing.

Now a little reflection will show that the dashed curve is identical in shape to the solid curve. The two curves are separated by the width of the line, so the ordinate differences could just as easily have been obtained by using one curve and obtaining ordinate differences between points on the curve whose abscissa difference is equal to the line width. This is illustrated in figure 20.2, b.

The ordinate difference ΔI is equivalent to the illumination at the corresponding point in the image; Δx is the width of the idealized geometrical image. These two methods for obtaining the image curve are equivalent.

2. *Variation in the line image with line width.* The edge gradient can be represented by a function $f(x)$. The line-image curve can be represented by a function $h(x)$, which will, in general, be different for each Δx, diminishing as Δx approaches zero. This is illustrated by the curves in figure 20.3, a. The marks on the curves indicate the magnitude of the Δx associated with each curve. If each curve is divided point by point by its associated Δx, the set of curves in figure 20.3, b, will be obtained (actually, the curves as shown were divided by $2\Delta x$ for convenience in presentation). The result of this process is that the area under each curve is equal to the area under any other curve in the set. The heavy curve at the bottom corresponds to the distribution of the illumination in the image of an infinitesimally fine line.

3. *The relation between the fine-line image and the edge gradient.* This last curve is the limit as Δx approaches zero of $h(x)/\Delta x$. But, as obtained, $h(x)=f(x+\Delta x)-f(x)$. Therefore, by the definition of a derivative, this limit is the first derivative of $f(x)$. That is, the curve

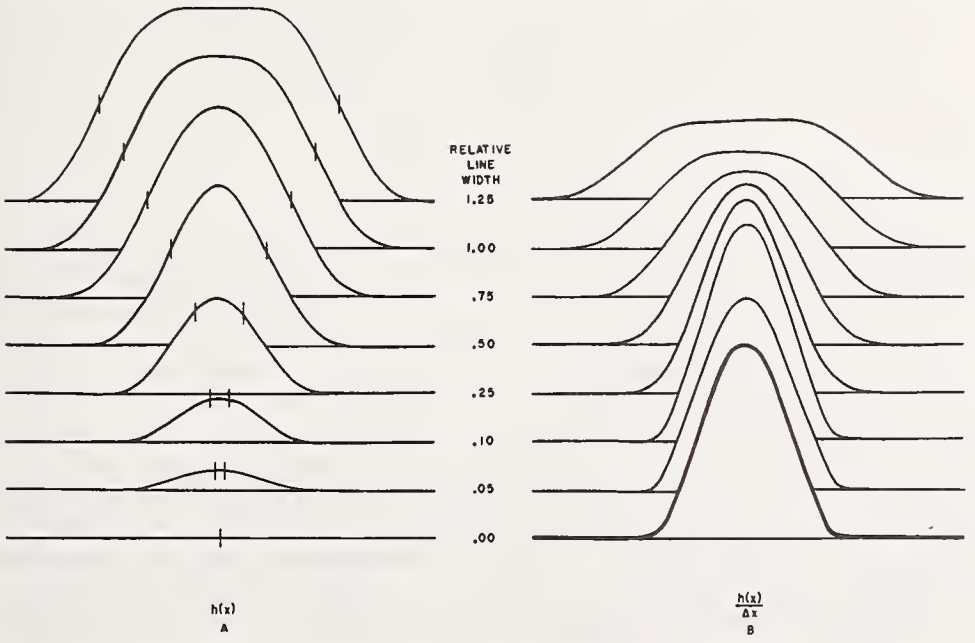


FIGURE 20.3. *Line-image variation with changing object line widths.*

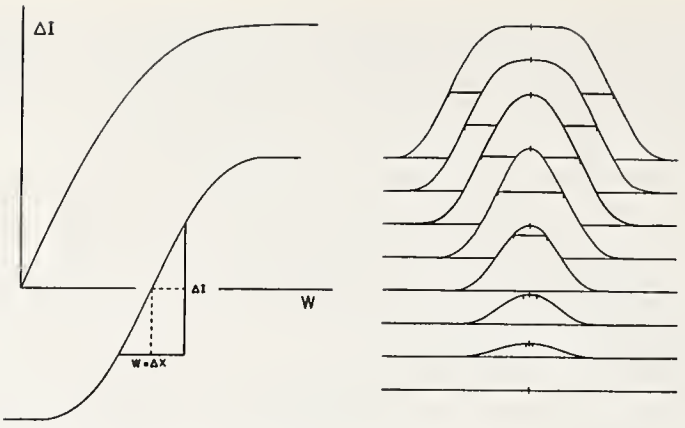


FIGURE 20.4. *Central illumination increment versus ideal image line width.*

of the derivative of the edge gradient has the same shape as the distribution of the illumination in a fine-line image.

4. *The relation between the illumination at the center of a finite line image and the edge gradient.* If the illumination at the centers of the line images in figure 20.3a are plotted against each corresponding Δx, a curve is obtained such as that in figure 20.4. Hereafter, Δx will be called W, the width of the ideal image line.

It will be noticed that, assuming the edge gradient to have point symmetry about its center, each half of the edge gradient may be obtained by plotting $\frac{1}{2}\Delta I$ against $\frac{1}{2}W$. If the edge gradient is not symmetrical, the plot of $\frac{1}{2}\Delta I$ against $\frac{1}{2}W$ will yield a curve of average ordinates between the two halves, assuming that the lower half is rotated 180° about the center to bring it into near coincidence with the upper half. This curve is identical to the curve obtained by plotting the line centers against the widths except that it is half the size.

It is convenient to normalize the ordinate in order to make the result independent of the brightness of the object line. The ordinate then becomes $K=\Delta I/(\Delta I)_{max}$, which varies from 0 to 1.

5. *Summary—correlations.* This curve of K versus W then is a simple curve, characteristic of the imaging process, and independent of variations in the contrast of the object, or, for that matter, any amount of nonimage-forming light. This is true because K is a ratio of the differences between two pairs of illumination levels and so is independent of change of illumination, which acts on all the levels either proportionally or additively. The curve is first of all a plot of the variation in the relative illumination at the center of a line image versus the ideal width of the line. By section 4, its shape is the same as that of the edge gradient, or very nearly so, depending on the symmetry. And by section 3, it is a plot of the relative amount of flux contained within the limits of W in the fine-line image. These relations are illustrated in figure 20.5. For example, at a point corresponding to a K of .5, the W obtained is the width of the ideal line image where the corresponding actual line image has only half the relative peak illumination. The same W is the width in the fine-line image, which contains 50 percent of the total flux. And if there were some good method of obtaining a numerical value of the edge gradient, it could also be applied to this K versus W curve.

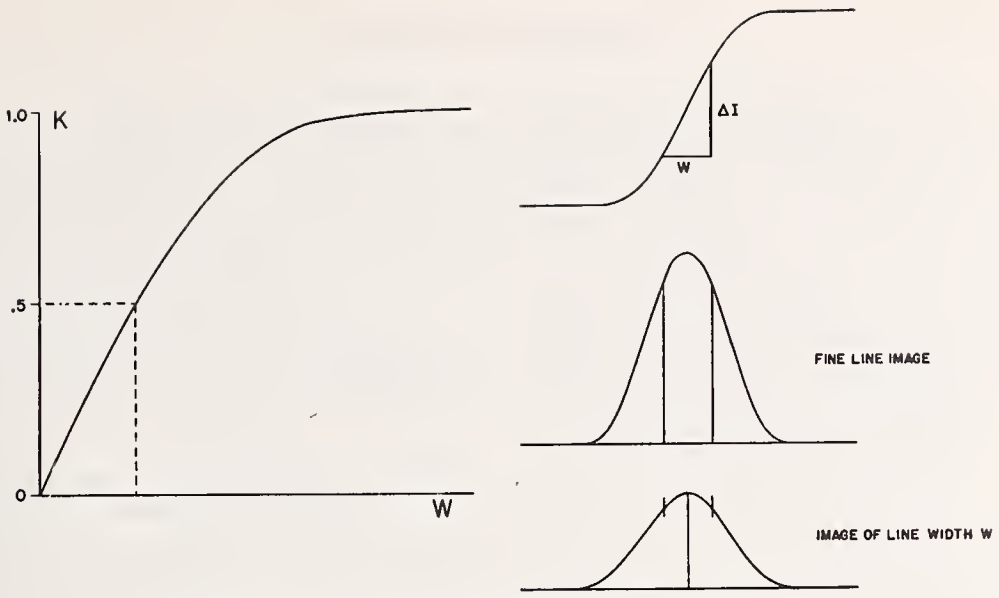


FIGURE 20.5. *Line image, edge gradient relationships.*

Therefore, such a curve is suggested as the basis for a new approach to image evaluation. The lens designer can plot the curve for a design if he knows the light distribution in the point image. This means that he can predict the performance in terms that convey considerable information to the user. This information is, of course, the nature of the image of an edge, and the ability to reproduce isolated fine detail.

6. *Image-object contrast relations.* Both the eye and a photographic emulsion can be assumed to be linearly sensitive to the logarithm of the illumination in the image. A convenient definition of the contrast between two illumination levels is the difference between their logarithms, or the logarithm of their ratio, which amounts to the same thing. Photographically, this is equivalent to the density difference between the corresponding parts of the image divided by γ.

The relation between the object brightness ratio R_o and the image illumination ratio R_i is given by the following equation,

$$K=(1+\alpha)\frac{(R_i-1)}{(R_o-1)}, \tag{1}$$

where K is the image evaluation function obtained above; α is a measure of the nonimage-forming light present; log R_i is the image contrast; and log R_o is the object contrast.

This equation, derived in appendix 1, applies to dark lines in a bright field as well as bright lines in a dark field.

All the quantities considered in eq 1 are ratios, and so anything affecting illumination proportionally, such as transmission loss in the lens, does not enter into the situation. In other words, such an effect does not alter the contrast.

The quantity $(R-1)$ is what is usually considered the optical contrast. Log R has the advantage of a better psychophysical correlation.

In dealing with photographic images, an emulsion contrast might be defined as γ log R. This would be simply the density difference in the emulsion.

518-743 O - 73 - 32

Multiple Line Image

1. *The formation of a multiple-line image.* In this section, only patterns having equal line and space widths will be considered.

The approach that was used to obtain the image of a single line can be extended to images of a series of parallel lines. As illustrated in figure 20.6, if the center of the edge gradient is located at the center of one of the lines, then the illumination in the image at that point is given by the sum of the components produced by each line within the effective range of the edge gradient. An equivalent illumination is obtained at the center of a space. Curves for both the line response and the space response may be plotted as K_l' and K_s' versus W, where the K's are analogous to the K obtained for a single line, the prime being used to distinguish them.

If the lines and spaces are equal in width, then $K_l'+K_s'=1$, and the curves are symmetrical about $K'=.5$, as shown in figure 20.6. The point where the curves first meet as W diminishes is the physical limit of resolution, for it is there that both line and space images have the same illumination. Regions where the curves have crossed each other show line patterns, but the resolution here is spurious. This occurs because of the repetitive nature of the object.

2. *Image-object contrast relations.* The contrast relations for a multiple-line image derived in appendix 2, are a little more complex than they are for the single-line image. They are as follows.

$$\rho=(1+\beta)\left(\frac{M_i}{M_o}\right). \tag{2}$$

ρ is a contrast reduction function defined in the same way as Schade's amplitude response factor.[3] It is equal to $2K_l'-1$. β is a measure of the nonimage-forming light present. M_i is an image contrast factor, where $M_i=(R_i-1)/(R_i+1)$, log R_i being the image contrast. M_o is the same type of factor for the object contrast.

That this equation is valid is indicated by figure 20.7. The three graphs on the left are three plots of the same experimental data. They were plotted this way so as to indicate the nature of the function involved. The image contrast is the contrast in the emulsion. The three graphs on the right, plotted from eq 2, are based on entirely theoretical material, an analytical function being chosen for ρ as a function of line frequency, which is similar to the ρ of the experimental

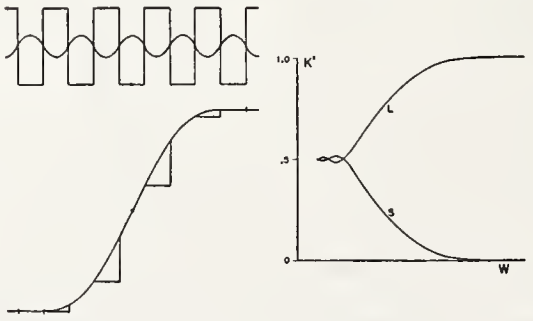


FIGURE 20.6. *Derivation of a multiple line image.*

[3] O. H. Schade, RCA Rev. June, 1948.

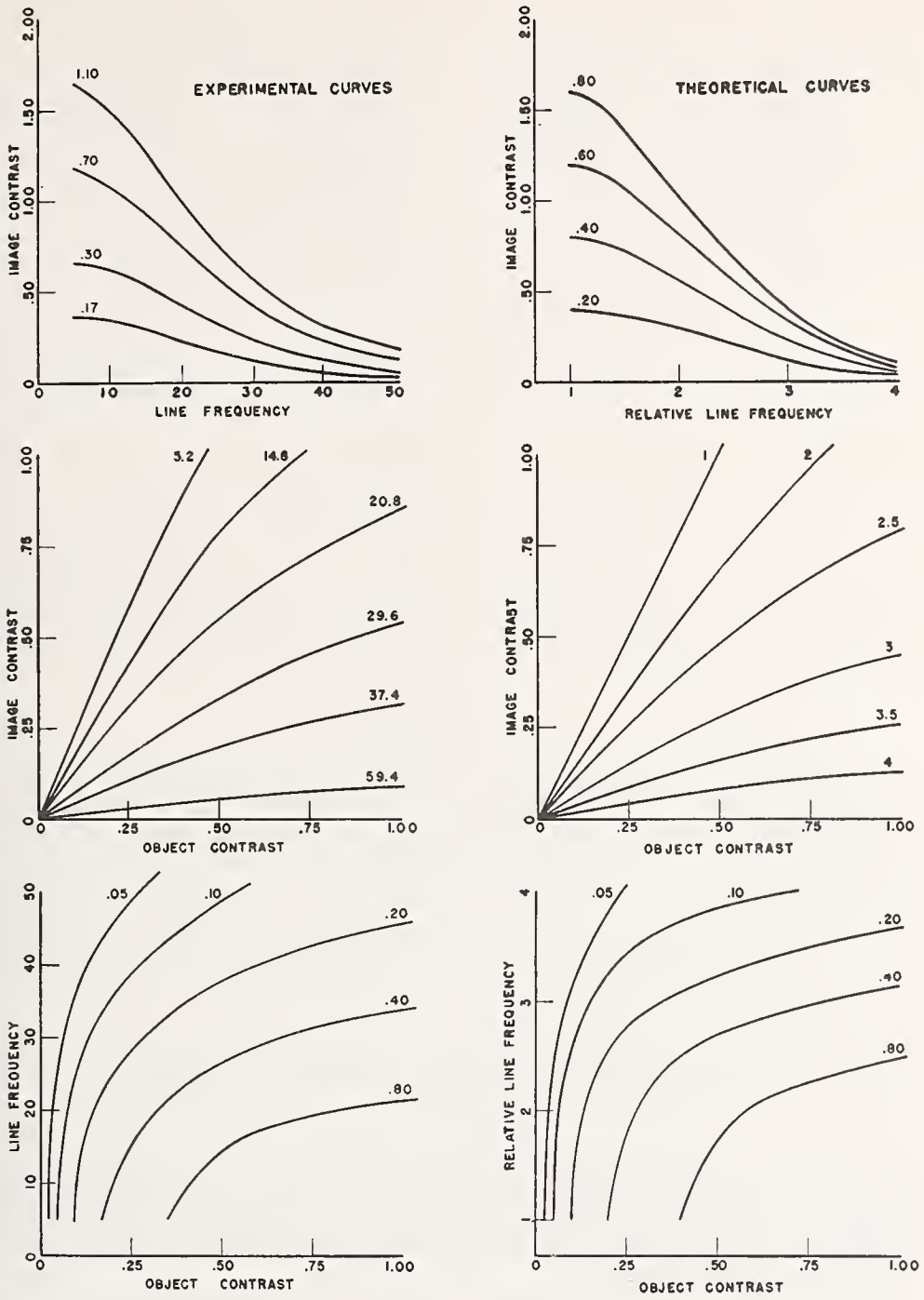


FIGURE 20.7. *Comparison of experimental and theoretical contrast relationships.* The experimental data is for a 127-millimeter Ektar lens on V-G emulsion.

curves. Also, for convenience, a γ of 2 was assumed. The nature of the theoretical curves seem to agree quite well with the nature of the experimental curves.

3. *Correlation between the multiple-line image and the one-line image.* Figure 20.8 shows the relationship between the K-curve for the one-line image, plotted as a solid line, and the K'-curves for the

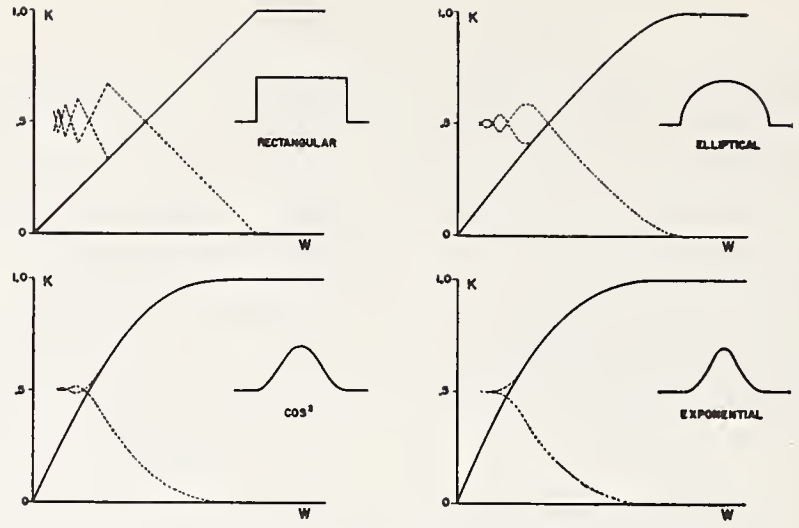


FIGURE 20.8. *Relation between single and multiple line images.*

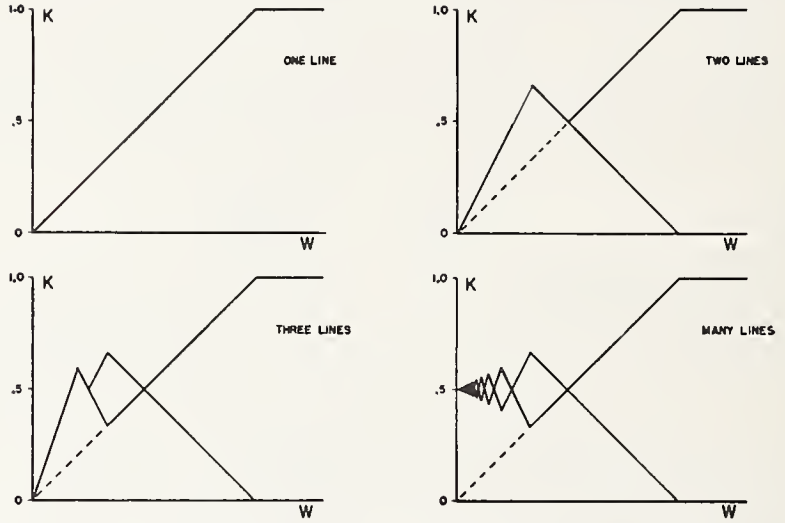


FIGURE 20.9. *Comparison of two- and three-line images with the multiple line image.*

multiple-line image, plotted as dotted lines. K_l' is identical to K except for a small region of W smaller than its value for a point near the limit of resolution. The inserts are profiles of the fine-line images associated with each graph.

4. *Two- and three-line images.* Figure 20.9 indicates the properties of two- and three-line images. In general they are identical with those of the multiple-line image until the pattern is so fine that the width of the edge gradient (or fine-line image) overlaps the entire pattern of two or three lines. For smaller patterns the effect is approximately as if the entire pattern were replaced by one broad line.

Figure 20.10 illustrates the nature of the image of a three-line pattern as the line width in the pattern relative to the width of the edge gradient diminishes. All the curves are drawn with the abscissae expanded so that the apparent widths of the object lines are the same for all the patterns.

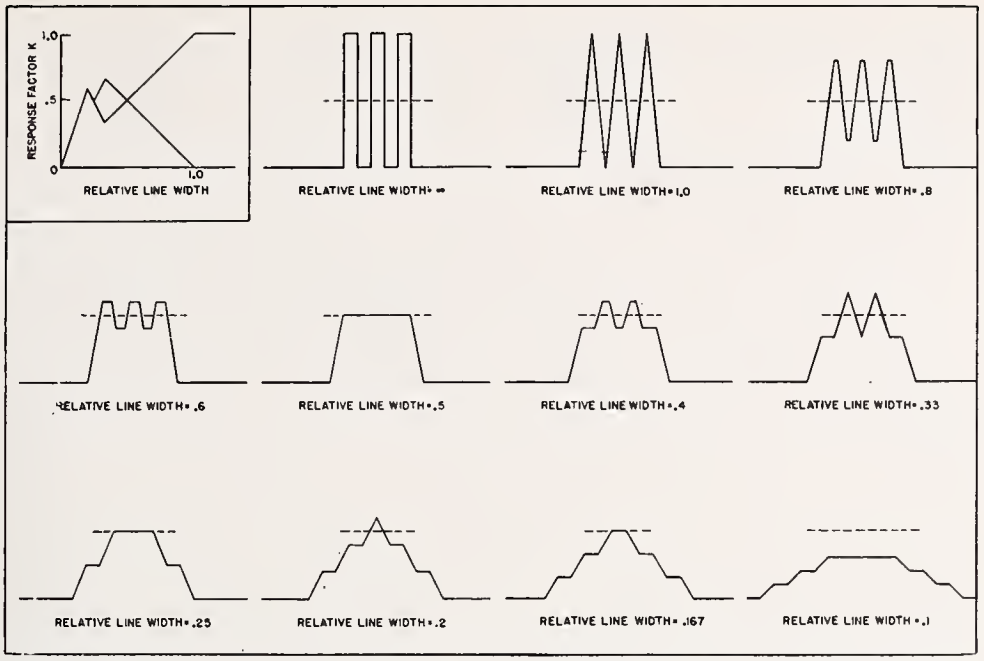


FIGURE 20.10. *Variation in appearance of three-line image with changing pattern size.*

Summary

This analysis accomplishes two things. First, it suggests a new approach to image evaluation, which conveys considerable information to the user and yet is close enough to the designer's approach for him to predict the performance of his design in terms of the new approach. Second, it provides an explanation for the phenomena involved in a resolution test, thereby linking the new approach to the older resolution test. Experimental work is going on at present to investigate the validity of the assumptions and conclusions obtained.

Appendix 1. Derivation of Image-Object Contrast Relations for an Isolated Single Line

Let B_1 be the background brightness and B_2 the line brightness in the object. The line may be either brighter or darker than the background. The corresponding illuminations in the image are $I_1=kB_1$ and $I_2=kB_2$, assuming no image degradation. Let I_l be the illumination in the center of the actual line image, and i be the illumination in the image plane produced by nonimage-forming light.

Furthermore, let the following symbols be defined as indicated.

$$K\equiv\frac{I_l-I_1}{I_2-I_1} \tag{1}$$

$$\alpha\equiv\frac{i}{I_1} \tag{2}$$

$$R_o\equiv\frac{B_2}{B_1}=\frac{I_2}{I_1} \tag{3}$$

$$R_i\equiv\frac{I_l+i}{I_1+i}. \tag{4}$$

Then

$$K=\frac{(I_l+i)-(I_1+i)}{I_2-I_1}=\left(\frac{I_1+i}{I_1}\right)\left(\frac{R_i-1}{R_o-1}\right)=(1+\alpha)\left(\frac{R_i-1}{R_o-1}\right). \tag{5}$$

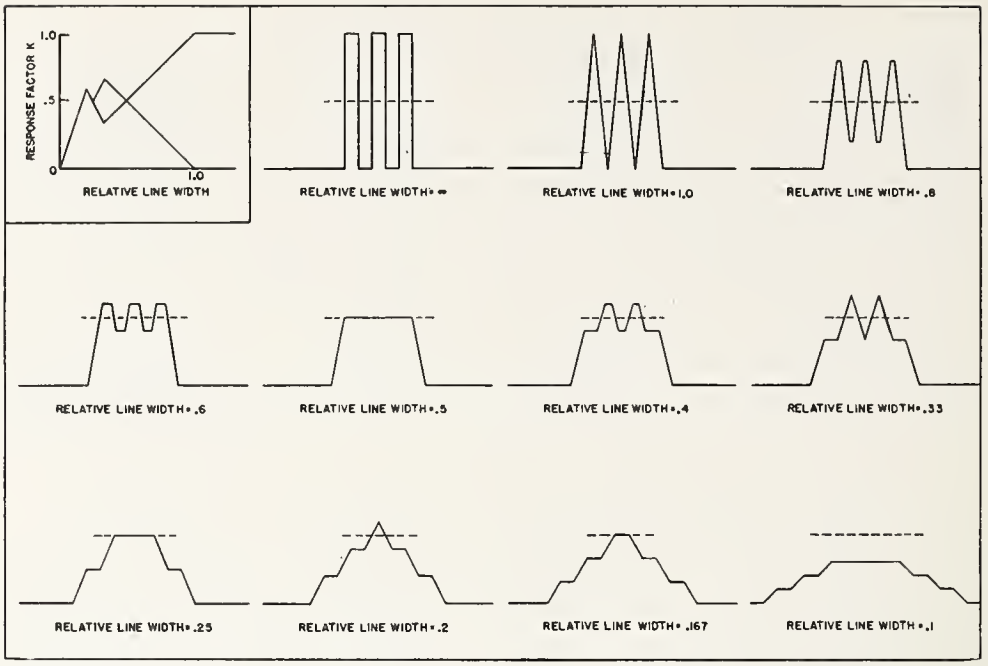


FIGURE 20.10. *Variation in appearance of three-line image with changing pattern size.*

Summary

This analysis accomplishes two things. First, it suggests a new approach to image evaluation, which conveys considerable information to the user and yet is close enough to the designer's approach for him to predict the performance of his design in terms of the new approach. Second, it provides an explanation for the phenomena involved in a resolution test, thereby linking the new approach to the older resolution test. Experimental work is going on at present to investigate the validity of the assumptions and conclusions obtained.

Appendix 1. Derivation of Image-Object Contrast Relations for an Isolated Single Line

Let B_1 be the background brightness and B_2 the line brightness in the object. The line may be either brighter or darker than the background. The corresponding illuminations in the image are $I_1=kB_1$ and $I_2=kB_2$, assuming no image degradation. Let I_l be the illumination in the center of the actual line image, and i be the illumination in the image plane produced by nonimage-forming light.

Furthermore, let the following symbols be defined as indicated.

$$K\equiv\frac{I_l-I_1}{I_2-I_1} \tag{1}$$

$$\alpha\equiv\frac{i}{I_1} \tag{2}$$

$$R_o\equiv\frac{B_2}{B_1}=\frac{I_2}{I_1} \tag{3}$$

$$R_i\equiv\frac{I_l+i}{I_1+i}. \tag{4}$$

Then

$$K=\frac{(I_l+i)-(I_1+i)}{I_2-I_1}=\left(\frac{I_1+i}{I_1}\right)\left(\frac{R_i-1}{R_o-1}\right)=(1+\alpha)\left(\frac{R_i-1}{R_o-1}\right). \tag{5}$$

Any method of calculating the light distribution is valuable. One advantage of the method that I happened to select for my paper is that this method gives the characteristic function, and thus a complete mathematical analysis of the optical system.

Dr. D. S. Grey, Polaroid Corp., Cambridge, Mass.: Dr. Herzberger just touched in his last remark on what I wanted to say. His method is the equivalent of getting response characteristics on the lens whereas what you get is resolving power. If you admit in the beginning that the response characteristic is what you want, you can replace the integration by multiplication. The response-characteristics method is not intended to make work, it is intended to save work.

Dr. R. E. Hopkins, Institute of Optics, University of Rochester, Rochester, N. Y.: I asked Dr. Grey if he was going to start computing response factors for every lens he designed and he made the statement that he was not going to do it for every lens but he was going to do it for some of the lenses. He has already done some, but what he hopes is that this theory is going to give him rules of thumb for designing better lenses. I thought that was a wonderful way of putting it. I think that studies such as Cox is doing on the high-speed computing machine are going to give valuable information, but I think we must remember that not everybody is going to be able to use these machines. Since you are spending taxpayers' money to get this data—and I am a taxpayer—I think it is up to you to publish this data so I will be able to get this rule of thumb for making better lenses.

Mr. A. Cox, Bell & Howell Co., Chicago, Ill.: The equipment we are using is an IBM Commercial Unit, not the rather glorified equipment that the Bureau of Standards can afford.

Chairman B. O'Brien, University of Rochester, Rochester, N. Y.: Even so, it is your duty to publish it.

Dr. G. H. Conant, Jr., Harvard College Observatory, Cambridge, Mass.: I would like to ask Mr. Cox a question, if I may. What digital accuracy do you find it necessary to carry in tracing the rays through the system to get the data that are later evaluated?

Mr. Cox: As a rule we planned on using machine 604. The work will be described in Chicago. We have used 60 program sometimes and for spherical surfaces, using the 604, we go to five-figure accuracy, which is quite sufficient for most photographic lenses. That takes about 16 seconds per surface. When we need to pick up the accuracy, then we do another run through four different boards, which provides us with an extra two-figure accuracy.

Dr. K. V. Pestrecov, Bausch & Lomb Optical Co., Rochester, N. Y.: My question applies to three or four speakers who talked about analysis of response factors, energy distribution and about predicting the resolution. However, perhaps I am presenting the question too soon. I am interested in why we did not have any tables here computed at a 25-degree resolution on double X. We actually obtained 13. Why is there no experimental confirmation or data? Maybe it is too early to have them. I would like to see them to know where we really stand in actual experimentation with these suggested methods for computing resolving power.

Mr. Cox: So far the results that have been computed have been on classified military instruments. We have found reasonably good agreement on a rather simple basis. With military instruments we would be working on or near the axis.

Dr. Pestrecov: It would be on or near the axis. I probably can design a lens that will resolve to the Rayleigh limit. I am a little lost as to what to expect from a lens of large field angle. I would like to know if you computed such lenses?

Mr. Cox: We computed an F.2 lens at 15 degrees and got close agreement.

Dr. Pestrecov: Thank you.

Dr. Grey: What else could have gone wrong? Are you asking whether we know our arithmetic?

Dr. Pestrecov: Well, on paper all of the theories sound very plausible. I don't know whether you have experimental confirmation but I would like to know how they work when you have extremely complicated aberrated images—I am still somewhat doubtful that your resolution measurements given in comparison will be correct. I am going to suggest a test at 25 degrees with an extremely complex image pattern and I would like to see somebody predict the resolution, at 15 degrees. Has anybody done anything like that?

Dr. Grey: I haven't had any complaints and I have used this. Even in tests 15 degrees off axis I would expect that the physical optics would enter in an order of magnitude that would be about the discrepancy you would expect. We don't know exactly what the response characteristic of the emulsion is so there are two sources of error. One is the emulsion and the other is the use of geometrical optics rather than physical optics.

Dr. J. G. Baker, Harvard College Observatory and Perkin-Elmer Corp., Norwalk, Conn.: I should like to add a few remarks. I have studied the effect of chromatic aberrations on resolving power because color addition is an important part of the problem. Also, in the presence of aberrations of large magnitude it is necessary to consider the particular resolving-power target rather than just the intensity distribution across an edge. One has to integrate over the pattern as a whole in order to determine what its photographic image will be.

There is another point I want to make. For a number of years I have been using at Harvard a method of testing related to the one Dr. Cox describes. I have called it the "Visual Hartmann Test." Instead of integrating across the image, in this particular form of test I have integrated across the exit pupil. The set-up requires a linear scale across the meridional diameter of the exit pupil and a micrometer controlled knife-edge movement in the vicinity of the mean image. Observing as in the Foucault test, one plots the positions of all bright-dark edges seen in the exit pupil according to scale readings against the micrometer reading. The knife-edge is moved in chosen small increments across the image. The resulting graph represents in effect the slopes of the wave-front. By integrating under the curve, one obtains in effect the shape of the wave-front error in fringes or in any other convenient units, and can correct to any other nearby focal setting. The test is particularly useful for aspheric figuring and I have used it during 1937 to 1939 perhaps several hundred times. I have used it also in the area figuring of a completely unsymmetrical optical system. The test is a useful one in the laboratory.

Excerpt From Letter From T. Smith [1]

Dr. Gardner: It was my sincere wish that Mr. T. Smith might be present at this symposium. He is so well known personally to all of you from across the Atlantic and so well known by his publications to those on this side that his presence would have added greatly to the meeting as a whole, and in particular his contributions to the discussions of the papers would have been most illuminating and helpful to all of us. I know that all here join me in regretting that Mr. Smith found it not possible to be with us. At my request he has prepared a letter for the symposium which I wish to read at this time.

"My hope is that on this occasion when you have present representatives of many countries, you may be able to present questions to them which they will like to consider with their colleagues when they have gone home, and that ultimately agreements may be reached which will be of service to all concerned with optical instruments.

"I will begin with a somewhat minor matter. I suppose that in these days almost all countries are interested in aerial surveys. We all desire the maps that are made to be highly accurate, and one of the factors necessary to this end is excellent correction in the photographic lenses used in the cameras. Perhaps each of the major countries has its own specification for the lenses used for this purpose. It would be interesting to know whether there is sufficient experience for all to agree on the relative importance of the various aberrations that must be kept within close limits. Since the focal lengths are greater than with most lenses intended to cover the same angular field with corresponding relative apertures, aberrations are more appreciable, and it becomes more difficult to keep the large number that may be significant within acceptable limits unless the construction becomes decidedly complex. Freedom from distortion seems to be an important property, and if this is achieved there seems to be no justification for demanding precise centering of the lens in the camera. The prints of the reference marks on the pressure plate enable every photograph to be placed in the same position relative to the lens, and this, rather than reference to the lens axis, is the essential requirement for the construction of accurate maps. It appears undesirable to insist on high accuracy in making an adjustment of some difficulty when this results in no improvement in the final product, and may possibly cause less perfect adjustment in ways that are of real importance, through the difficulty of securing close observance of many adjustments simultaneously.

"Another question that might be considered is whether tele-photo lenses should be used for taking photographs for maps. Unless the centre of the lens aperture coincides with a nodal point or its real image in part of the system the axis of the refracted cone of rays lies in a different direction from that of the incident cone, and this causes a displacement on the plate of the point representing a feature of the ground which lies out of the plane focused on the plate. The use of telephoto lenses for purposes other than precise mapping seems free from objection. I should be greatly interested in knowing what opinions are held now on these questions—it is some years now since I was in touch with these matters.

"The second subject on which I would like to say a word is the graphical representation of aberrations. It was natural that the representations of aberrations used by von Rohr and others should have been adopted in the early days when new constructions were being evolved, but it is less clear that we should now continue to employ them rather than record our results in a different way. We ought, I think, to consider in the first place what the purpose of these records is. If we only mean to give some idea to the general public of the state of correction of lenses of different designs it may not matter much how these outstanding defects are recorded. But if we are concerned in giving information that will be of value

[1] Roselyn, Holton, Wincanton, Somerset, England.

to technical workers the position is different. For example, if longitudinal central aberration is recorded with the first power of the aperture zone as independent variable, the central part of the curve means nothing, and the outer parts which are of real importance are considerably squashed up; a diagram of this kind no doubt appeals to the makers of lenses that are less good than the best obtainable. But apart from this a diagram of this kind is not particularly useful to technical workers. The prime consideration for work of high quality is that path differences should not exceed some definite limit—say a quarter of a wavelength of the light taken as a reference standard. If the longitudinal aberration is plotted with the cosine of the angle made by the ray with the axis as ordinate, the position of focus where the differences of path are a minimum can be found by drawing a straight line parallel to the ordinate axis so as to cut off equal areas between this line and the aberration curve, the axis and the extreme aperture being the other limits of these areas. In many cases the same use can be made of diagrams with the square of the aperture rather than the first power as the independent variable. Second-order ordinates should also be used to record coma. Similar criticisms can be made of the usual representations of curvature and astigmatism as well as of distortion. The central parts of these diagrams are useless, and the important parts made to look of little consequence. If the square of the angular field, or better still the cosine of the inclination of the principal ray of a pencil to the axis, were taken as variables the diagrams would gain greatly in value. For distortion the transverse displacement would be plotted as a fraction of the ideal distance from the axis.

"But there is another point to be considered bearing on the graphical representation of aberrations. In all the early work it seems to have been assumed that it would suffice to record only those aberrations which are of types represented in the aberrations which I call the first order but are very frequently named third-order aberrations. It happens that these can be exhibited as dependent on only one of the fundamental variables—some on the aperture and others on the field. Among the higher order aberrations are some which necessarily depend on both, and the way in which these aberrations—and they are becoming increasingly important in the development of modern instruments—can be represented graphically so as to be of value to lens designers is not clear. For this reason I am inclined to think that it is of more importance to give aberrations by means of the values of coefficients than by means of diagrams. The coefficients that occur in one of the Hamiltonial Characteristic functions—the directional function T has distinct advantages for this purpose—would meet all needs conveniently. In this connection it is not superfluous to point out that the theory of these calculations has been fully worked out; the real bar to their use in the past for all but the best known aberrations has been the numerical computation; but with the advent of the very powerful and rapid computing engines developed in recent years the labor aspect of this suggestion seems no longer important. It would, of course, become important to reach agreement on standard forms for the representation of those aberrations to which in the past so many of us have been willing to shut our eyes.

"I have taken the liberty of mentioning two or three optical topics that seem interesting to me, but I realize they may not fit at all into the program that you have in mind for the celebrations you are holding. I should like to leave them in your hands to deal with exactly as you think fit. I should in any case be most interested to hear your views on these and indeed on any other optical subjects that are discussed at your meetings."

Prof. F. Zernike, Natuurkundig Laboratorium, Groningen, Netherlands: Yes. Mr. Chairman, and attendants of the symposium. As one of the foreign guests who have so generously been invited to come to this country and attend this wonderful symposium that has I think, and you will all agree, far surpassed our best expectations, I want to express our great gratitude to all who have given their time and efforts to the organization of this symposium as well as to the National Bureau of Standards and the other organizations who have backed it and altogether enabled us to come here.

Of course, we see all these personified in Dr. Gardner and therefore I expressly address Dr. Gardner and tell him how much we have

enjoyed being here. We have not only listened to so many, perhaps too many, scientific papers, all presented in a very congenial way I would say, but we have especially, also, met old friends and have made new ones, and I think this personal note is also of great importance to the progress of the science we represent. Thank you very much.

Dr. I. C. Gardner, National Bureau of Standards, Washington, D. C.: Dr. O'Brien, I would like to say that the success of this symposium has depended in the main on the fact that fortunately a very timely subject was chosen and those who participated have cooperated most wonderfully well.

Reprinted from JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, Vol. 44, No. 8, 610–615, August, 1954
Printed in U. S. A.

Light Distribution in the Image of an Incoherently Illuminated Edge*

W. WEINSTEIN†
Technical Optics Section, Imperial College of Science and Technology, London, S. W. 7, England
(Received May 4, 1954)

The image of an incoherently illuminated edge is shown to have a simple structure with no diffraction fringes for any state of correction of the imaging system. The light distribution in the defocused image is calculated and numerical results are given. A comparison is made with the predictions of geometrical optics.

I. INTRODUCTION

A CONSIDERABLE amount of experimental and theoretical work has been carried out recently on the diffraction images of point sources formed by optical systems with varying amounts of aberration; a bibliography is given by Wolf.[1] From this work it appears that the main effect of aberrations less than a quarter of a wavelength is to decrease the flux density in the center of the diffraction pattern; this drop in central flux density forms the basis of a useful measure of the effect of small aberrations on the imaging quality, the "definitions-helligkeit" of Strehl. In many optical systems, among the most important being photographic objectives, the aberrations are much greater than $\lambda/4$ and when this limit is passed the diffraction image of a point becomes increasingly complex, with very beautiful fringe structures, as may be seen from the samples reproduced in Wolf's review. A consequence of this complexity is that point images are quite useless as a basis for the study of the influence of moderate or large amounts of aberration on imaging quality, since there is no region in the image where the flux density decreases more or less regularly with increasing aberration. Very little trace of this fine structure can be discerned in the normal use of such optical systems; it therefore appears reasonable to study the images produced, not of point sources, but of objects such that increase in aberration in the optical system causes general deterioration of image sharpness and contrast, rather than complicated fringe formation. Such an object is the incoherent half-plane, i.e., an incoherently illuminated plane divided by a straight line into two regions of different uniform flux density; usually one region would be of zero flux density but this is not essential.

It will be shown that under the usual assumptions of scalar diffraction theory[2] the gradient of the flux density in the image of an incoherent half-plane has the same sign over the whole image plane. This theorem is true whatever the aberrations, pupil shape or variation of amplitude over the pupil, so that the image has no fringe structure and preserves a relatively simple character. Furthermore, it may be expected that the effect of increasing aberration will be chiefly to cause a greater spreading in the image. Thus the incoherent edge, as it may be called, is a suitable test object to form the basis of a theoretical study of the influence of large aberrations on image quality,[3] since the condition of incoherent illumination is appropriate to most of the applications of photographic objectives.

II. GENERAL FEATURES OF THE EDGE IMAGE

Let (ξ_0,η_0) be rectangular coordinates in the object plane. The flux density $I_0(\xi_0,\eta_0)$ in the object can without loss of generality be taken as

$$I_0=\begin{cases}1, & (\eta_0\leqslant 0)\\ 0, & (\eta_0>0).\end{cases} \tag{1}$$

Let (ξ,η) be rectangular coordinates in any chosen image plane. The origin will be assumed to lie on the principal ray from the origin of the (ξ_0,η_0) system, and the ξ and η axes will be parallel to the images of the ξ_0 and η_0 axes. Let $A(\xi,\eta)$ be the flux density distribution in the image plane due to a luminous point at the origin in the object plane. Then the flux density distribution $I(\eta)$ in the edge image is obtained by integrating the effect of a distribution of such point sources over the half-plane $\eta_0\leqslant 0$. It will be assumed that the aberrations of the pencils from all points of the field making appreciable contributions to the flux density in the region of the edge image are the same. This is a reasonable assumption since only points very near the edge make such contributions, and the relevant variations of the aberrations with field angle for a system of moderate numerical aperture are very small, even when the system is not isoplanatic in the usual sense. It then follows that

$$I(\eta')=\int_{\eta'}^{\infty}\left\{\int_{-\infty}^{\infty}A(\xi,\eta)d\xi\right\}d\eta, \tag{2}$$

neglecting a scale factor. This integral may be interpreted in the usual way as the volume of the diffraction solid contained in the space $\eta\geqslant\eta'$.

Since $A(\xi,\eta)$ is a non-negative function, being a flux density distribution, $I(\eta')$ must be nonincreasing with

* Invited paper given before the Optical Society of America, Spring meeting, 1954.

† Guest worker, Optics and Metrology Division, National Bureau of Standards, Washington, D. C., from January to July 1954.

[1] E. Wolf, Repts. Progr. Phys. **14**, 95 (1951).

[2] Theimer, Wassermann, and Wolf, Proc. Roy. Soc. (London) **A212**, 426 (1952).

[3] W. Weinstein, Phot. J. **91B**, 138 (1951).

ncreasing η', that is, the gradient of I is always negative or zero. This proves the theorem stated in Sec. I.

In most cases the gradient will not be zero for any value of η' since this would require that $A(\xi,\eta')$ should vanish for all values of ξ; this could only occur in relatively few cases, of which perhaps the most likely is that of astigmatic imagery with a rectangular pupil, where it is known that $A(\xi,\eta)$ can be expressed in the form $F(\xi)\cdot G(\eta)$. In the case of photographic objectives, for which the edge test object is particularly suited, it seems very unlikely that this situation could occur. Thus, it may be expected that the flux density distribution with any aberrations would be strictly monotonic and that its general features could be summarized by some such quantity as the maximum gradient or the distance within which a specified fraction of the total flux density change across the edge takes place. It was decided to investigate the light distribution in the edge image in more detail in order to see how nearly this expectation was fulfilled.

III. THE INTEGRAL FOR THE FLUX DENSITY DISTRIBUTION

An expression for $I(\xi,\eta)$ can be obtained from the double integral in Eq. (2) by substituting the appropriate expression for A. However, manageable explicit expressions for A have not been obtained for aberrations greater than about $\frac{1}{2}\lambda$ with the exception of pure defocusing and even in this case the required double integral seems intractable. An indirect method was therefore adopted.

Take rectangular coordinates (x,y) in the exit pupil parallel to the (ξ,η) axes. Let $\phi(x,y)$ be the aberration of the wave front from the (ξ_0,η_0) origin in the object plane, in radians, where it is met by the ray passing through the point (x,y). The aberration is taken with respect to a reference sphere centered on the (ξ,η) origin, so that ϕ is $2\pi/\lambda$ times the optical path between this sphere and the wave front. Let $G(\xi,\eta)$ be the complex amplitude in the image produced by this wave front.

Duffieux and Lansraux[4] showed by an application of the Parseval formula for Fourier transforms[5] that

$$\int_{-\infty}^{\infty}|G(\xi,0)|^2d\xi=\int\left|\int\exp i\phi(x,y)dy\right|^2dx, \qquad (3)$$

neglecting a factor depending on the scales in the two coordinate systems. The integrals on the right-hand side are taken over ranges defined by the pupil area. Alternatively they may be taken between infinite limits if the integrand is defined to be zero outside the pupil. By definition G is the amplitude corresponding to a finite energy distribution. Thus G is square integrable and the application of the Parseval formula is valid.

[4] P. M. Duffieux and G. Lansraux, Rev. opt. 24, 76 (1945).
[5] E. C. Titchmarsh, *Theory of Fourier Integrals* (Oxford University Press, London, 1937).

The integrand on the left-hand side of Eq. (3) is equal to $A(\xi,0)$, so that the integral is the area of the section of the diffraction solid by the plane $\eta=0$. Thus, from Eq. (2) it gives the negative of the gradient of flux density in the edge image at $\eta=0$ to a certain scale, or, to another scale, the central flux density in the image of an incoherently illuminated line.

Next let the integrand with respect to y on the right-hand side of Eq. (3) be replaced by $\exp i[\phi(x,y)+vy]$, where v is a variable parameter. The integral then represents the central gradient in the image when the aberration is $\phi(x,y)+vy$ instead of simply $\phi(x,y)$; but these two aberration functions are essentially the same, the former being taken with respect to a reference sphere of which the center is displaced laterally a distance $\eta=f\lambda v/2\pi$, where f is the radius of the reference sphere, i.e., the distance from the exit pupil to the image plane. Thus the new integral represents the gradient in the flux density distribution of the edge image at the point η, or the flux density of the line image at η,

$$-\frac{dI(\eta)}{d\eta}=\int\left|\int\exp i\left(\phi(x,y)+\frac{2\pi}{f\lambda}\cdot\eta y\right)dy\right|^2dx. \qquad (4)$$

Integrating with respect to η, the flux density in the edge image is then [see Eq. (2)]

$$I(\eta')=\int_{\eta'}d\eta\int\left|\int\exp i\left(\phi(x,y)+\frac{2\pi}{f\lambda}\cdot\eta y\right)dy\right|^2dx. \qquad (5)$$

The theorem of Sec. I also follows from the fact that the integrand with respect to η is non-negative.

The flux density distributions with perfect in-focus imagery for the incoherent edge and line were obtained by Struve[6] by direct application of Eq. (2); in this paper the case of aberrationless but defocused imagery is treated by evaluating the integral in Eq. (5).

V. FLUX DENSITY DISTRIBUTION IN THE DEFOCUSED EDGE IMAGE

For simple defocusing the aberration function $\phi(x,y)$ takes the form $p(x^2+y^2)$ where the parameter p is related to the geometrical displacement ξ of the image plane by $\lambda f^2p=\pi\xi$. If the exit pupil is circular and of unit radius the integral (5) may be written

$$I(p,v)=A\int_0^v dv \times\int_{-1}^{1}\left|\int_{-a}^{a}\exp i\{p(x^2+y^2)+vy\}dy\right|^2dx+B, \qquad (6)$$

where $a=+(1-x^2)^{\frac{1}{2}}$ and the parameters p and v are used as coordinates instead of ξ and η; the constants A and B will be used to normalize the result so that the total flux density step is unity and I tends to zero as

[6] H. Struve, Wiedemann's Ann. 17, 1008 (1882).

v tends to infinity. This integral is conveniently evaluated in different ways according to the ranges of p and v in question.

VI. POWER-SERIES EXPANSION

For small values of p and v a simple power-series solution can be obtained. The factor $\exp ipx^2$ in the integrand in Eq. (6) can be ignored since it vanishes in taking the modulus. The squared modulus can then be written as a repeated integral

$$\left|\int_{-a}^{a}\exp i\{py^2+vy\}dy\right|^2 = \int_{-a}^{a}\int_{-a}^{a}\exp i\{p(y^2-y'^2)+\xi(y-y')\}dydy'. \quad (7)$$

On making the change of variables $t=(y-y')/2a$, $z=y+y'$ and carrying out the integration with respect to z the result is

$$2\int_0^1 \frac{\cos(2avt)\sin[4a^2pt(1-t)]}{pt}dt$$

plus an imaginary term which must vanish identically since the result is to be a squared modulus. Substituting in Eq. (6),

$$I(p,v)=2A\int_0^v dv\int_1^1 dx \times\int_0^1 \frac{\cos(2avt)\sin[4a^2pt(1-t)]}{pt}dt+B. \quad (8)$$

Integrating with respect to v and expanding the second sine in the integrand as a power series,

$$I(p,v)=A\sum_{m=0}^{\infty}\frac{(-)^m G_{2m}(v)}{(2m+1)!}\cdot p^{2m}+B, \quad (9)$$

where

$$G_m(v)=2^{2m+2}\int_{-1}^{1}dx \times\int_0^1 a^{2m+1}t^{m-1}(1-t)^{m+1}\sin(2avt)dt. \quad (10)$$

Expanding the sine, integrating term by term and using the factorial function duplication formula,

$$G_m(v)=2\pi(m+1)!\times\sum_{n=0}^{\infty}\frac{(-)^n(2n+m)!v^{2n+1}}{(2n+1)!(n+m+\frac{1}{2})!(n+m+\frac{3}{2})!}. \quad (11)$$

The constants A and B can now be chosen, since the result must reduce to Struve's expression for the in-focus flux density distribution when $p=0$; it is found that $A=-1/(2\pi^2)$, $B=\frac{1}{2}$, giving for the normalized flux density distribution

$$I(p,v)=\frac{1}{2}-\frac{1}{\pi}\sum_{m=0}^{\infty}(-)^m p^{2m} \times\left\{\sum_{n=0}^{\infty}\frac{(-)^n(2n+2m)!v^{2n+1}}{(2n+1)!(n+2m+\frac{1}{2})!(n+2m+\frac{3}{2})!}\right\}. \quad (12)$$

In the above derivation the series expansions can be dominated by sine-type power series, so that the term-by-term integrations are justified by uniform convergence.

The functions $G_m(v)$ of Eq. (11) can be expressed in terms of hypergeometric functions of ${}_2F_3$ type[7] and they can be shown by a number of differentiations and eliminations to satisfy the following recurrence relations:

$$G_{m+2}(v)=\frac{m+3}{(m+2)(m+4)v^2}\{4(m+1)^2(m+2)G_{m+1}(v) -(5m+8)vG_{m+1}'(v)-4m(m+1)(m+2)G_m(v)\}, \quad (13)$$

$$G_{m+2}'(v)=\frac{4(m+3)}{(m+4)v^3}\{3(m+1)(m+2)G_{m+1}(v) +(m^2+7m+9)vG_{m+1}'(v)-3m(m+2)G_m(v) -(m+2)(m+4)vG_m'(v)\}. \quad (14)$$

These relations are useful for computing numerical values when $v>m$.

Equation (12) is only suitable for computation with moderate values of the variables, although the series are convergent everywhere. In the next section an expansion suitable for large values of v is obtained.

VII. EXPANSION IN DESCENDING POWERS OF v

When v is large compared with p an expansion in descending powers of v can be obtained in the following way.

Let

$$f(x)=\int e^{ivy}e^{i\phi(x,y)}dy, \quad (15)$$

so that from Eq. (5), using the parameter v instead of η,

$$I(v)=\int^v dv\int|f(x)|^2dx; \quad (16)$$

the ranges of integration with respect to x and y are again determined by the pupil edge. Put $\exp i\phi(x,y)=U$

[7] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, 1944).

and integrate Eq. (15) by parts, giving

$$f(x)=\left[e^{iv y}\left\{-\frac{i}{v}U+\frac{1}{v^2}U'-\frac{i}{v^3}U''+\cdots+\frac{(i)^{n+2}}{v^n}U^{(n-1)}\right\}\right]-\frac{(i)^{n+2}}{v^n}\int e^{iv y}U^{(n)}dy. \quad (17)$$

If after taking the squared modulus the integrations with respect to x and v van be carried out, this will provide an expansion in descending powers of v; whether it will converge or be asymptotic depends on the form of $\phi(x,y)$. For the case of simple defocusing, where as in the foregoing $\phi=p(x^2+y^2)$, it is found that

$$|f(x)|^2=\frac{2}{v^2}+\frac{24a^2p^2}{v^4}-\frac{40p^2}{v^6}+\frac{160a^4p^4}{v^6}+\frac{32ap^2}{v^5}\sin 2av+\left\{-\frac{2}{v^2}-\frac{8a^2p^2}{v^4}+\frac{40p^2}{v^6}-\frac{32a^4p^4}{v^6}\right\}\times\cos 2av+O\left(\frac{1}{v^7}\right). \quad (18)$$

This can be integrated with respect to x by means of functions introduced by Lommel;[7] these are expressed

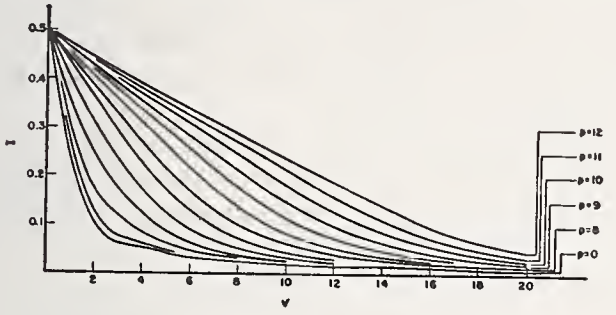


FIG. 1. Flux density in the image of an incoherently illuminated edge.

in descending powers of v and the result is integrated with respect to v from v to ∞. The first few terms of the expansion for the flux density distribution are as follows:

$$I(p,v)=\frac{1}{\pi^2}\left\{\frac{2}{v}+\frac{6}{v^3}-\frac{3}{40v^5}+\frac{16p^2}{3v^3}-\frac{8p^2}{v^5}+\frac{256p^4}{15v^5}+\left(\frac{\pi}{v}\right)^{\frac{1}{2}}\sin\left(2v-\frac{\pi}{4}\right)\left(\frac{1}{v^2}-\frac{937}{512v^4}+\frac{2p^2}{v^4}\right)+\left(\frac{\pi}{v}\right)^{\frac{1}{2}}\cos\left(2v-\frac{\pi}{4}\right)\times\left(-\frac{17}{16v^3}+\frac{16813}{4096v^5}+\frac{39p^2}{8v^5}\right)\right\}+O\left(\frac{1}{v^6}\right). \quad (19)$$

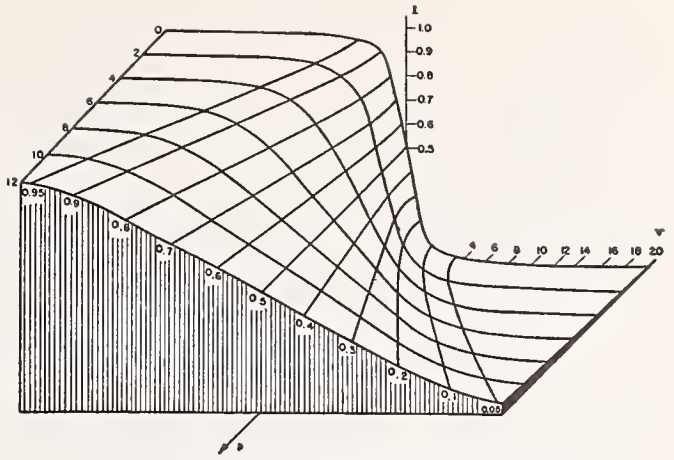


FIG. 2. Relief of flux density in edge image.

It is shown in the appendix to this paper that this is an asymptotic expansion.

VIII. THE CENTRAL FLUX DENSITY GRADIENT IN THE EDGE IMAGE

The flux density gradient $S(p)$ in the edge image at the center of the geometrical image can be found by differentiating the expression (12) for $I(p,v)$ with respect to v and setting $v=0$; the result is

$$S(p)=-\frac{1}{\pi}\sum_{m=0}^{\infty}\cdot\frac{(-)^m(2m)!p^{2m}}{(2m+\frac{1}{2})!(2m+\frac{3}{2})!}. \quad (20)$$

A formula equivalent to Eq. (20) was given by Steel[8] as the central flux density in a line image.

Equation (20), although convergent for all values of p, is only suitable for computation with moderately low values. An alternative expression, which is suitable for all values of p, can be obtained by direct integration across the diameter of the diffraction solid. Starting from the series of Nijboer[9] for the complex amplitude in the out-of-focus Airy disk, the squared modulus is taken, giving for the flux density distribution

$$A(p,v)=\frac{\pi}{v^2p}\sum_{\lambda=0}^{\infty}\sum_{\mu=0}^{2\lambda}(-)^{\lambda+\mu}(4\lambda-2\mu+1)(2\mu+1)\times J_{2\lambda-\mu+\frac{1}{2}}(\tfrac{1}{2}p)J_{\mu+\frac{1}{2}}(\tfrac{1}{2}p)J_{4\lambda-2\mu+1}(v)J_{2\mu+1}(v). \quad (21)$$

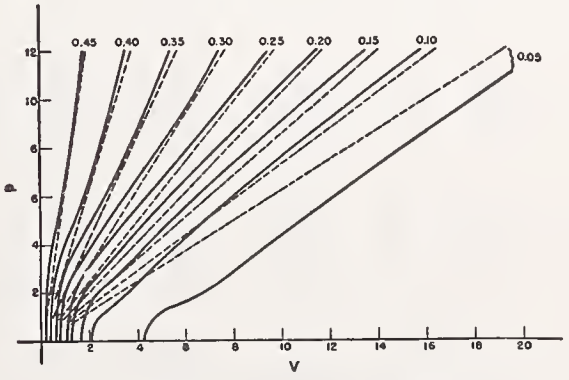


FIG. 3. Isophots in edge image.

[8] W. H. Steel, Rev. opt. 31, 334 (1952).
[9] B. R. A. Nijboer, Physica 13, 605 (1947).

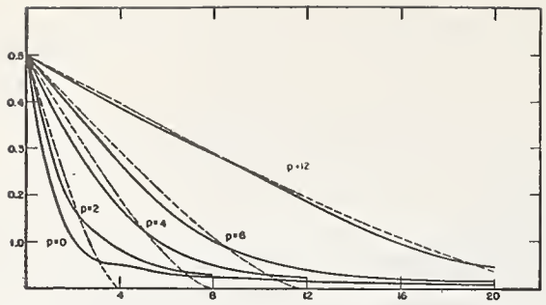


FIG. 4. Comparison of predictions of geometrical optics and diffraction theory for edge image.

This must be integrated from 0 to ∞ with respect to v to give the central gradient. On making use of the Bessel function recurrence formulas the integration can be carried out by means of an indefinite integral due to Lommel;[7] the result normalized to correspond to unit flux density step at the edge is

$$S(p)=\frac{8}{\pi p}\sum_{\lambda=0}^{\infty}\frac{(-)^{\lambda}}{(4\lambda+1)(4\lambda+3)}\times\left\{\sum_{\mu=0}^{\lambda}\frac{(-)^{\mu}\epsilon_{\lambda-\mu}(4\lambda-2\mu+1)(2\mu+1)}{(4\lambda-4\mu-1)(4\lambda-4\mu+1)}\times J_{2\lambda-\mu+\frac{1}{2}}(\tfrac{1}{2}p)J_{\mu+\frac{1}{2}}(\tfrac{1}{2}p)\right\},\quad(22)$$

where ϵ_n is Neumann's factor, defined as equal to 1 if n is zero or 2 if n is not zero. This expression is suitable for computation with large values of p.

IX. NUMERICAL RESULTS

The results of numerical computations using the expressions obtained in Secs. 6 and 7 are shown in Figs. 1 to 4 and are given in Table I. Figure 1 shows the flux density distribution for $v\gtrless 0$ for a number of different values of p, the defocusing parameter; the values for $v<0$ are, of course given by $I(p,-v)=1-I(p,v)$ and for $p<0$ by $I(-p,v)=I(p,v)$. The largest value of p taken corresponds to 1.9 wavelengths defocusing. Figure 2 shows a perspective relief of the I, p, v relationship and Fig. 3 shows the isophots, or lines of constant flux density, in the p, v plane. It can be seen that the $I-v$ curves are, as expected, very smooth and even the faint indication of an inflection which appears in the in-focus image at about $v=4$ cannot be seen in the out-of-focus images.

The flux density distribution in the defocused edge image on the basis of geometrical optics, i.e., neglecting diffraction effects, can be shown to be

$$\bar{I}(p,v)=\frac{1}{2}-\frac{1}{\pi}\left\{\arcsin\left(\frac{v}{2p}\right)+\frac{v}{2p}\left(1-\frac{v^2}{4p^2}\right)^{\frac{1}{2}}\right\},\quad(-2p\leqslant v\leqslant 2p)\quad(23)$$

$$\bar{I}(p,v)=0,\quad(v<-2p,\ v>2p).$$

For comparison the values obtained from this expression are plotted in Figs. 3 and 4 with broken lines. It can be seen that for values of p greater than about 2 (defocusing $>\lambda/3$) the agreement between the two sets of values is fairly good until the region of the geometrical shadow ($v=2p$) is reached; beyond this region, however, there is a considerable "tail" of light in the true image which remains unaccounted for by geometrical optics.

The central flux density gradient computated from the expressions in Sec. 8 is plotted with reversed sign in Fig. 5; for comparison the gradient on the basis of

TABLE I. Flux density $I(p,v)$ in the defocused image of an incoherently illuminated edge.

v \ p	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
1	0.252	0.266	0.304	0.352	0.398	0.431	0.447	0.452	0.456	0.459	0.464	0.470	0.474
2	0.107	0.127	0.178	0.246	0.312	0.362	0.389	0.410	0.417	0.424	0.430	0.439	0.446
3	0.063	0.078	0.119	0.178	0.238	0.291	0.329	0.359	0.378	0.391	0.400	0.409	0.417
4	0.053	0.061	0.085	0.124	0.171	0.222	0.266	0.310	0.338	0.357	0.370	0.379	0.388
5	0.041	0.045	0.059	0.084	0.121	0.172	0.211	0.262	0.298	0.323	0.339	0.351	0.361
6	0.033	0.035	0.044	0.061	0.088	0.130	0.169	0.217	0.257	0.285	0.309	0.323	0.337
7	0.029	0.031	0.037	0.048	0.068	0.096	0.132	0.173	0.213	0.248	0.276	0.299	0.313
8	0.026	0.027	0.030	0.037	0.052	0.071	0.101	0.136	0.175	0.212	0.246	0.270	0.287
9	0.022	0.023	0.025	0.031	0.040	0.055	0.078	0.107	0.142	0.178	0.214	0.242	0.263
10	0.020	0.021	0.023	0.026	0.033	0.044	0.061	0.083	0.114	0.147	0.181	0.211	0.238
11	0.019	0.019	0.020	0.023	0.028	0.036	0.048	0.066	0.090	0.119	0.151	0.183	0.212
12	0.017	0.017	0.018	0.020	0.023	0.030	0.039	0.052	0.072	0.097	0.126	0.157	0.186
13	0.016	0.016	0.017	0.018	0.020	0.026	0.033	0.044	0.059	0.080	0.105	0.134	0.164
14	0.015	0.015	0.015	0.017	0.018	0.022	0.027	0.035	0.048	0.064	0.086	0.111	0.141
15	0.014	0.014	0.014	0.015	0.016	0.019	0.023	0.029	0.038	0.051	0.068	0.090	0.116
16	0.013	0.013	0.013	0.014	0.015	0.017	0.020	0.025	0.032	0.042	0.056	0.074	0.095
17	0.012	0.012	0.012	0.013	0.014	0.016	0.018	0.022	0.027	0.035	0.046	0.061	0.080
18	0.011	0.011	0.012	0.012	0.013	0.014	0.016	0.019	0.023	0.029	0.038	0.050	0.066
19	0.011	0.011	0.011	0.011	0.012	0.013	0.015	0.017	0.020	0.025	0.032	0.042	0.055
20	0.010	0.010	0.011	0.011	0.011	0.012	0.014	0.015	0.018	0.022	0.027	0.035	0.046

geometrical optics, which from Eq. (23) is given by $S(p)=-1/(\pi p)$, is also shown with broken line. It can be seen that the gradient decreases more or less steadily, so that for the case of simple defocusing the gradient is a measure of image quality which fulfills the requirements outlined in Sec. 1.

APPENDIX: PROOF THAT EQUATION (19) IS AN ASYMPTOTIC EXPANSION

The expression for $G_{2m}'(v)$ which can be obtained by differentiating Eq. (10) can be integrated by parts with respect to t; on making the change of variable $s=1-2t$ and using Poisson's integral for the Bessel function it follows after some reduction that

$$G_{2m}'(v)=\frac{2^{2m+\frac{3}{2}}(2m)!}{v^{2m+\frac{1}{2}}}\int_0^1 \times a^{2m+\frac{3}{2}}\{\cos avJ_{2m+\frac{1}{2}}(av)+\sin avJ_{2m+\frac{3}{2}}(av)\}dx.$$

Thus

$$G_{2m}'(v)=\frac{\pi^{\frac{1}{2}}2^{2m+5/2}(2m)!}{v^{2m+\frac{1}{2}}}g_{2m}(v), \quad (|g_{2m}(v)|<1,\ v>1),$$

and from Eq. (9), ignoring a constant factor,

$$\frac{\partial}{\partial v}I(p,v)=\sum_{m=0}^{\infty}\frac{(-)^m}{(2m+1)}\left(\frac{2p}{v}\right)^{2m}g_{2m}(v).$$

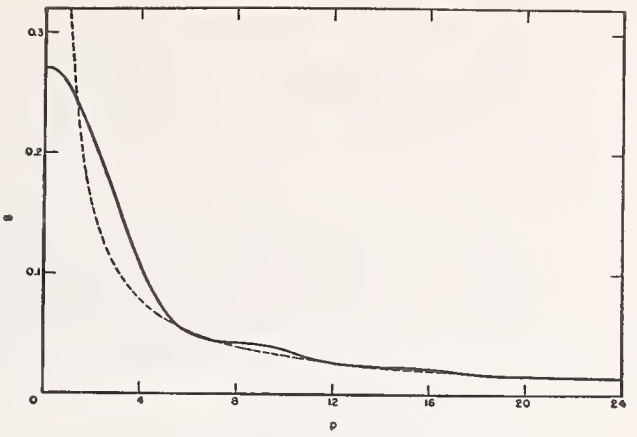


FIG. 5. Central flux density gradient in edge image.

If $v>2p$, this series converges more rapidly than the geometric series $\sum(2p/v)^{2m}$; in order to evaluate the expressions $g_{2m}(v)$, the Bessel functions in the integrand are expanded as finite descending series in v, and the result is equivalent to the right-hand side of Eq. (18), which is therefore a convergent series for $v>2p$. The subsequent operations in which asymptotic expansions of the Lommel functions are introduced are all permissible, so that Eq. (19) is a true asymptotic expansion for $v>2p$.

ACKNOWLEDGMENTS

Thanks are due Miss E. M. Cook, Miss J. M. Drewitt. and Mr. Loyd Sutton, for help with the computations,

Journal of Research of the National Bureau of Standards Vol. 54, No. 3, March 1955 Research Paper 2574

Variation in Distortion with Magnification

Arthur A. Magill

The distortion introduced in the image by a lens for a given axial inclination of the chief ray is a linear function of the magnification. Specifically, if D_0 represents the distortion with parallel light incident on one side of the lens (zero magnification) and D_∞ the distortion with parallel light incident on the other, then the distortion D_m at any magnification m is given by $D_m = D_0 - {}^m D_\infty$. This equation has been experimentally verified for examples of three types of symmetrical lenses.

1. Introduction

The growth of photogrammetry with its high-precision imaging systems has necessitated accurate determination of the distortion introduced by a lens. Because the distortion varies with the magnification, it has been customary heretofore to have the distortion measured for each object-to-image ratio employed. To avoid making such a series of measurements, a simple linear equation for a lens with a single effective stop has been developed, from which the distortion at any magnification may be computed from the measured values obtained at two magnifications. The two distortions that would normally be used are those obtained with parallel light incident, in turn, on the front and on the back of the lens. These not only require the simplest experimental setup, but also have special significance as the limits of real image formation. Virtual images (i. e., negative magnifications) will be excluded as having little practical value.

2. Theoretical Development

In 1907 E. Wandersleb [1] developed a hyperbolic relation between magnification and the ratio of the lateral displacement of the image from its distortionless position to the height of the distortionless image. By defining the distortion as the later displacement of the image, the relationship may be reduced to a linear function, fully expressible by the introduction of the two values obtained with parallel light.

In figure 1, a lens of focal length f with surface 1 and final surface k has the paraxial foci F and F'. It is equipped with a single effective stop. The paraxial pupils are located at E_0 and E'_0 at distances p_0 and p'_0 from the foci. The spherical aberration of the lens elements preceding and following the stop changes the positions of the pupils for nonparaxial rays. For the chief ray with a slope angle of u to the optic axis in the object space and u' in the image space, the pupils are shifted by amounts δ and δ' to E and E' at distances p and p' from the foci.

Now consider the object point O with this same chief ray located at a height y from the optic axis, as shown in figure 2. The sign convention is that of taking figure 2 as an all-positive diagram. The intersection of the emergent chief ray with the Gaussian image plane at O′, a height y' from the axis, will be taken as the image of O. For low-aperture lenses with negligible zonal aberrations, this is certainly a legitimate definition of the image position. Because most precision projection is done

[1] E. Wandersleb, Uber die Verzeichnungsfehler Photographischer Objectiv, Z. Instrumentenk. **27**, 33-37, 75-85 (1907). The equation is also used in many optical texts.

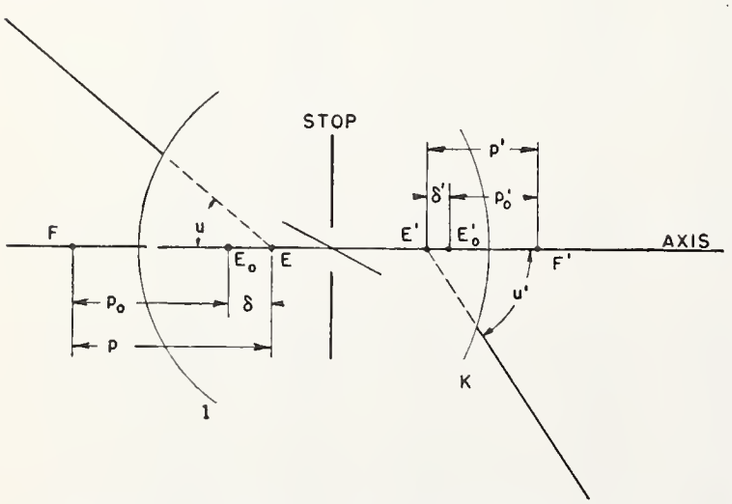


FIGURE 1. *Diagram of a lens and a given chief ray.*

The lens elements preceding and following the stop image is at E_0 and E'_0, respectively. For the indicated nonparaxial chief ray the spherical aberration of the elements relocates these images at E and E'.

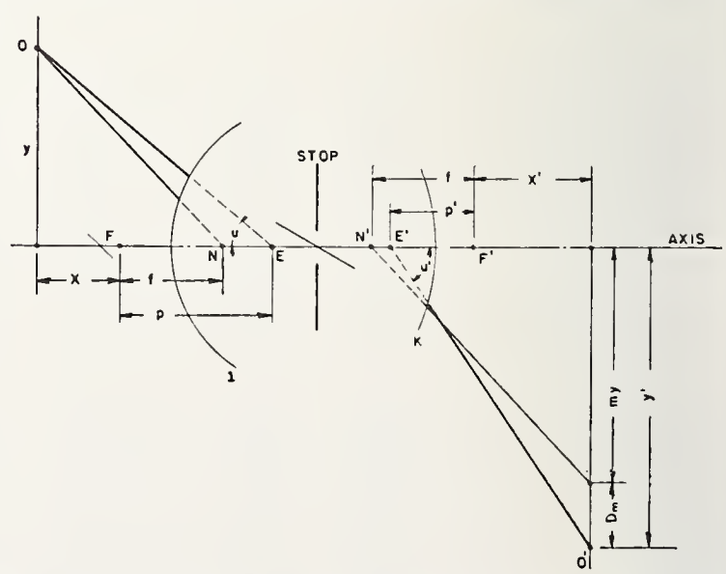


FIGURE 2. *Same lens as figure 1 with object and image planes included.*

The object O is imaged at O′ in the Gaussian image plane. The distortion D_m is the lateral displacement of O′ from its Gaussian position, $D_m = y' - my$. All distances and angles are to be considered positive, as shown.

at f-numbers of 22 or larger, the definition is also in accord with photogrammetric practice. Thus, the magnification m, can be defined in terms of Gaussian optics as

$$m=\lim_{y\to 0}\frac{y'}{y}=\frac{x'}{f}=\frac{f}{x}, \tag{1}$$

where x and x' are the distances of the object and image planes from their respective foci. A ray in the object space directed toward the first nodal point N, will emerge at the same slope angle in the image space as though emanating from the second nodal point, N'. This ray intersects the Gaussian image plane at the distortionless image position, a height my from the optic axis.

The linear distortion for any magnification D_m is defined as the difference between the actual image position and this distortionless image position, that is,

$$D_m=y'-my. \tag{2}$$

From figure 2 it is apparent that the values of y and y' referred to the chief ray are

$$y=(x+p)\tan u, \tag{3}$$

$$y'=(x'+p')\tan u'. \tag{4}$$

By substituting the values of x and x' from eq (1) and multiplying y through by m, the following equations are obtained:

$$my=(f+mp)\tan u, \tag{5}$$

$$y'=(mf+p')\tan u'. \tag{6}$$

Substitution of eq (5) and (6) in eq (2) yields

$$D_m=(p'\tan u'-f\tan u)-(p\tan u-f\tan u')m. \tag{7}$$

For any given chief ray, all values in eq (7) are constants with the exception of the magnification, and D_m is a linear function of m.

The magnification for an object located an infinite distance to the left is zero and eq (7) reduces to

$$D_0=p'\tan u'-f\tan u. \tag{8}$$

This is geometrically illustrated in figure 3 where parallel light incident on the left at the slope angle u forms an image in the paraxial focal plane on the right at a height of $p'\tan u'$. The distortionless image position is given as the limiting value of my as m approaches zero and y approaches infinity, which from eq (5) is $f\tan u$.

For an object in the focal plane at the left, the magnification and the linear distortion are both infinite. However, if an object point is considered to be located at the position of the infinite image, it would in turn be imaged at the intersection of the chief ray with the focal plane on the left. The linear distortion D_∞, is here defined as that obtained

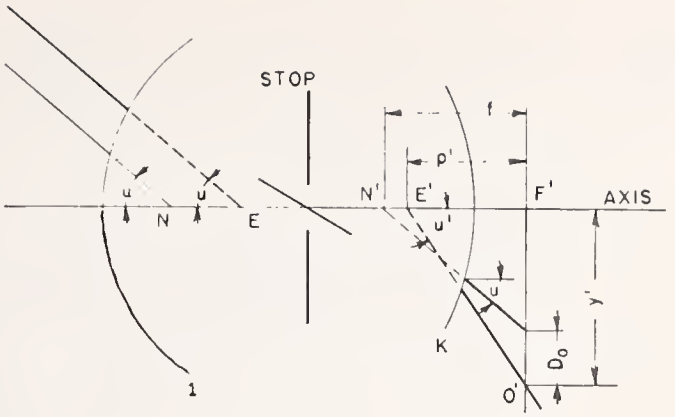


FIGURE 3. *Same lens as figure 1 at a magnification of zero.*
An infinite point object on the left is imaged in the focal plane at the right at O'. For this magnification of zero the distortion is given by $D_0=y'-f\tan u$.

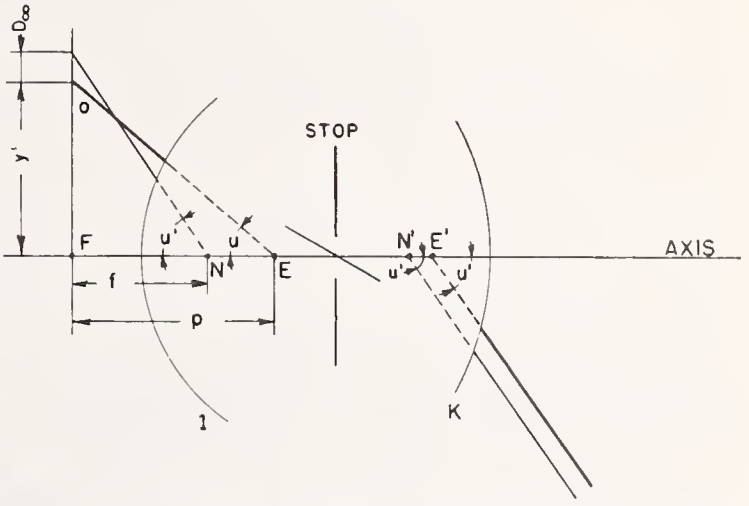


FIGURE 4. *Same lens as figure 1 at a magnification of infinity.*
An object O in the focal plane at the left is imaged at infinity on the right. If the infinite image were replaced by an object, it would in turn be imaged at O The resulting distortion is designated D_∞ and in this case is negative.

when parallel light is incident on the right at a slope angle of u'. Referring to figure 4, it is clear that

$$D_\infty=p\tan u-f\tan u'. \tag{9}$$

By dividing eq (7) by m, it is also true that

$$D_\infty=-\lim_{m\to\infty}\frac{D_m}{m}=p\tan u-f\tan u', \tag{10}$$

where the negative sign indicates that D_∞ is measured with the light going backward through the lens. From eq (7), (8), and (9),

$$D_m=D_0-mD_\infty. \tag{11}$$

This equation gives the linear distortion at any magnification in terms of the two distortions obtained with parallel light for any given chief ray. The only restrictions are that there be a single effective stop and a requirement that the chief ray adequately determine the center of gravity of the image or that, if it does not, the discrepancy be a linear function of the magnification.

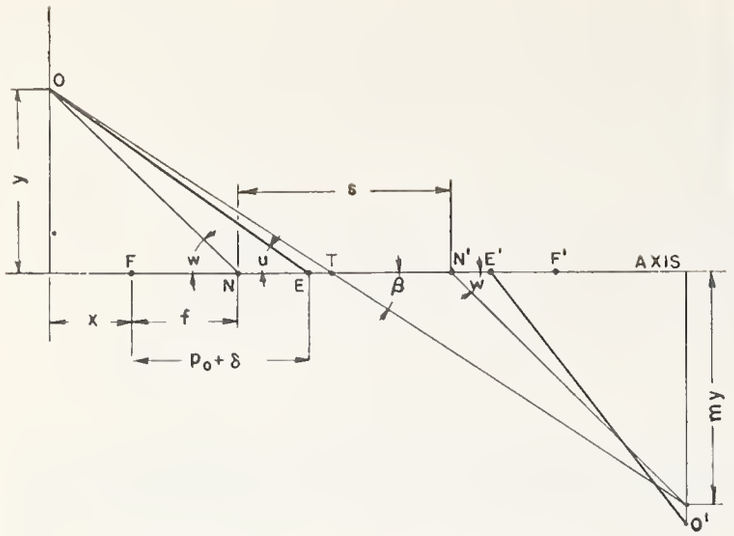


FIGURE 5. *Same lens as figure 1 illustrating other reference angles.*

It is often more convenient to refer the distortion to the slope angles w or β than it is to the slope angle, u. For a symmetrical lens, u, w, and β may be used interchangeably with little error.

3. Experimental Method

The equation has been verified for three types of commonly used symmetrical objectives at relative apertures of $f/22$. The experimental method is the same as that described by Bennett,[2] using a nodal slide and optical bench. The probable error of measurement varies with the slope angle, ranging from approximately 1 to 5 μ at 45° with parallel incident light. For finite object distances the probable error is greater by the nominal factor of $(m+1)$.

For the measurement of D_∞, the value of u' that corresponds to any given u may be found from eq (8), providing the distance $p'=(p_0'+\delta')$ is known. Furthermore, for the user of the lens to be able to refer distortion to the chief ray requires that he know either $(p_0+\delta)$ or $(p_0'+\delta')$. The measurement of the location of the paraxial pupils presents no particular problem. To determine the values of δ and δ', the stop could be equipped with a cross hair and the displacement of its image measured as the lens is rotated around each paraxial pupil in turn. The magnitude of δ could also be computed from eq (9) as a check on the measured values.

In practice it is often more convenient to refer the distortion to some other basis than the slope angle of the chief ray. In the nodal slide method the distortion is referred to the slope angle of the straight line joining the object and its Gaussian image, indicated as β in figure 5. The slope angle, w, of the ray in the object space directed toward the first nodal point might serve as a more convenient basis for the user of the lens, who normally computes the conjugate distances from the nodal points. The equations relating these two slope angles to that of the chief ray are

$$\tan w=\frac{p_0+\delta+x}{f+x}\tan u, \tag{12}$$

$$\tan \beta=\frac{p_0+\delta+x}{f+x+\dfrac{s}{m+1}}\tan u, \tag{13}$$

where s is the separation of the nodal points.

For the ideal symmetrical lens, the pupils coincide with the corresponding nodal points and $p_0=p_0'=f$, $u=u'$, and $\delta=\delta'$. Also from eq (8) and (9), $D_0=D_\infty=\delta\tan u$. Remembering that $x=f/m$, eq (12) and (13) may be rewritten

$$\tan w=\tan u+\frac{D_0}{f}\left(\frac{m}{m+1}\right), \tag{14}$$

$$\tan \beta=\frac{1}{1+\dfrac{s}{f}\dfrac{m}{(m+1)^2}}\tan w. \tag{15}$$

That u, w, and β may be used interchangeably with little error is superficially apparent because D_0 is normally a very small percentage of f and the discrepancy between β and w maximizes for $m=1$, for which magnification the distortion is zero. Although nominally symmetrical lenses are actually slightly asymmetric resulting in obvious differences between D_0 and D_∞, the effect on the reference slope angle has been ignored in obtaining the data for this paper; that is, the assumption has been made that u, u', and β are all equal. A detailed defense of this assumption for each lens tested would appear to be trivial in view of the good experimental agreement that was obtained. It should be pointed out, however, that if u' is not equal to u, the assumption of equality will in itself tend to minimize the errors introduced by employing a constant β for finite object distances.

4. Experimental Verification

The experimentally determined distortions are given for examples of the three types of symmetrical lenses in tables 1, 2, and 3. In all cases the value of the distortion at a slope angle of 5° is assumed to be zero. The errors in magnification are so small with respect to the other errors involved that all magnifications are considered exact. The measurements were made with a tungsten light source and a narrow-band filter having a dominant wavelength of 575 mμ.

The italicized digits represent the differences between the computed and observed values, and give a measure of the error involved in predicting distortion from D_0 and D_∞. In general these differences are what would be expected from the probable error of measurement. The greatest discrep-

[2] A. H. Bennett, The distortion of some typical photographic objectives, J. Opt. Soc. Am. and Rev. Sci. Instr. **14**, 235 (March 1927).

TABLE 1. *Distortion of two Hypergon-type lenses*

The measured values of the linear distortion D_m in millimeters for two Hypergon-type lenses are given for the indicated magnifications m and slope angles. The italicized digits beneath each distortion give the differences in microns between the values computed from D_0 and D_∞ and the measured values.

m	Slope angle							
	10°	15°	20°	25°	30°	35°	40°	45°
Lens A1, f/22, 127 mm								
0	0.000	0.001	0.005	0.012	0.024	0.041	0.067	[a]
∞	.001	.002	.004	.003	.000	−.007	−.013	[a]
0.2	.000	.001	.004	.013	.021	.044	.074	[a]
	0	*0*	*0*	*−2*	*+3*	*−2*	*−4*	
0.5	−.001	−.001	.003	.010	.025	.044	.074	[a]
	+1	*+1*	*0*	*0*	*−1*	*0*	*0*	
1	−.001	.000	.003	.014	.029	.054	.087	[a]
	0	*−1*	*−2*	*−5*	*−5*	*−6*	*−7*	
2	−.002	−.005	−.006	.003	.022	.044	.085	[a]
	0	*+2*	*+3*	*+3*	*+2*	*+11*	*+8*	
5	−.001	−.014	−.013	.001	.025	.068	.138	[a]
	−4	*+5*	*−2*	*−4*	*−1*	*+8*	*−6*	
Lens A2, f/22, 127 mm								
0	0.001	0.002	0.003	0.003	0.002	−0.001	−0.004	−0.008
∞	.001	.002	.007	.013	.031	.048	.064	.093
0.2	.000	.000	.000	.000	−.004	−.009	−.019	−.025
	+1	*+2*	*+2*	*0*	*0*	*−2*	*+2*	*−2*
0.5	.000	−.001	−.003	−.006	−.011	−.020	−.036	−.054
	0	*+2*	*+2*	*+2*	*−3*	*−5*	*0*	*0*
1	.000	.000	−.002	−.007	−.020	−.040	−.062	−.096
	0	*0*	*−2*	*−3*	*−9*	*−9*	*−6*	*−5*
2	−.002	−.004	−.012	−.028	−.049	−.090	−.146	−.209
	+1	*+2*	*+1*	*+5*	*−11*	*−7*	*+14*	*+15*
[b]5	−.002	−.003	−.017	−.046	−.092	−.160	−.262	−.375
	−2	*−5*	*−15*	*−16*	*−61*	*−81*	*−62*	*−98*

[a] The special barrel in which this lens was mounted limited the effective total field to about 84°.

[b] The differences obtained at this magnification are an example of the systematic error introduced by a poor setting of the transverse axis over the center of rotation of the nodal slide. The differences could be materially reduced by employing a calibrating procedure such as that illustrated in figure 10.

TABLE 2. *Distortion of three Dagor-type lenses*

The measured values of the linear distortion D_m in millimeters for three Dagor-type lenses are given for the indicated magnifications m and slope angles. The italicized digits beneath each distortion give the differences in microns between values computed from D_0 and D_∞ and the measured values.

m	Slope angle						
	10°	15°	20°	25°	30°	35°	40°
Lens B, f/22, 6½ in.							
0	0.004	0.018	0.048	0.110	0.232	0.462	0.919
∞	.005	.017	.041	.101	.212	.413	.838
0.25	.002	.011	.032	.081	.173	.355	.702
	+1	*+3*	*+6*	*+4*	*+6*	*+4*	*+7*
1	.001	.001	.004	.010	.020	.052	.095
	−2	*0*	*+3*	*−1*	*0*	*−3*	*−14*
4	−.009	−.048	−.120	−.280	−.603	−1.194	−2.380
	−7	*−2*	*+4*	*−14*	*−13*	*+4*	*−53*

TABLE 2. *Distortion of three Dagor-type lenses*—Continued

The measured values of the linear distortion D_m in millimeters for three Dagor-type lenses are given for the indicated magnifications m and slope angles. The italicized digits beneath each distortion give the differences in microns between values computed from D_0 and D_∞ and the measured values.

m	Slope angle						
	10°	15°	20°	25°	30°	35°	40°
Lens B3, f/22, 6½ in.							
0	0.004	0.017	0.049	0.116	0.244	0.488	0.967
∞	.002	.011	.032	.077	.171	.346	.796
0.25	.005	.013	.043	.100	.208	.412	.787
	−1	*+1*	*−2*	*−3*	*−7*	*−19*	*+3*
1	.001	.005	.016	.038	.073	.144	.265
	+1	*+1*	*+1*	*+1*	*0*	*−2*	*−4*
4	−.004	−.021	−.072	−.188	−.439	−.883	−1.884
	0	*−6*	*−7*	*−4*	*−1*	*−13*	*+27*
Lens B4, f/22, 6½ in.							
0	0.004	0.018	0.049	0.111	0.232	0.464	0.914
∞	.002	.012	.038	.086	.185	.372	.745
0.25	.005	.015	.039	.089	.191	.379	.739
	−1	*0*	*0*	*0*	*−5*	*−8*	*−11*
1	.002	.005	.013	.026	.047	.089	.163
	0	*+1*	*−2*	*−1*	*0*	*+3*	*+6*
4	−.009	−.031	−.100	−.216	−.526	−1.042	−2.098
	+5	*+1*	*−3*	*−17*	*+18*	*+18*	*+32*

TABLE 3. *Distortion of two Artar-type lenses*

The measured values of the linear distortion D_m in millimeters for two Artar-type lenses are given for the indicated magnifications m and slope angles. The italicized digits beneath each distortion give the differences in microns between the values computed from D_0 and D_∞ and the measured values.

m	Slope angle				
	10°	15°	20°	25°	30°
Lens, C1, f/22, 12 in.					
0	0.010	0.033	0.088	0.196	0.392
∞	.005	.024	.051	.121	.222
0.25	.008	.026	.073	.163	.331
	+1	*+1*	*+2*	*+3*	*+5*
0.5	.008	.023	.064	.137	.287
	0	*−2*	*−2*	*−1*	*−6*
1	.004	.008	.034	.073	.154
	+1	*+1*	*+3*	*+2*	*+16*
2	.000	−.006	−.016	−.054	−.061
	0	*−9*	*−2*	*+8*	*+2*
4	−.008	−.056	−.120	−.281	−.509
	−2	*−7*	*+4*	*−7*	*+13*
Lens C2, f/22, 16½ in.					
0	0.014	0.043	0.106	0.214	0.354
∞	.011	.040	.098	.215	.386
.5	.008	.022	.054	.115	[a]
	0	*+1*	*+3*	*−9*	
1	.001	.000	−.001	−.004	−.036
	+2	*+3*	*+9*	*+3*	*+4*
2	−.006	−.031	−.090	−.232	[a]
	−2	*−6*	*0*	*+16*	

[a] The conjugate distances were too great for the range of the optical bench.

TABLE 4. *Distortion of twenty Hypergon-type lenses*

The measured values of the linear distortion D_m in microns for 20 Hypergon-type lenses are given for parallel light and the indicated slope angles. These lenses have a nominal focal length of 127 mm and a nominal aperture of *f*/22.

Lens	m	Slope angle							
		10°	15°	20°	25°	30°	35°	40°	45°
A_3	0	2	3	3	7	12	19	29	54
	∞	2	4	6	7	12	17	21	27
A_4	0	−1	−2	−6	−11	−19	−24	−42	−63
	∞	1	5	10	18	32	52	82	124
A_5	0	2	3	8	13	20	29	41	64
	∞	0	1	3	3	2	0	−2	1
A_6	0	1	1	0	1	3	4	6	12
	∞	2	5	9	15	23	34	47	64
A_7	0	0	1	−5	−7	−6	−5	−4	6
	∞	0	4	6	15	23	33	49	67
A_8	0	3	7	12	19	28	42	65	91
	∞	−2	−4	−7	−9	−12	−16	−19	−19
A_9	0	0	2	3	6	11	14	20	26
	∞	0	2	4	6	11	16	27	45
A_{10}	0	0	2	6	10	16	25	38	56
	∞	1	3	5	5	2	1	1	1
A_{11}	0	0	1	3	6	8	12	12	12
	∞	0	1	2	6	11	17	30	52
A_{12}	0	0	0	3	6	13	20	32	54
	∞	0	1	3	7	10	14	16	22
A_{13}	0	0	0	1	3	5	7	11	22
	∞	1	3	5	9	17	27	39	56
A_{14}	0	0	2	3	6	11	19	30	53
	∞	0	0	0	2	6	12	18	23
A_{15}	0	1	3	5	6	9	12	14	22
	∞	0	−2	−2	4	11	22	39	55
A_{16}	0	−2	−4	−5	−6	−6	−3	1	14
	∞	1	3	7	12	25	32	40	58
A_{17}	0	0	1	2	5	9	13	18	25
	∞	0	0	1	5	10	17	26	42
A_{18}	0	1	3	4	6	7	7	9	10
	∞	−1	−2	−1	6	13	24	40	65
A_{19}	0	0	0	0	5	8	13	18	29
	∞	1	2	3	8	17	24	36	60
A_{20}	0	1	0	1	5	9	17	23	42
	∞	1	4	6	11	15	21	35	49
A_{21}	0	3	7	9	14	17	28	32	48
	∞	1	1	1	4	9	18	26	47
A_{22}	0	0	1	3	5	8	13	22	38
	∞	2	6	9	16	22	34	44	57

ancy occurs for lens A_2 at a magnification of 5, representing about three times the nominal probable error. The good agreement at other magnifications for this lens indicates that the greatest part of the discrepancy is a result of errors in the observed values. That the differences have the same algebraic sign is evidence of a systematic error, which can be attributed to the method of measurement. This consisted of making a run at 5° intervals of the slope angle without resorting to the time-consuming process of resetting the lens on the nodal slide for each angle measured. Thus, any longitudinal displacement of the transverse axis (located at T in fig. 5) from the center of rotation of the nodal slide for a given run results in a systematic error in the distortion. The magnitude of this error is very closely proportional to the tangent of the slope angle. With the exception of B_2 and B_5, all of the lenses were submitted to the Bureau for routine tests and were not available for sufficient time to allow more than 2, and in some cases 3, such runs to be made at each magnification.

The data obtained on four other lenses are presented graphically in figures 6, 7, 8, and 9. For convenience in plotting, the equation has been defined in two regions around unit magnification. D_m is plotted against m for $0 \leq m \leq 1$ and D_m/m against $1/m$ for $1 \leq m \leq \infty$. This allows the point $-D_\infty$ to be included as the $\lim_{m \to \infty} D_m/m$. The points give the experimental distortions and the straight lines the computed values from D_0 and D_∞. For the perfectly symmetrical lens, all the lines would intersect at zero distortion and unit magnification; only in this case would there be no discontinuity. For the lenses measured, minimum distortion occurs at some other magnification than unity, indicating slight asymmetries introduced in manufacture. For very low distortion lenses, these slight asymmetries can result in pronounced differences in distortion between lenses of the same type. This is clearly illustrated in table 4, where the measured values of D_0 and D_∞ in microns are given for 20 Hypergon-type lenses of the same nominal focal length submitted to the Bureau for test.

Equation (11) may also be applied to minimized distortions that have been calibrated by a given cri-

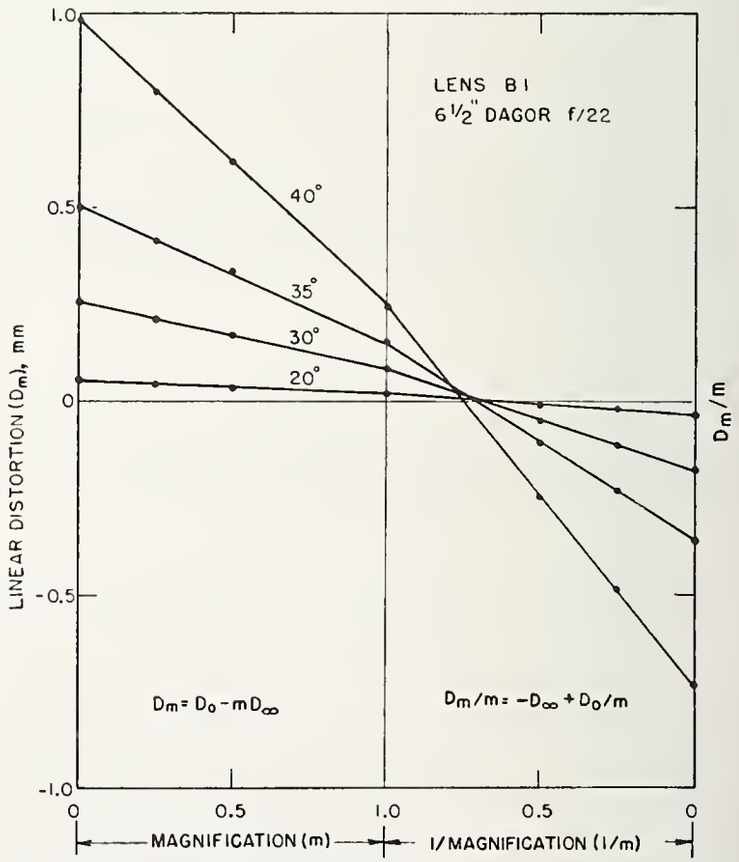


FIGURE 6. *Distortion plotted against magnification for the indicated slope angles.*

In order to include infinite magnifications, the graph has been split as shown into two regions at unit magnification. The straight lines represent the values computed from the two equations shown in their respective regions. The points indicate the experimentally determined values. The same graphical form has been employed in figures 7, 8, 9, 10, and 11.

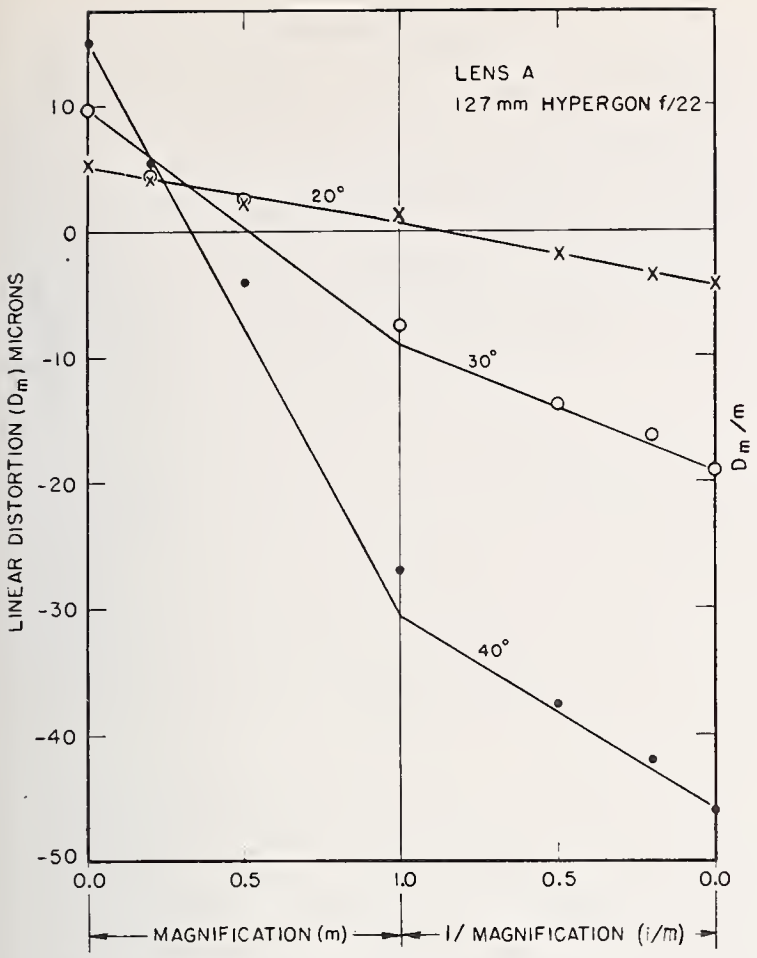


FIGURE 7. *Distortion plotted against magnification for the indicated slope angles.*

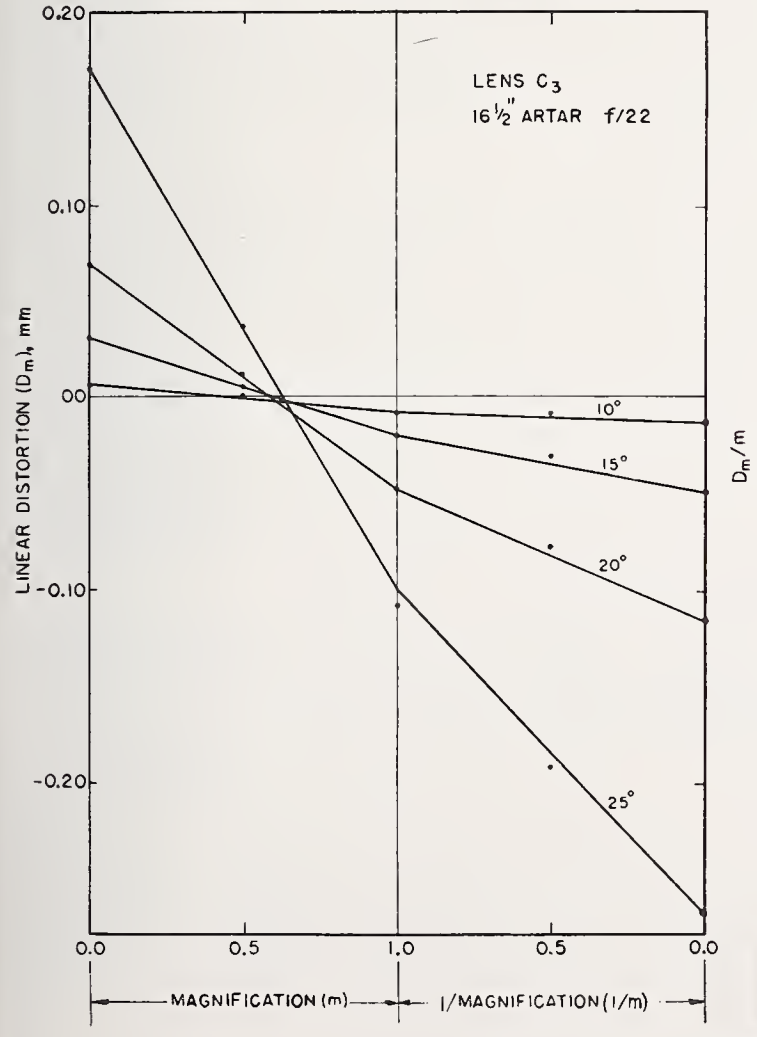


FIGURE 8. *Distortion plotted against magnification for the indicated slope angles.*

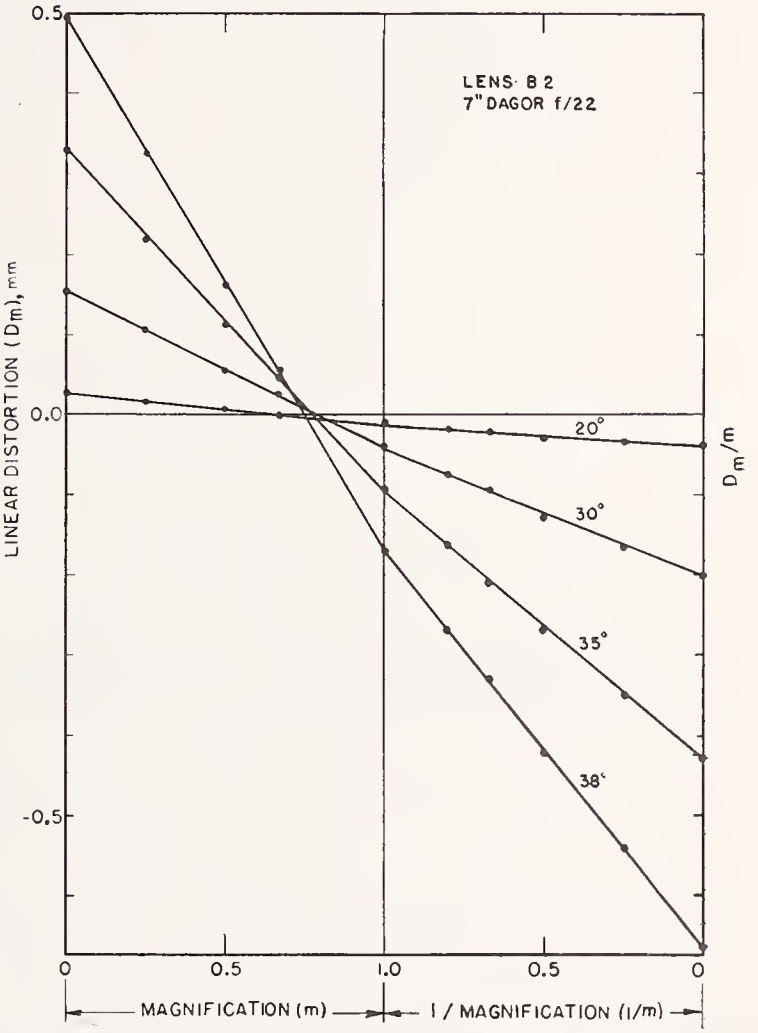


FIGURE 9. *Distortion plotted against magnification for the indicated slope angles.*

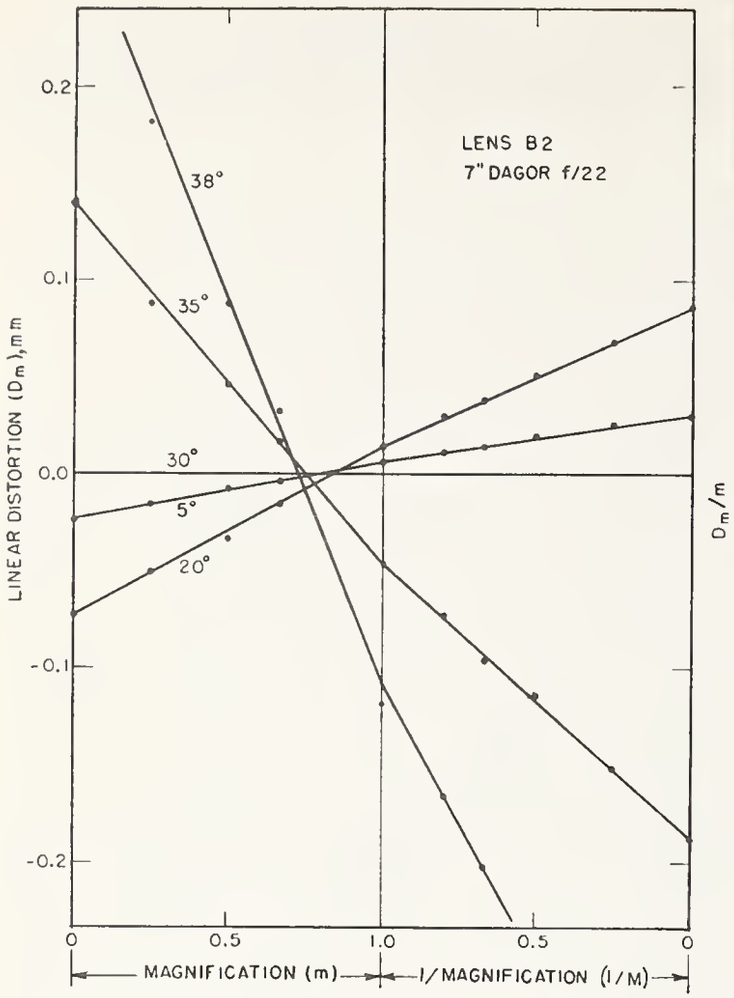


FIGURE 10. *Calibrated distortion plotted against magnification for the indicated slope angles.*

The data for the same lens as presented in figure 8 has been calibrated to give zero distortion at a slope angle of 30°. The magnitude of the change in distortion for any given magnification is proportional to the tangent of the slope angle.

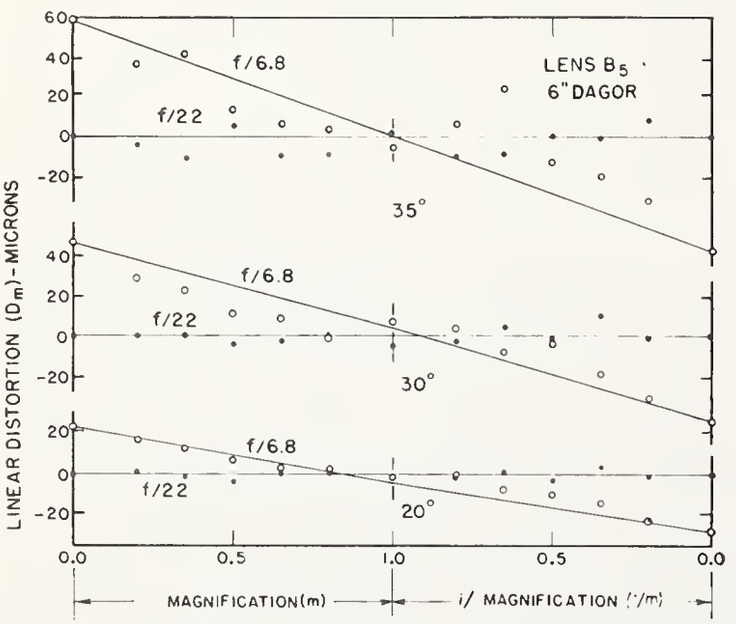


FIGURE 11. *Difference between the experimentally determined distortions for the indicated slope angles and those computed from the measured values of D_0 and D_∞ at f/22.*

The open circles represent data obtained at an aperture of *f*/6.8, and the solid circles at an aperture of *f*/22.

terion. At any given magnification, the magnitude of the change in distortion introduced by this calibration is proportional to the tangent of the slope angle, representing an effective shift in the position of the nodal points. For parallel incident light this results in calibrated focal lengths, which are equal only for a perfectly symmetrical lens. At finite object distances this results in a calibrated image distance or, more conveniently, a calibrated magnification. The calibrated focal lengths and magnifications are applied solely as scale factors and are not used in the determination of conjugate distances. Figure 10 shows the distortion presented in figure 8 calibrated to zero at a slope angle of 30°.

A detailed discussion of the effects on distortion of zonal aberrations introduced at apertures larger than *f*/22 is beyond the scope of this paper. However, their net effect can only result in a shift of the center of gravity of the nonparaxial image and a change in the plane of best focus. If the center of gravity shift is a linear function of the magnification and if the Gaussian identities expressed in eq (1) adequately determine the image plane of best focus when the higher aperture focal length is used, then eq (11) will still be valid.

In figure 11 the experimental distortions of a symmetrical-type lens at *f*/22 and the maximum aperture of *f*/6.8 have been plotted as the difference from the values computed from D_0 and D_∞ at *f*/22. The focal lengths and the values of D_0 and D_∞ for both apertures are given in table 5. No attempt was made to determine whether the image planes computed from the measured focal lengths were actually the planes of best over-all definition. At maximum aperture the effective stops were the outer lens retaining rings.

For this lens the experimental points for *f*/6.8 seem to lie along an S-shaped curve rather than being distributed about the straight line given by eq (11). The probable error of measurement is certainly higher at *f*/6.8 than at *f*/22, particularly for the larger slope angles. More data would have to be taken to determine if figure 11 is typical of all lenses of the same type, so at the present time the results must be considered as applying only to this particular lens. It is of interest that the difference between the distortion at *f*/22 and *f*/6.8 minimizes in the region about unit magnification, and elsewhere is generally greater than the deviation of the experimental distortions at *f*/22 from the computed values.

TABLE 5. *Change in focal length and distortion with aperature*

The change in focal length and distortion with aperature for two apertures of the Dagor-type lens B_5, with a nominal focal length of 6 in. The distortion is given in millimeters at the indicated slope angles.

Aperture	Focal length	m	Distortion		
			20°	30°	35°
	mm				
f/22	150.52	0	0.028	0.152	0.302
		∞	.047	.224	.435
f/6.8	150.12	0	.052	.199	.360
		∞	.075	.267	.492

5. Conclusion

The findings reported show that for low-aperture symmetrical lenses, the distortion at finite object distances can be simply computed from the two values obtained with parallel incident light with about the same accuracy as the direct experimental measurement. The necessary modifications required in carrying out this computation on other types of lenses have been indicated, and it is hoped that experimental verification of these procedures will soon be available.

Acknowledgment is made to F. E. Washer, who suggested the project and to W. R. Darling, who made many of the calculations and experimental measurements.

Washington, October 5, 1953.

Journal of Research of the National Bureau of Standards Vol. 56, No. 5, May 1956 Research Paper 2672

Characteristics of an Image-Forming System

Roland V. Shack

Two general approaches to the analysis of an image-forming system are considered. One depends on the image of a point object and the other on the Fourier transform of this image. The two are developed independently and then coordinated, a practical characteristic function being determined for each approach. The relative merits of the two approaches are considered.

1. Introduction

For the past few years considerable energy has been expended in the search for an objective procedure for evaluating the quality of images formed by optical instruments. Existing procedures have been found to be not entirely satisfactory, and much work has been done in measuring previously unused physical parameters, which are objectively determinable, for the purpose of correlating them with the existing quality criteria.

The objection to this is that such an empirical, and therefore much limited, correlation eliminates only one of the faults of the existing criteria, and this is the subjectivity of their determination. Any other weakness is ignored.

A better approach is to analyze the image-forming process as a phenomenon, the aim being to characterize the process in as general and inclusive a way as possible, consistent with practical instrumentation. New criteria of image quality would of course be expected to be developed. Many approaches have been made in this direction also, and the present paper is to be included among them. However, in contrast to some of the published material, the emphasis here is on the practicality and usefulness of the results obtained rather than on mathematical rigor, although the treatment should be rigorous enough to include all essential factors.

Let us consider this matter of practical instrumentation. The heart of the test instrument is the photosensitive detector, for it is this which provides the data by which the tested instrument is evaluated. Three practical photosensitive detectors are available—the eye, the photographic emulsion, and the photocell.

The only test of image quality for which the eye is capable of making quantitative measurements is the resolving-power test. This test is rapid and relatively inexpensive, but the information obtained is incomplete, the precision is low, and the results are subject to variation from individual to individual.

A photographic detector allows quantitative measurements to be made under nonthreshold conditions, but time is required for processing, the processing conditions must be rigidly standardized, the granularity and diffusion in the emulsion affect the results, the response of the film is nonlinear with respect to incident flux, and, in the end, an additional sensing mechanism, such as a microdensitometer, is needed to reduce the emulsion properties to numerical values.

The photocell is probably the most satisfactory photosensitive detector for the test instrument. Within its proper operating range, its characteristics remain reasonably constant, its response is linear with respect to incident flux, its spectral response can be adjusted so as to approximate that of the eye, and its output can very easily produce graphical or numerical results. However, it must be used in conjunction with an aperture that limits the spatial integration of the detail in the image being examined, and there must be provision for relative displacement between the aperture and the image so that various portions of the image may be sampled.

It should also be pointed out that the report is illustrated throughout by the characteristics of an aberration-free system with a circular aperture in monochromatic light, diffraction being the sole source of image degradation, and the light from various points in the object space being noncoherent. This has been selected as an interesting and informative type of system, which real systems tend to approximate as their quality improves. It must be emphasized,

however, that this is used for illustration only. The material covered applies to all types of images, assuming noncoherence.

There are two viewpoints from which image evaluation can be approached. One, the classical viewpoint, considers the point image to be the most fundamental element in an image process. Any object can be regarded as a summation of points, and its image as the summation of the corresponding point images. An evaluation of the point image would be sufficient to characterize the system.

The other viewpoint involves the concepts of Fourier analysis. Here the object is considered to be the summation of a set of sinusoidal waves distributed in the object plane. These component waves, differing from each other in amplitude, frequency, phase, and direction, are spatially distributed waves. That is, they are spatial, not temporal, sinusoidal variations in brightness throughout the object plane. The image consists of the summation of the images of these component waves. A description of the manner in which the optical system forms images of these component waves would also be sufficient to characterize the system.

In section 2 a way of describing the point image is developed, which can be obtained, at least in principle, from a variety of test objects, namely, a point, a fine line, a variable slit, and a knife edge. Data from any or all of these objects can be represented by a single common curve, which can be interpreted in terms of any of the objects.

Section 3 deals with the Fourier type of approach. An imaging system does not affect the frequency, direction, or sinusoidal character of the component waves. It can only affect their amplitudes and phase relationships. The function that describes these modifications as a function of the frequency and direction of the component waves is also characteristic of the imaging system. The test object required to obtain this information consists of a series of patterns in which the luminance varies sinusoidally, each pattern having a different spatial frequency and all oriented in the same direction. The directional variation can be obtained by rotating the test object with respect to the system being tested.

Section 4 is concerned with the coordination between the two viewpoints. It is shown that it is possible to obtain from either approach the characteristics for a periodic test object consisting of alternate dark and light stripes of equal width, such as is commonly used in resolving-power work. It is also shown that it is possible to transform the characteristics of either approach into those of the other.

Section 5 discusses the application to practice, the interpretation of the results and the relative merits of the two approaches.

The appendix gives the mathematical formulation of the diffraction images used as illustrations.

In the following mathematical statements, constant coefficients are ignored in the integrations unless otherwise indicated. The functions are assumed to be normalized after integration. Also, the object-plane coordinates are reduced to the image plane by application of the magnification.

2. Evaluation of the Point Image

2.1. General Image Formation with the Point Image

The point image is the flux-density distribution in the image plane when the object is a point source (figs. 1, 2). A general object can be considered to consist of a summation of points, and its image the summation of the corresponding point images.

Let $O(x,y)$=general object function,
$\varphi(x,y)$=point image function,
$I(x,y)$=general image function.

Then

$$I(x',y')=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} O(x,y)\varphi(x'-x,y'-y)dxdy. \tag{1}$$

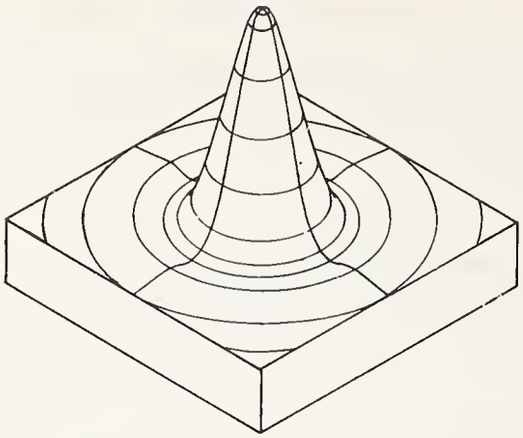


FIGURE 1. *Point-image model.*

In this model the vertical dimension represents flux density (per unit area). It is understood that the rings actually continue indefinitely, whereas only the first bright ring is shown completely.

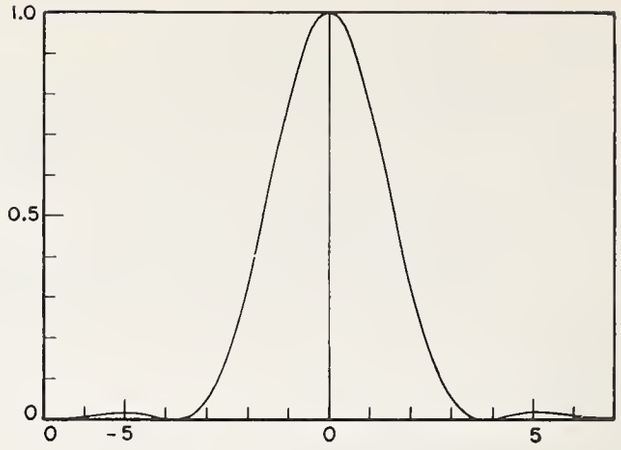


FIGURE 2. *Section of point image.*

The normalized distance in the image plane is measured in z-units, as explained in appendix 1.

The primed variables represent the displacement between the O and φ functions required for the integration. For each displacement, the integral, being a definite integral, establishes a specific value for I. The image function, I, then, is a function of the displacement involved in the integration. The space described by x',y', however, is the same as that described by x,y. O and I can be compared point for point.

The integration can also be written in such a form that the object function is the displaced function, that is,

$$I(x',y')=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)O(x'-x,y'-y)dxdy. \tag{2}$$

Either of these forms is perfectly valid, and either may be transformed into the other, provided one recognizes that x,y in eq (1) is not the same as x,y in eq (2). To distinguish them, one might use subscripts on the symbols, but this would make the equations more confusing, and is not necessary if one understands the situation.

The image function is to be sampled with a space-integrating detector in the image plane, that is, a photocell behind an aperture.

Let $A(x,y)$=detector aperture transmission function,

$E(x,y)$=total flux passing through A as a function of the position of A.

Then

$$E(x'',y'')=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}I(x',y')A(x''-x',y''-y')dx'dy'. \tag{3}$$

Combining this with eq (2) we obtain

$$E(x'',y'')=\int_{-\infty}^{\infty}\int\int\int\varphi(x,y)O(x'-x,y'-y)A(x''-x',y''-y')dxdydx'dy'. \tag{4}$$

From this it can be seen that the functional characteristics of O and A can be interchanged without affecting the measured flux, E. For example, suppose the object were a point source and the aperture a circular hole centered on the point image. The output from the photocell will indicate amount of flux passing through the hole. Then interchange the object and aperture. Now the object is a uniformly luminous disk with a reduced diameter equal to that of the previous aperture, and the new aperture is a minute hole, equal in size to the reduced geometric size of the previous point source. The output from the photocell will be the same as before.

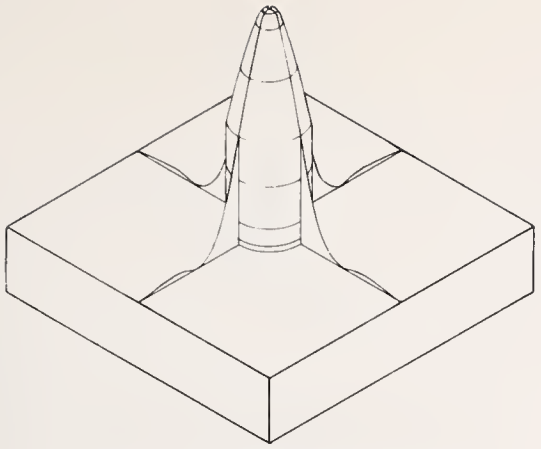


FIGURE 3. *Point-image evaluation—Hopkins' method.*

The solid represents the portion of the total flux that passes through the circular aperture.

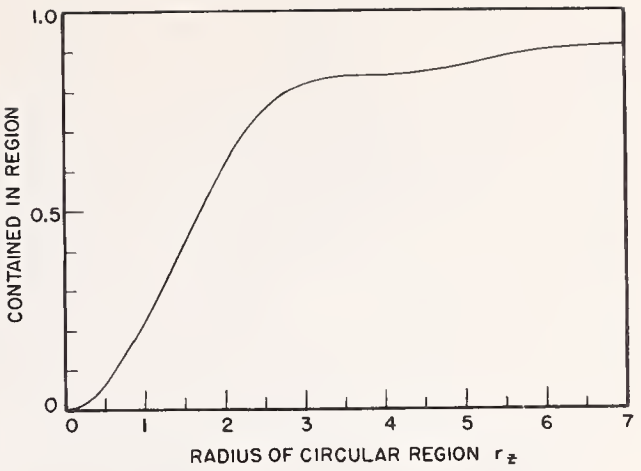


FIGURE 4. *Radial flux distribution of point image.*

This curve shows the normalized volume of the solid of figure 3 as a function of the radius of the limiting cylinder.

Mathematically, and in a more general sense, this situation is as follows:
If O is a point source, then

$$E(x'',y'')=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A(x''-x',y''-y')\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)O(x'-x,y'-y)dxdy\right]dx'dy'$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x',y')A(x''-x',y''-y')dx'dy', \tag{5}$$

and the flux detected depends on the nature of A.

On the other hand, if A is a point aperture, then

$$E(x'',y'')=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}O(x'-x,y'-y)A(x''-x',y''-y')dx'dy'\right]dxdy$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)O(x''-x,y''-y)dxdy, \tag{6}$$

which is identical in form with eq (5), except that A is replaced by O. If O in eq (6) had the same functional characteristics as A in eq (5), then they would be mathematically indistinguishable, and the same E will be obtained from either. Also note that eq (6) has the same form as eq (2). As one would expect, the use of a point aperture would produce an undistorted map of the flux-density distribution in the image plane.

2.2. Determination of a Characteristic Function of the Point Image

It should be clear from the above that the function which distinguishes one image-forming system from another is the point-image function φ. The problem is to find some way of describing φ, which can be obtained experimentally and which provides significant information to the user.

The direct mapping of the flux-density distribution in the image of a point object with a point aperture is impractical because of the very low-energy levels involved. A practical method must involve in some manner the integration of the flux over an area.

One method, used by Hopkins [7] [1] is to measure the flux contained within successive concentric circular regions about the center of the point image (fig. 3). The ratio of the flux contained within a circular region to the total flux of the point image is plotted as a function of the radius of the circle (fig. 4).

[1] Figures in brackets indicate the literature references at the end of this paper.

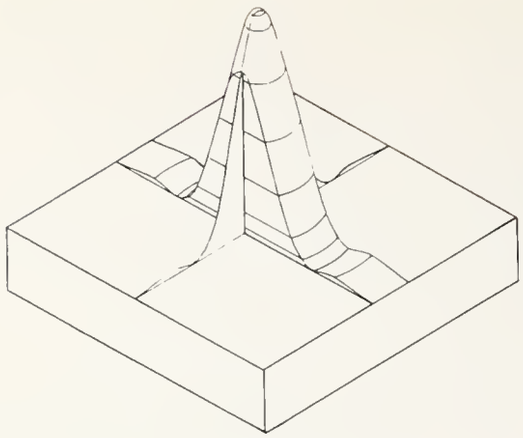


FIGURE 5. *Determination of point-image characteristic.*

The solid represents the portion of the total flux that passes through the receiver slit, the point image being centered on the slit.

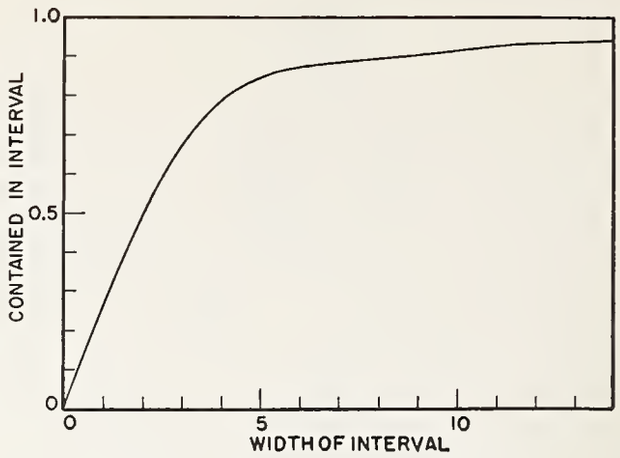


FIGURE 6. *Point-image characteristic.*

This curve shows the normalized volume of the solid in figure 5 as a function of the width of the slab.

This method has several advantages over mapping the flux-density distribution directly. Of course, there is an increase in the energy involved, which increases the signal-to-noise ratio of the measurements. Also, there is a reduction of a three-dimensional function to two dimensions, which is more convenient. Further, this method provides the user with an idea of the contrast with which small detail will be imaged, for, by the interchangeability of the object and aperture function, the resulting curve (fig. 4) can be considered to represent the flux density at the center of the image of a disk as a function of the size of the disk. This disk corresponds to a small object in the scene, later observed by the user.

This method does have disadvantages, however. It is satisfactory for systems in which the point image is radially symmetrical, but is insensitive to the presence of radial asymmetry, such as exists in an astigmatic or comatic image. It also presents the practical difficulty of locating the centers of the apertures or disks in two dimensions with precision.

A different integrative method that allows a considerable increase in the available energy is one in which the integration is limited in one direction only. This is done by measuring the flux contained in successive widths about the center of the point image (fig. 5). The normalized flux contained in a region as a function of the width of the region (fig. 6) is the function that is here proposed as the most useful and practical characteristic function describing the point image. It will hereafter be called the point-image characteristic $K(w)$, where w represents the width of the region. For example, if the integration is limited in the x-direction, then

$$K(w_x)=\int_{-w_x/2}^{w_x/2}\int_{-\infty}^{\infty}\varphi(z,y)\,dy\,dx. \tag{7}$$

It should be noticed that this function does not actually involve a reduction of three dimensions to two, because the curve obtained is a function of the direction in which the integration is taken. This makes the data somewhat less convenient than is true of the previous method, but this is not objectionable because of the additional information obtained. The new method will detect a lack of radial symmetry. However, the inconvenience involved is not too great because most images are bilaterally symmetrical, or nearly so, and two mutually perpendicular orientations are all that are necessary to characterize the image.

The point-image characteristic has other virtues beyond the relatively large amount of available energy and the abilĭty to detect lack of radial symmetry. These arise out of the unidirectional limitation of the integration.

For example, the point-image characteristic is closely connected with the fine-line image. The fine-line image is the image of a line of infinite length but infinitesimal width. The flux density is constant along the length of the line image and varies in a direction perpendicular

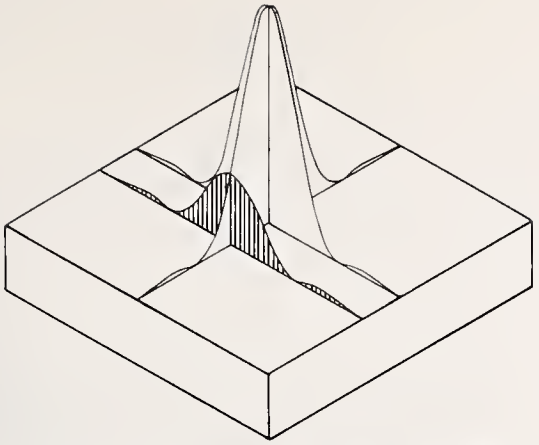


FIGURE 7. *Generation of fine-line image.*

The flux density at a point in a fine-line image is proportional to the area of the section of the point image orientated in the direction of the line.

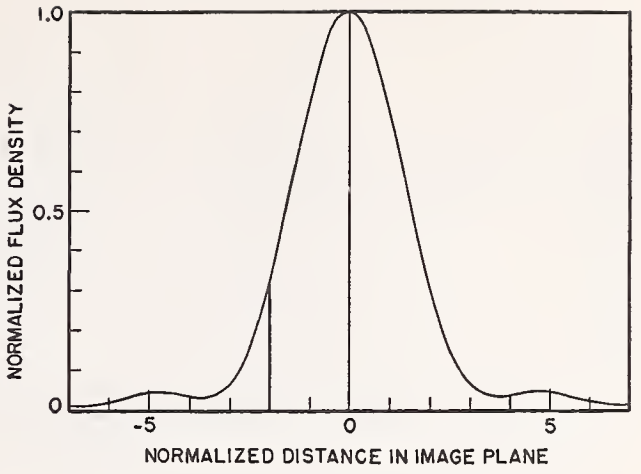


FIGURE 8. *Cross section of fine-line image.*

The vertical line corresponds to the area of the section shown in figure 7.

to the length. It is obtained by integrating the point image in one direction only. For example, the image function for a fine line, the length of which is in the y-direction, is given by

$$L(x)=\int_{-\infty}^{\infty}\varphi(x,y)dy. \tag{8}$$

This is illustrated in figures 7 and 8.

The point-image characteristic can be obtained from the fine-line image by

$$K(w_x)=\int_{-w_x/2}^{w_x/2}L(x)dx. \tag{9}$$

This is illustrated in figure 9.

Up to this point we have been considering the object to be a point or a fine line and the detector aperture to be a variable slit, the operating mechanism that produces the variation in w. But, because of the interchangeability of the object and aperture, an illuminated variable slit could be used as an object and a fine slit centered on the variable slit image as the detector aperture. This method of obtaining the point-image characteristic is illustrated in figure 10.

The point-image characteristic is also closely connected with the knife-edge image. The latter is related to the point image as follows:

$$S(x')=\int_{-\infty}^{x'}\int_{-\infty}^{\infty}\varphi(x,y)\,dy\,dx. \tag{10}$$

This is shown in figures 11 and 12.

The relationship between the knife-edge image and the fine-line image is given by

$$S(x')=\int_{-\infty}^{x'}L(x)dx. \tag{11}$$

The point-image characteristic is obtainable from $S(x)$ by observing the values of $S(x)$ at $x=-w_x/2$ and $x=+w_x/2$. Then

$$K(w_x)=S(w_x/2)-S(-w_x/2)=\int_{-\infty}^{w_x/2}L(x)dx-\int_{-\infty}^{-w_x/2}L(x)dx=\int_{-w_x/2}^{w_x/2}L(x)dx. \tag{12}$$

These relationships are illustrated in figures 13 and 14.

In summary, it may be pointed out that the point-image characteristic provides a simple yet practical and informative way of evaluating an imaging system. It may be obtained from a variety of test objects, namely, a point, a fine line, a variable-width line, or a knife edge.

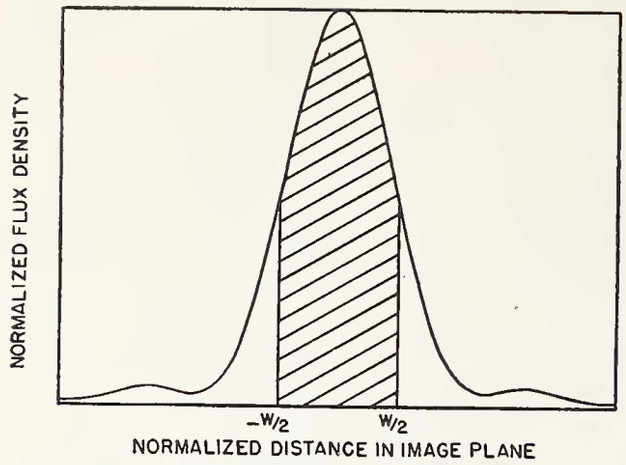


FIGURE 9. *Determination of point-image characteristic from fine-line image profile.*

The cross-hatched area corresponds to the solid in figure 5.

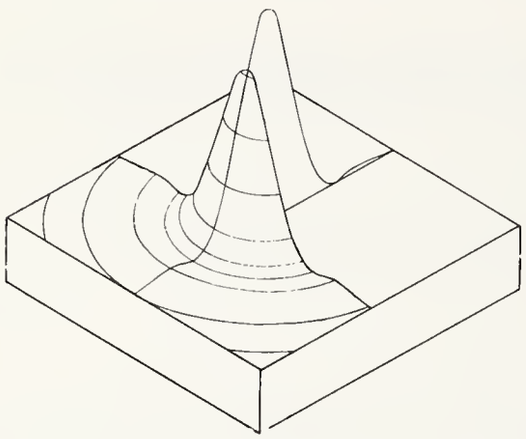


FIGURE 11. *Generation of knife-edge image from point image.*

The flux density at a point in the knife-edge image is proportional to the volume of the solid indicated where the limiting plane has the same orientation as the knife edge and passes through the point in question.

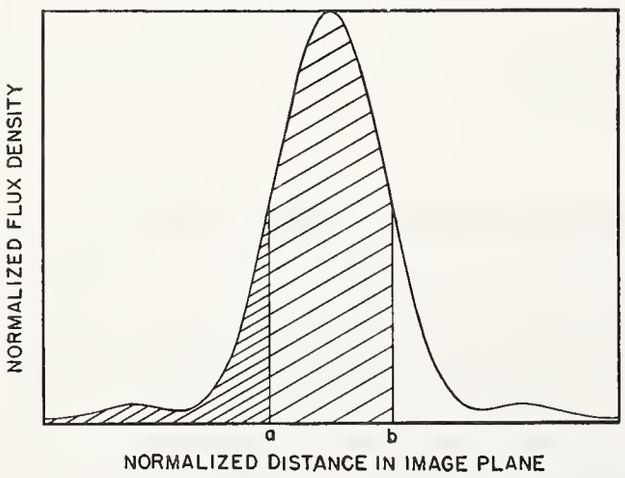


FIGURE 13. *Generation of knife-edge image from fine-line image.*

The flux density at a point in the knife-edge image is proportional to the area under the fine-line image to the left of the corresponding abscissa. The area under the fine-line image between any two abscissa values is proportional to the difference between the corresponding ordinate values of the knife-edge image.

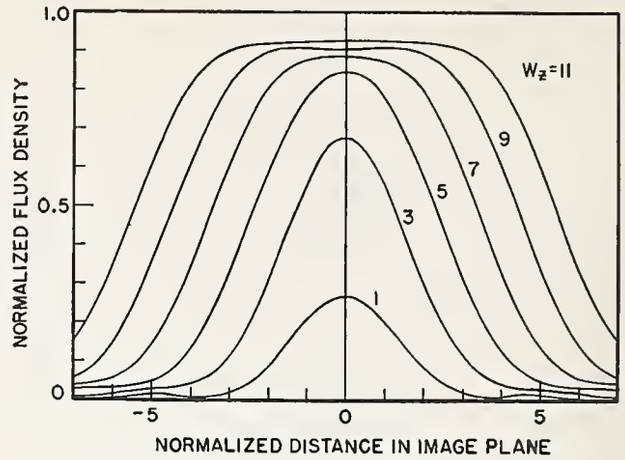


FIGURE 10. *Cross sections of images of finite-width object lines.*

The reduced object line widths are 1, 3, 5, 7, 9, and 11 x-units. Plotting the central flux density as a function of the reduced object line width results in the point-image characteristic curve shown in figure 6.

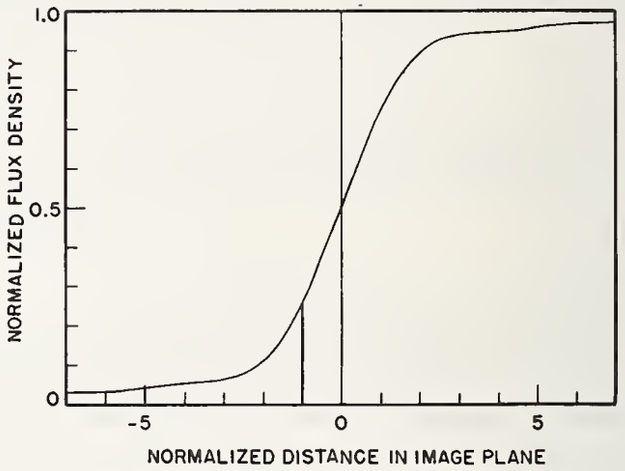


FIGURE 12. *Cross section of knife-edge image.*

The vertical line corresponds to the solid shown in figure 11.

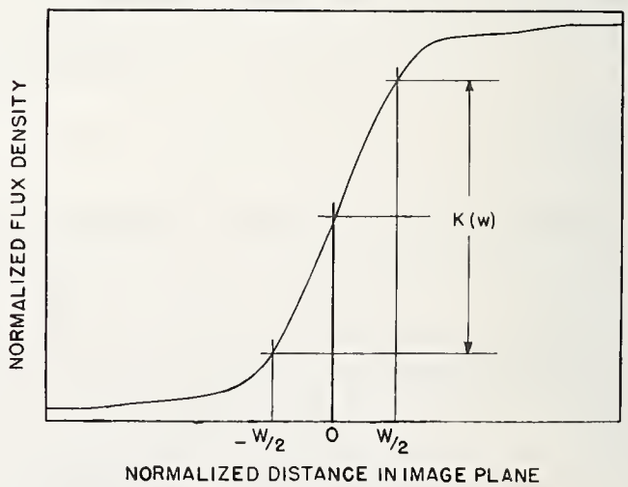


FIGURE 14. *Determination of point-image characteristic from knife-edge image.*

If the abscissa values a and b in figure 13 are made equal to $-w/2$ and $+w/2$ as in this figure, then the ordinate difference shown here is proportional to the corresponding area in figure 13, which itself is proportional to the corresponding value of the point-image characteristic, as indicated in figure 9.

Furthermore, regardless of which way it is obtained, it can be interpreted in terms of any of the objects. This is discussed further in a later section.

3. Evaluation by Fourier Analysis

A general object $O(x,y)$ can be analyzed into a continuum of sinusoidal spatial waves, differing from each other in direction, frequency, amplitude, and phase. The characteristics of these waves are given by the Fourier transform of the object,

$$T_o(\omega_x,\omega_y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} O(x,y)\ \exp\ [-i(\omega_x x+\omega_y y)]dxdy, \tag{13}$$

where $T_o(\omega_x,\omega_y)$ is complex, containing both amplitude and phase factors, and ω_x and ω_y are directional frequency components of the component waves.

A component wave itself is represented by

$$W_c=T_o(\omega_x,\omega_y)\ \exp\ [i(\omega_x x+\omega_y y)]. \tag{14}$$

Consider this to be an object. Then by applying eq (2),

$$I_W(x',y')=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)T_o(\omega_x,\omega_y)\ \exp\ [i\{\omega_x(x'-x)+\omega_y(y'-y)\}]dxdy$$

$$=T_o(\omega_x,\omega_y)\ \exp\ [i(\omega_x x'+\omega_y y')]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)\ \exp\ [-i(\omega_x x+\omega_y y)]dxdy. \tag{15}$$

The integral is the Fourier transform $\Phi(\omega_x,\omega_y)$ of $\varphi(x,y)$, and therefore

$$I_W(x',y')=\Phi(\omega_x,\omega_y)T_o(\omega_x,\omega_y)\ \exp\ [i(\omega_x x'+\omega_y y')]. \tag{16}$$

The function modified by Φ is seen to be simply the component object wave. Each component image wave then will be the product of Φ and the corresponding component object wave. Therefore,

$$T_I(\omega_x,\omega_y)=\Phi(\omega_x,\omega_y)T_o(\omega_x,\omega_y), \tag{17}$$

where T_I is the Fourier transform of the image.

If $I(x,y)$ can be considered to be the object of a second imaging precess, then

$$T_{I_2}(\omega_x,\omega_y)=\Phi_2(\omega_x,\omega_y)T_{I_1}(\omega_x,\omega_y)=\Phi_2\Phi_1 T_o. \tag{18}$$

This can be extended to as many imaging processes as desired.

The transform

$$\Phi(\omega_x,\omega_y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)\ \exp\ [-i\omega_x x+\omega_y y)]dxdy \tag{19}$$

is seen to be characteristic of the imaging process. Let us examine it in more detail.

The transform is the double integral of the product of two functions, one being the point image $\varphi(x,y)$ and the other a two-dimensional wave $exp[-i(\omega_x x+\omega_y y)]$.

This two-dimensional wave is sinusoidal in one direction and constant in the perpendicular direction, like a corrugated roof. The direction of the sinusoidal variation is inclined to the x-axis by an angle θ, where $\tan\theta=\omega_y/\omega_x$, and the frequency is given by $\omega_\theta=\sqrt{\omega_x^2+\omega_y^2}$ (see fig. 15). The transform can then be considered to be $\Phi\,(\theta,\omega_\theta)$.

For $\theta=0$, $\omega_y=0$ and the transform becomes

$$\Phi(\omega_x)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)\exp[-i(\omega_x x)]dydx=\int_{-\infty}^{\infty}L(x)\exp[-i(\omega_x x)]dx, \tag{20}$$

518-743 O - 73 - 34

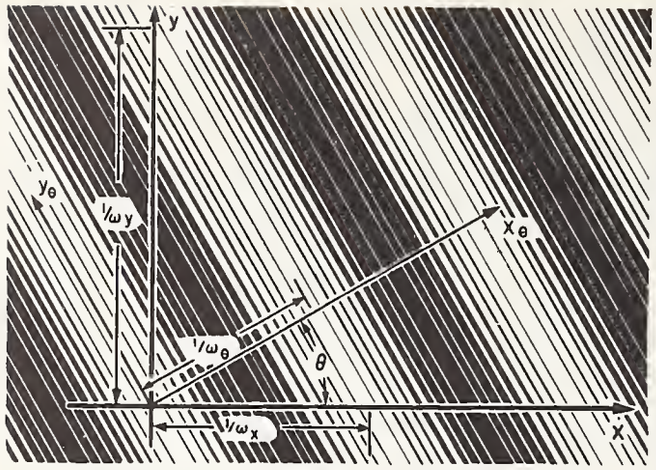


FIGURE 15. *Transformation of coordinates.*

If the direction of the pattern is inclined to the original x-axis by an angle θ, then, in the new coordinate system in line with the pattern, $\omega_\theta = \sqrt{\omega_x^2 + \omega_y^2}$ and $\tan\theta = \omega_y/\omega_x$.

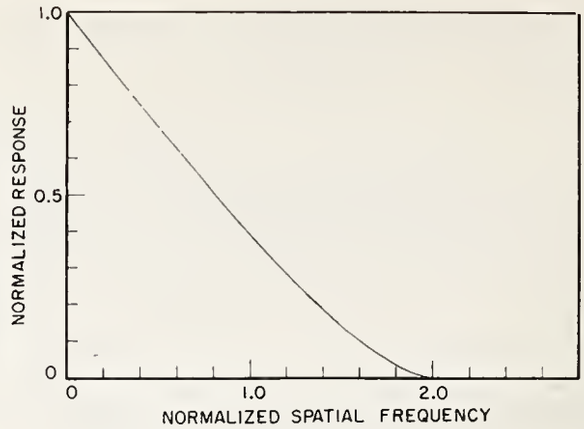


FIGURE 16. *Sine-wave response curve.*

This is the positive side of the cross-section of the Fourier transform of the point image shown in figure 1. It is also the Fourier transform of the fine-line image shown in figure 8.

where $L(x) = \int_{-\infty}^{\infty} \varphi(x,y)dy$ is the fine-line image (fig. 16).

If the x, y coordinates are rotated in x_θ, y_θ through the angle θ (fig. 15), then eq (20) can be generalized into

$$\Phi(\theta,\omega_\theta) = \int_{-\infty}^{\infty} L(x_\theta)\exp[-i(\omega_\theta x_\theta)]dx_\theta. \tag{21}$$

For any θ, the transform Φ of eq (21) is the cross section in the θ-direction of the solid representing the transform $\Phi(\omega_x, \omega_y)$. This is shown in figure 17.

Equation (21) serves as the basis for the experimental determination of Φ, since for each θ it is the transform of the fine-line image oriented with its length perpendicular to the θ-direction. The object can be one in which the flux density varies sinusoidally in one direction, and the detector aperture can be a fine slit.

The object can be represented by

$$O(x_\theta) = A + B\ \exp[i\omega_\theta x_\theta], \quad A \geq B. \tag{22}$$

The constant term is necessary because all flux densities must be positive. The image is given by

$$I(x_\theta) = A + \Phi(\omega_\theta)B\ \exp[i\omega_\theta x_\theta]. \tag{23}$$

Let $M_o = B/A$ be defined as the modulation in the object. The corresponding modulation in the image is $M_I = \Phi B/A$. The transform Φ is obtained by

$$\Phi = \frac{M_I}{M_o}. \tag{24}$$

The use of the modulation automatically compensates for any factor that changes the signal output of the testing device proportionally, such as the transmission factor of the lens, change in brightness of the source, change in gain of the detector amplifier, etc.

It must be remembered that Φ is actually complex, involving both an amplitude and a phase factor. However, if the origin of $\varphi(x,y)$ is properly selected, the phase shift is small. If $\varphi(x,y)$ is symmetrical about its origin, then there is no phase shift involved. In most cases, for evaluation purposes, the phase-shift factor can be neglected.

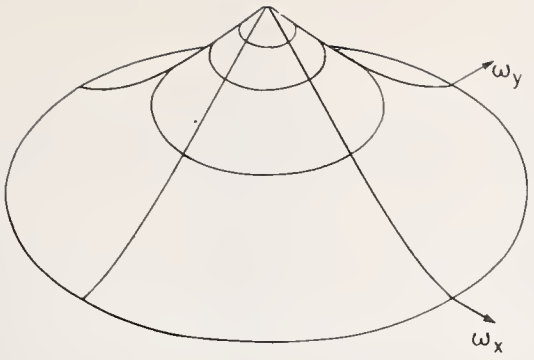


FIGURE 17. *Three-dimensional model of Fourier transform of point image.*

Each radial cross section is the sine-wave response in that direction. Note that the space is frequency space.

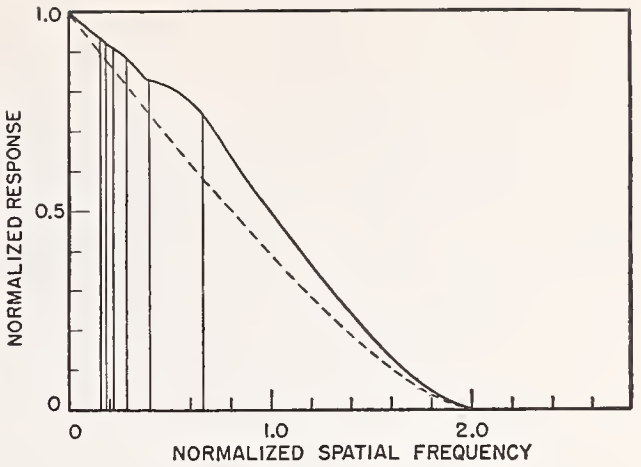


FIGURE 18. *Square-wave response curve.*

The dashed line is the sine-wave response curve from which this was derived. The vertical lines separate regions in which the square-wave image contains different numbers of harmonic components, each region having one more component than its neighbor to the right. In the region farthest to the right, the square-wave image contains only the fundamental.

4. Coordination

If $\Phi(\omega)$ is given, it is possible to predict a similar response curve in which the test object consists of alternate dark and light stripes, a spatial square wave. The square-wave test object can be analyzed into its component harmonics, each of which is attenuated by the value of Φ corresponding to its frequency, and the image is obtained by adding together these attenuated harmonics. For the response curve we are only interested in the peak-to-peak values that are obtained from the values at the centers of the lines and spaces. The square-wave response curve then is given by

$$\psi(\omega)=\frac{4}{\pi}\left[\Phi(\omega)-\frac{1}{3}\Phi(3\omega)+\frac{1}{5}\Phi(5\omega)-\cdots\right]. \tag{25}$$

It should be noticed that there will be only a finite number of terms in the sum because there is a limiting value of ω beyond which Φ remains at zero. This limit exists because of the finite dimensions of the aperture of the system; the larger the aperture the higher the limiting frequency.

Because of the finite range in ω, it is possible to obtain $\Phi(\omega)$ from $\psi(\omega)$ by the inverse process. Here it is necessary to start at the limiting value of ω and work backward. From this limiting value ω_c back to $\omega_c/3$ the sine-wave response is given by

$$\Phi(\omega)=\pi/4\ \psi(\omega) \tag{26}$$

because Φ and ψ are both zero for the odd multiples of ω for ω greater than $\omega_c/3$.

From $\omega_c/3$ to $\omega_c/5$,

$$\Phi(\omega)=\pi/4\ \psi(\omega)+\tfrac{1}{3}\Phi(3\omega), \tag{27}$$

where $\Phi(3\omega)$ may not be zero, and has already been found.

From $\omega_c/5$ to $\omega_c/7$,

$$\Phi(\omega)=\pi/4\ \psi(\omega)+\tfrac{1}{3}\Phi(3\omega)-\tfrac{1}{5}\Phi(5\omega), \tag{28}$$

and so forth.

The relationship between the square-wave response and the sine-wave response is illustrated in figure 18.

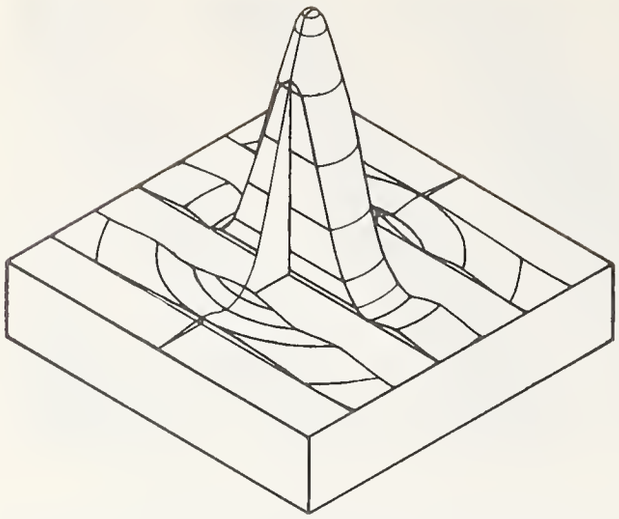


FIGURE 19. *Generation of square-wave response from point image.*

The sum of the volumes of the slabs is proportional to the flux density at the center of a bright line in the image. The sum of the volumes of the slabs that have been removed corresponds to the dark line. The difference between the two is proportional to the square-wave response.

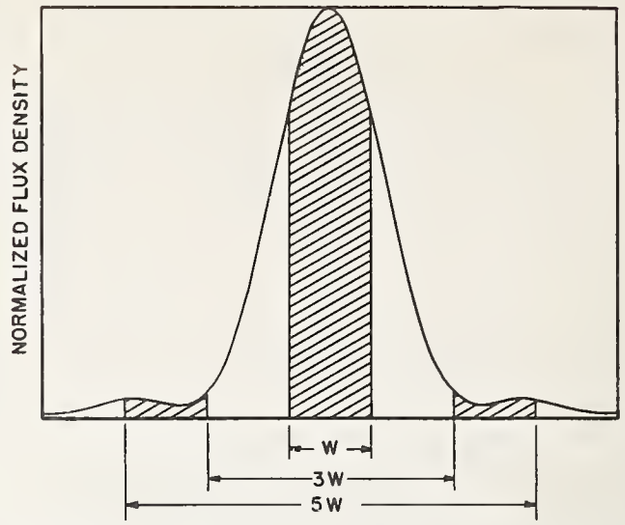


FIGURE 20. *Generation of square-wave response from fine-line image.*

The cross-hatched areas correspond to the slabs in figure 19. The marked widths indicate how the square-wave response can be obtained from the point-image characteristics.

The square-wave response may also be obtained from the point-image characteristic. If we consider the imaging of a square-wave test object as a point-image integration, then the flux density at the center of the image of a bright line in the pattern is proportional to the sum of the volumes of the slabs illustrated in figure 19, or the areas of the stripes in figure 20. The flux density at the center of the image of a dark line is proportional to the sum of the volumes of the slabs (or areas of the stripes) which lie between those illustrated. The sum of these two flux densities is proportional to the total flux in the point or fine-line image, and the square-wave response is given by

$$\Psi(\omega)=E\left(\frac{1}{2\omega}\right)-\left[1-E\left(\frac{1}{2\omega}\right)\right]=2E\left(\frac{1}{2\omega}\right)-1, \tag{29}$$

where $E(1/2\omega)$ is the flux density at the center of the image of a bright line and $1/2\omega=w$ is the width of a line, bright or dark, in the pattern.

It can be seen from figure 20 that

$$E\left(\frac{1}{2\omega}\right)=E(w)=K(w)+[K(5w)-K(3w)]+[K(9w)-K(7w)]+\ldots, \tag{30}$$

where $K(w)$ is the point-image characteristic, so

$$\Psi(\omega)=2\{K(w)+[K(5w)-K(3w)]+\ldots\}-1. \tag{31}$$

It is now apparent that $\Phi(\omega)$, the sine-wave response, may be obtained from $K(w)$, the point-image characteristic. This is done by determining $\psi(\omega)$, the square-wave response, from $K(w)$ and then applying the procedure indicated in eq (26), (27), and (28) to obtain the sine-wave response.[2]

To obtain the point-image characteristic from the sine-wave response is more direct. Consider the test object to be a variable-width line. The transform of the image of this object is obtained by multiplying the transform of the object by the sine-wave response of the system.

[2] This calculation for the case of the unaberrated image was made by the author before he was aware of Steel's equation (see appendix). It was done for 30 different spatial frequencies, using the tabulated values for the knife-edge image given by Struve (see appendix) as well as his approximation for values beyond those tabulated. The agreement between this calculated sine-wave response and Steel's equation is well within the error of calculation.

The transform of this object is of the form

$$T_o = w \frac{\sin \pi \omega w}{\pi \omega w} \tag{32}$$

shown in figure 21, and the transform of the image is given by $T_I = \Phi T_o$. The image itself would be given by the inverse transform

$$I(x_\theta) = \int_{-\infty}^{\infty} \Phi T_o \exp\left[i(\omega_\theta x_\theta)\right] d\omega_\theta, \tag{33}$$

but we are interested only in the value at the center of the line image where $x_\theta = 0$, so

$$K(w_\theta) = I(O) = \int_{-\infty}^{\infty} \Phi T_o d\omega_\theta = 2w \int_0^{\infty} \Phi(\omega_\theta) \frac{\sin \pi \omega w}{\pi \omega w} d\omega_\theta. \tag{34}$$

This equation shows the direct manner in which $K(w)$ may be obtained from $\Phi(\omega)$. For each w selected, the sine-wave response is multiplied by the proper calculated function of the form $\sin x/x$, and the product is integrated graphically or numerically. The result is then multiplied by $2w$, and the value obtained is the desired $K(w)$.

5. Application

It has been shown that an imaging system can be evaluated by means of a description of the point image or, alternatively, a description of its Fourier transform. Each of them has its advantages.

The Fourier approach allows us to combine several systems or to analyze a system into its components, under certain conditions. The principal condition is that each intermediate image formed by each component must be equivalent to a real luminous object having the same flux-density distribution in the image plane. Thus the combining or analyzing process can be applied to a photographic process, a television system, or a system in which each intermediate image is formed on a diffusing surface. It cannot be applied to a telescopic system because the light emerging from the primary image still contains the wave-front deformations produced by the objective and is subject to aberration correction by the ocular.

This does not mean that the Fourier approach cannot be applied to the entire telescopic system. It is just that the whole telescope cannot be precisely evaluated by simply evaluating the components.

Using the Fourier approach also enables one to apply the concepts developed in information and communication theory to imaging processes.

Another virtue of the Fourier approach is that the transform determined for any orientation of the sine-wave pattern in the image plane is a true section of the solid representing the entire transform in the image plane. A similar property does not hold for the point-image characteristic.

It should be noticed also that the sine-wave response of a lens has a finite boundary determined by the aperture of the system, whereas the point-image characteristic is unbounded.

The point-image characteristic, on the other hand, directly provides two basic types of information about the performance of the system, apart from the distribution of flux in the point image. These are the contrast at which an isolated object will be imaged as a function of the object size, and the gradient in the image of the edge of an extended object. The first holds because the point-image characteristic can be obtained by measuring the flux densities at the centers of finite-width line images. The second holds because the characteristic can also be obtained from the image of a knife-edge object.

However, if one knows either the point-image characteristic or the sine-wave response, he can calculate the other, as has been shown. It would probably be preferable to have a research instrument that would determine the sine-wave response, because the transformation to the point-image characteristic is more direct and more suitable for calculating machines than the reverse transformation.

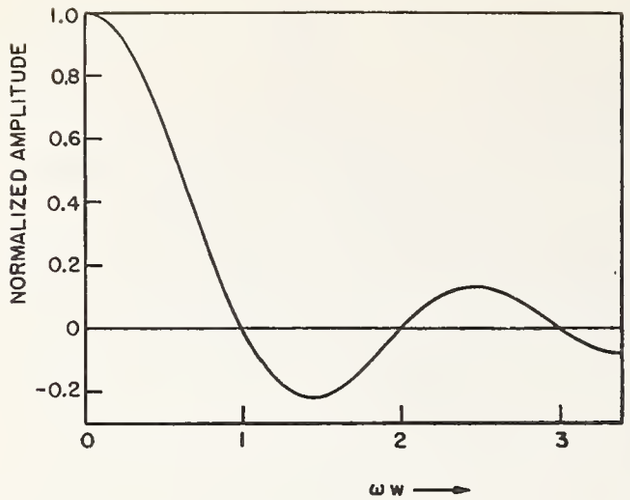


FIGURE 21. *Transform of finite-width line object.*

ω is the frequency of the component wave, and w is the reduced object-line width.

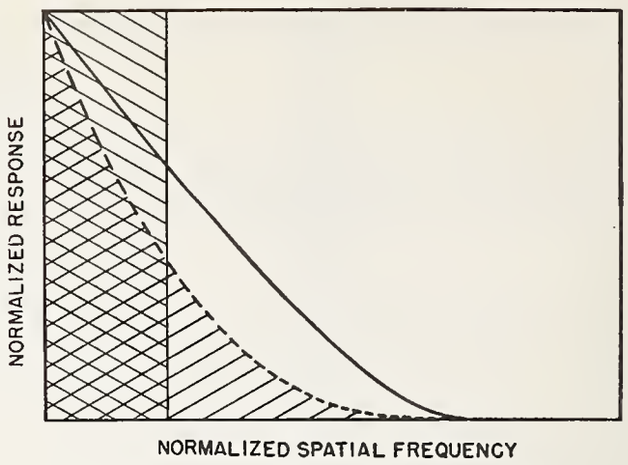


FIGURE 22. *Schade's equivalent bandwidth.*

The solid curve is the sine-wave response curve; the dashed curve is the square of the solid curve; and the rectangular area is equal to the area under the dashed curve. This rectangular area is measured by a single number, its limiting frequency, thereby providing a single number to describe the sine-wave-response curve.

Another problem that comes up is one that is involved in routine testing. It would be desirable to reduce the evaluating curve to a single number with as little loss of significant information as possible, and it would be desirable that this reduction be done automatically in the test procedure.

With respect to the sine-wave response curve, Schade has suggested a reduction in which the curve is squared, ordinate by ordinate, and then integrated. The resulting number is equal to the limiting frequency of a rectangular "response" curve having the same area as the squared sine-wave response curve. This establishes an equivalent bandwidth, shown in figure 22.

The mechanism that would permit this determination directly would be similar to the sine-wave response mechanism, except that a "noise" pattern instead of a sinusoidal pattern would be used. The ideal noise pattern contains all frequencies at equal amplitude but with random phase relations. The fluctuations in the photoelectric output produced by this pattern are fed through a squaring circuit and then integrated. The resulting steady current, indicated on a meter, is proportional to Schade's equivalent bandwidth. One trouble with this system is the difficulty involved in producing an acceptable noise pattern.

Equation (34) indicates another approach to the problem of representing the sine-wave response curve with a single number. The object transform T_0 in this equation can be considered to be a weighting factor for the sine-wave response curve, and the integral to represent the equivalent bandwidth Ω. Then this equivalent bandwidth can be determined by the use of a variable slit or a knife edge, for

$$\Omega=\frac{K(w)}{2w}. \tag{35}$$

For this to be single-valued, w must be fixed, and two convenient possible values appear evident. One is to set $w=1/\omega_c$, where ω_c is the limiting frequency for a theoretically perfect lens having the same aperture as the lens under test. Then the weighting function goes to its first zero at ω_c. This method is illustrated in figure 23.

The other convenient value for w is zero, for the limit of $K(w)/w$ as w goes to zero is the slope at the center of the knife-edge image, as can be seen with the aid of figure 14. As indicated in eq (35), this equivalent bandwidth is given by one-half the slope at the center of the knife-edge image. The weighting function approaches unity for small values of ω as w approaches zero, so this equivalent bandwidth is given by the area under the sine-wave response curve itself. This is illustrated in figure 24. It might be emphasized that this latter relationship is an important and fundamental condition. The slope at the center of a knife-edge image is exactly proportional to the area under the sine-wave response curve.

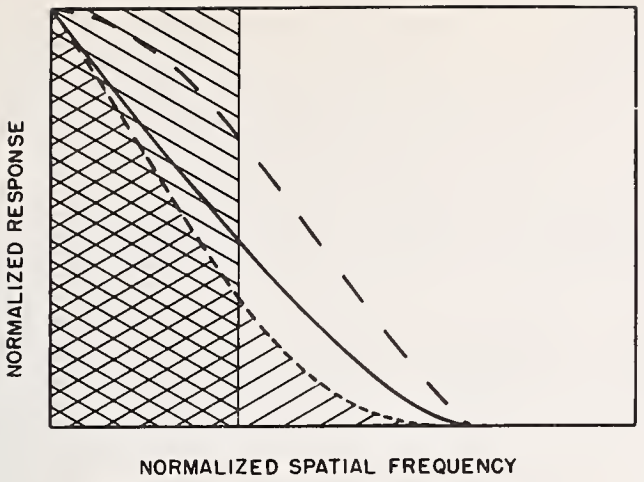


FIGURE 23. *Another equivalent bandwidth.*

The solid curve is the sine-wave response curve; the long-dashed curve is the transform of the finite-width-line object; the short-dashed curve is the product of the other two; and the rectangular area is equal to the area under the short-dashed curve. This provides a different single characteristic number, which can be obtained more easily than Schade's equivalent bandwidth.

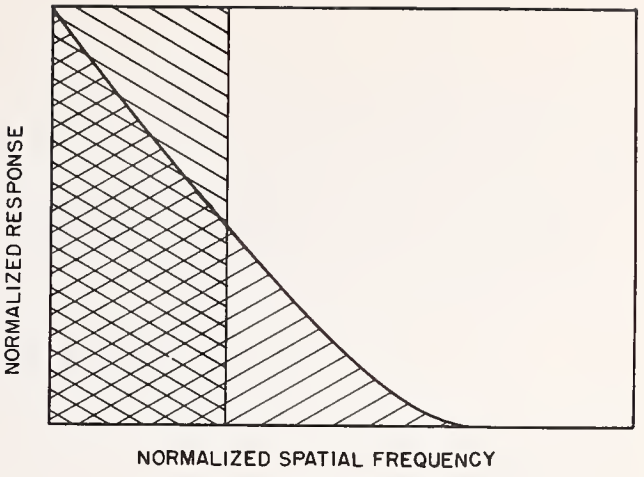


FIGURE 24. *A third equivalent bandwidth.*

This handwidth is measured by the area under the sine-wave-response curve itself. The area is exactly proportional to the slope at the center of a knife-edge image.

The idea of measuring an equivalent bandwidth may be a good way of reducing the sine-wave response curve to a single number, but this process eliminates one of the advantages of the sine-wave approach, and that is the ability to combine directly a sequence of imaging processes. For most telescopic systems this may be unimportant, but if such a combination is desirable, then perhaps the sine-wave response curve can be characterized well enough by some particular frequency. If the response to some such frequency were established as a measure of quality, then the measure of the quality of the combination is simply the product of the measures of quality of the components.

To summarize, the Fourier approach seems to be more desirable for research and detailed testing, but the determination of the point-image characteristic lends itself to rapid routine testing.

6. Appendix

6.1. Images Produced by an Aberration-Free System With a Circular Aperture in Monochromatic Light

The extended objects are assumed to be illuminated noncoherently. In the following expressions, the unit of displacement in the image is:

$$z=\frac{2\pi ax}{\lambda d}, \tag{36}$$

where a is the radius of the circular aperture, x is the distance in the image plane from the center of the image, d is the distance from the image plane to the aperture, and λ is the wavelength of light.

a. Point Image

This is well known, and its section is given by

$$\varphi(z)=4\left(\frac{J_1(z)}{z}\right)^2. \tag{37}$$

It is illustrated in figures 1 and 2.

b. Fine-Line Image

An expression for the cross section of the fine-line image, implicit in the original work of Struve [1], is explicitly given by Steel [2] as

$$L(z)=\frac{3\pi}{8}\frac{H_1(2z)}{(2z)^2}, \tag{38}$$

where H_1 is what is known as the Struve function, and has been tabulated.[3] The cross section of the fine-line image is illustrated in figure 8.

c. Knife-Edge Image

Struve [1] developed this in a series expansion and tabulated it for z up to 15. If the edge is oriented so that the gradient for the image is positive, then the image is given by

$$S(z)=\frac{1}{2}+\frac{2}{\pi^2}\left\{\frac{3}{1}\frac{2^2z}{1^2\cdot3^2}-\frac{5}{3}\frac{2^4z^3}{1^2\cdot3^2\cdot5^2}+\frac{7}{5}\frac{2^6z^5}{1^2\cdot3^2\cdot5^2\cdot7^2}-\cdots\right\}. \tag{39}$$

Struve also gave a simple approximation, which is in error by less than 0.1 percent for $Z>7$ and 1.0 percent for $Z>3$.

$$S(z)\cong1-\frac{2}{\pi^2z}. \tag{40}$$

The knife-edge image is illustrated in figure 12.

d. Finite-Width Line Image

If the width of the Gaussian image is w_z, then

$$\overline{L(z)}=S\left(z+\frac{w_z}{2}\right)-S\left(z-\frac{w_z}{2}\right). \tag{41}$$

This family of images is illustrated in figure 10.

e. Point-Image Characteristic

This is given by

$$K(w_z)=S\left(\frac{w_z}{2}\right)-S\left(-\frac{w_z}{2}\right)=2S\left(\frac{w_z}{2}\right)-1. \tag{42}$$

Combining this with eq (39) gives

$$K(w_z)=\frac{4}{\pi^2}\left\{\frac{3}{1}\frac{2w_z}{1^2\cdot3^2}-\frac{5}{3}\frac{(2w_z)^3}{1^2\cdot3^2\cdot5^2}+\frac{7}{5}\frac{(2w_z)^5}{1^2\cdot3^2\cdot5^2\cdot7^2}-\cdots\right\}. \tag{43}$$

This is illustrated in figure 6.

f. Sine-Wave Response

Steel [2] gives the following expression for the sine-wave response:

$$\Phi(\omega)\begin{cases}=\frac{2}{\pi}\left(\arccos\frac{\omega}{2}-\frac{\omega}{2}\sqrt{1-\frac{\omega^2}{4}}\right), & 0<\omega<2\\ =0, & \omega>2,\end{cases} \tag{44}$$

where $\omega=\pi/w_z$, and w_z is the wavelength of the sinusoid. Notice that there is an absolute limit to the frequency of the pattern that a lens can form, this limit being where the wavelength of the pattern in z units is equal to π.

The frequency, ν, in the image plane is given by

$$\nu=\frac{\omega}{2N\lambda}, \tag{45}$$

where $N=d/2a$ and λ is the wavelength of light. If λ is in millimeters, ν is in cycles per millimeter.

Figures 16 and 17 show the sine-wave response.

---

[3] E. Jahnke and F. Emde, Tables of functions with formulas and curves, 4th ed. (Dover Publications, New York, N. Y., 1945).

7. References

The data for the curves in the illustrations were calculated from the information contained in the following references:

[1] H. Struve, Beitrag zur theorie der diffraction an fernrohren, Ann. Physik Chemie **17,** 1008 (1882).

[2] W. H. Steel, Étude des effets combinés des aberrations et d'une obturation centrale de la pupille sur le contraste des images optiques. Premiére Partie, Rev. opt. **32,** 4 (1953).

Steel provides information on the point, disk, fine-line, knife-edge, sine wave, and square wave images. Material on the sine wave or Fourier transform approach is contained in the following references.

[3] O. H. Schade, A new system of measuring the specifying image definition, NBS Circular 526 (April 29, 1954).

[4] O. H. Schade, Image gradation, graininess, and sharpness in television and motion picture systems. Part II, J. Soc. Motion Picture Television Engrs. **58,** 181 (1952).

[5] P. Elias, D. S. Grey, and D. Z. Robinson, Fourier treatment of optical processes, J. Opt. Soc. Am. **42,** 127 (1952).

[6] E. W. H. Selwyn, The photographic and visual resolving power of lenses. I. Visual resolving power, Phot. J. **88B,** 6 (1948).

Schade is the first to build a practical machine for obtaining a sine-wave analysis of an optical system. Selwyn applies the Fourier treatment to the prediction of resolving power.

Work on the photoelectric examination of point images through circular apertures is reported in the following paper:

[7] R. E. Hopkins, Measurements of the energy distribution of optical images, NBS Circular 526 (April 29, 1954).

Interest in the knife edge is shown in the following references:

[8] W. Weinstein, Criteria of image-forming quality in photographic objectives, Phot. J. **91B,** 138 (1951).

[9] R. V. Shack, A proposed approach to image evaluation, NBS Circular 526 (April 29, 1954).

Shack also discusses the variable slit. Hopkins is now using knife edges in his work.

Washington, June 1, 1955.

Reprinted from JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, Vol. 46, No. 9, 755–757, September, 1956
Printed in U. S. A.

Outline of Practical Characteristics of an Image-Forming System*

ROLAND V. SHACK
National Bureau of Standards, Washington, D. C.

(Received September 6, 1955)

(*Presented at the Symposium on Formation and Evaluation of Images, University of Rochester, June 15–18, 1955.*)

The relationships between certain fundamental characteristic functions relating to an optical image-forming system are indicated, some being observable experimentally and others being unobservable. Of the observable characteristics two are especially useful in the analysis of the image-forming properties of the system. One of these represents the image of a point object and the other the Fourier transform of this point image. The two are considered for their practicality as bases for objective laboratory techniques in image evaluation. The transform of the point image is practical as it stands whereas the point image itself is not. A function directly related to the point image and to a number of other simple objects is suggested as a practical representative function. The practical functions are correlated and the relative merits of the two approaches are considered.

1. INTRODUCTION

DURING the past few years the National Bureau of Standards, as well as a number of other laboratories, has been concerned with the problem of developing an objective procedure for evaluating the quality of images formed by optical instruments. One general approach has been to try to find some physical parameter which can be measured objectively and then to try to correlate this parameter with some already existing quality criterion. Although it is true that, if successful, such an empirical approach will eliminate the subjective element in the determination of image quality, it is made under the assumption that the chosen criterion is in itself a satisfactory index of image quality. There is reason to doubt the validity of this assumption.

A more fundamental approach is to analyze the image-forming process as a phenomenon, the aim being to characterize the process in as general and yet informative way as possible, consistent with practical instrumentation. New criteria of image quality would of course be expected to be developed. The present paper is intended to be an outline of such an approach.†

2. OBSERVABLE AND UNOBSERVABLE CHARACTERISTICS

Figure 1 illustrates the relationships existing between some of the more significant or characteristic functions of an optical image-forming system. The actual functions are functions of two spatial dimensions, but to avoid complexity in the figures, they are drawn as functions of one dimension. These curves can be considered to represent cross sections of the more complete functions.

Perhaps the most elementary characteristic function is that represented by Fig. 1(a). This curve represents the variation in optical path length between an object point and its Gaussian image point plotted across the pupil of the system. It can also be interpreted as the wave-front deformation across the pupil.

From this function can be derived the complex amplitude of the wave function in the pupil [Fig. 1(b)], another characteristic function. The Fourier transform of this complex amplitude in the pupil gives the complex amplitude of the wave function in the image plane [Fig. 1(c)]. The product of this third characteristic function with its complex conjugate yields a function which has been traditionally called the intensity distribution in the image of a point [Fig. 1(f)].

Of the functions so far considered, only the last is an observable characteristic. That is, it can be directly ascertained by experimental means. The others cannot, and must be considered to be unobservable characteristics.

However, the point image is not the only observable characteristic. The product of the complex amplitude in the pupil [Fig. 1(b)] with its complex conjugate

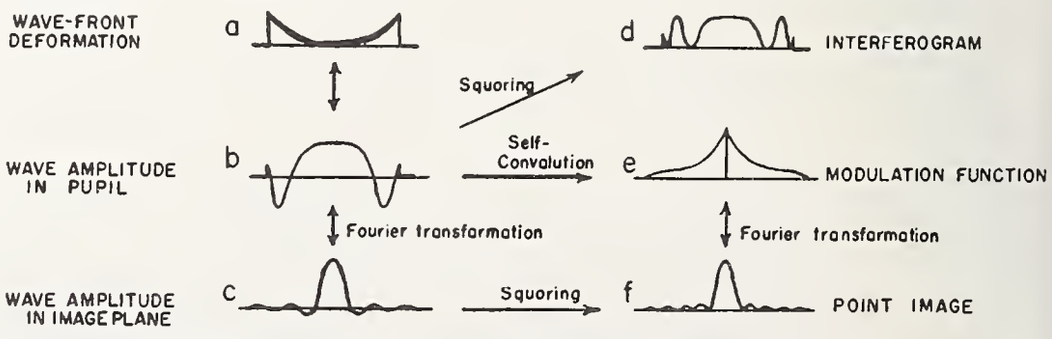


FIG. 1. Characteristic functions of an optical image-forming system.

* This work was done in connection with a project sponsored by Army Ordnance, Frankford Arsenal.

† Most of the material in this paper is somewhat more thoroughly dealt with in another paper by the author, "Characteristics of An Image-Forming System," to be published in the *Journal of Research of the National Bureau of Standards.*

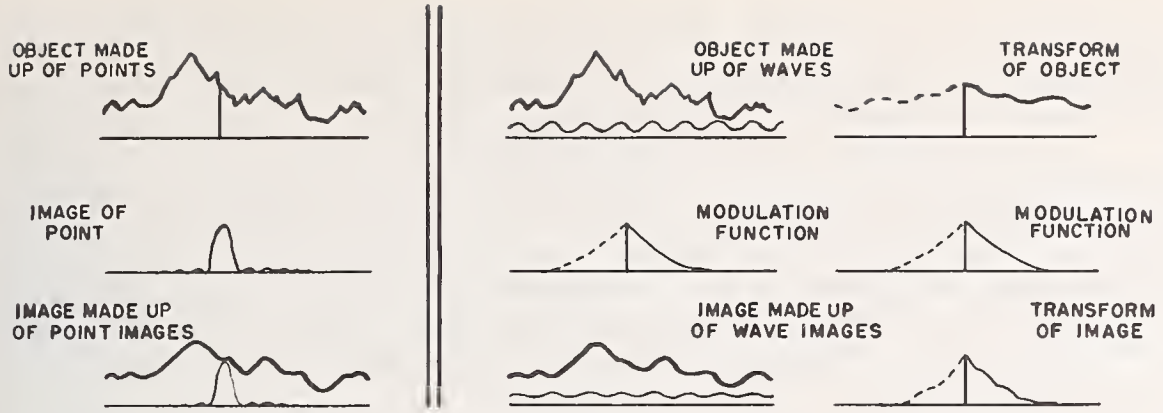


FIG. 2. Object-image relationships.

yields the function represented by the curve of Fig. 1(d), the kind of curve obtainable by means of an interferometer. This last function is also observable, and from it can be inferred the nature of the wave-front deformation [Fig. 1(a)], although not unambiguously.

A third observable function, only recently investigated, is that represented by Fig. 1(e). This is the Fourier transform of the point image [Fig. 1(f)]. It can also be obtained by convolving the complex amplitude in the pupil [Fig. 1(b)] with its complex conjugate.

We therefore have six interrelated functions characteristic of an optical image-forming system. Three of these (on the left of Fig. 1) are unobservable, cannot be directly obtained experimentally. The other three are observable. The interferogram [Fig. 1(d)] is valuable primarily because of its indication of the nature of the wave front in the pupil, but it is very difficult to infer from it the manner in which an image will be formed by the system. However, both of the remaining observable functions are concerned with object-image relationships.‡

3. OBJECT-IMAGE RELATIONSHIPS

That both are equally applicable to the problem of relating the image to the object is indicated in Fig. 2. Any object can be looked at as a summation of points and its image as the summation of the corresponding point images. The point image function is the characteristic modifying function.

On the other hand, the object can be considered to be the summation of a set of sinusoidal waves distributed in the object plane. These component waves, differing from each other in amplitude, frequency, phase, and orientation, are spatially distributed waves. That is, they are spatial, not temporal, sinusoidal variations in brightness throughout the object plane. The image consists of the summation of the images of these component waves.

Now an image-forming system, if linear, does not affect the frequencies, orientations, or sinusoidal character of the component waves. It can only affect their amplitudes and phase relationships. The function which describes these modifications as a function of the spatial frequencies and orientations of the component waves is that represented by the curve in Fig. 1(e), the Fourier transform of the point image. This function has many names in the literature, e.g., sine-wave response, system function, modulation function, transmission factor, contrast reduction factor, etc.

It might be noted, as indicated in Fig. 2, that, since an object can be considered to be made up of component sinusoidal waves, it can also be represented by a function in which the amplitudes and phase relationships of the component waves are plotted against their frequencies and orientations. This function is the Fourier transform of the object function. Multiplying this by the modulation function yields the Fourier transform of the image. An inverse transformation will yield the image function itself.

Now an image-forming process is useless unless the image is picked up by a detector, which may be the eye, a photographic emulsion, or a photoelectric detector. The detector itself modifies the image and must be considered to be a secondary image-forming process. Its inclusion in the over-all process is easily done with the approach involving the modulation function because the modulation function of an entire system is the product of the modulation functions of the components.

The approach involving the point image, however, requires a convolution for every component image-forming system, rather than a simple product.

4. EXPERIMENTAL DETERMINATION

Although any given optical system may be intended for use with one type of detector (i.e., the eye, a photographic emulsion, or a photoelectric detector). it is not necessary that a laboratory performance test use the same detector, provided that the effect of the detector on the image is predictable. The choise of detector for laboratory use should be made on the basis of convenience, controllability, and objectivity of the data obtained. The photoelectric detector in most cases will best meet the requirements.

However, to avoid integrating the fine detail in the

‡ Strictly speaking, both are significant only when the objects with which the system is to be used can be considered to be noncoherent extended luminous sources.

image, the photoelectric detector must be used behind a small aperture, and provision must be made for producing a relative displacement between the aperture and the image in order to explore various parts of the image.

The required smallness of the detector aperture places a premium on the amount of energy available, but the development of the photomultiplier tube has made this technique feasible. The shape of the detector aperture can be rigidly controlled and the results obtained should be highly reproducible. However, the photomultiplier will generate noise which will interfere with weak signals.

It should be pointed out that the electrical signal obtained from the detector cannot distinguish between the effect of the object and the effect of the detector aperture. The test object and the detector aperture are interchangeable. For example, the signal obtained by using a luminous disk for a test object and a pinhole for the detector aperture will be the same as the signal obtained by using a pinhole for a test object and a circular hole for the detector aperture, provided that the Gaussian geometry is the same. Also there are a number of ways of producing the relative motion between the image and the detector aperture. The detector aperture can be moved across a stationary image, or the image can be moved across a stationary detector aperture by moving the object or by rotating or rocking some element in the optical train.

The modulation function of an optical system can be obtained experimentally by using an object whose transform is known and a detector aperture whose modulation function is known, and determining the transform of the output signal. The modulation function of the optical system is given by the ratio of the latter to the product of the former. There are a number of ways of doing this, but the maximum energy in the signal for any given frequency will be obtained by using a sine-wave or a square-wave pattern for one of the functions.

The point image is not so easily obtained. The effect of the detector aperture cannot be divided out as in the case of the modulation function, and so the detector aperture must be made as minute as possible. This reduces drastically the energy available. To make matters worse, a large part of the energy in the point image may be distributed in the flare about the central peak, the area of the flare region being so large that the intensity is very low at any given point. The effect of this flare is noticeable with extended objects, although it may seem to be negligible in the image of a point.

A test object which is more easily handled and which shows the effect of the flare is the well-known knife-edge. If the data obtained with a knife-edge object is recorded as a curve in which the intensity difference between two points located at equal distances on either side of the geometric image is plotted against the distance between the points, it can be shown that this curve can be obtained from or interpreted in terms of a variety of objects, namely, a point, a fine line, a variable slit, and, of course, a knife-edge. Such a curve would be a useful characteristic, related to the point image, for determining image quality, and in the rest of this paper will be called the point-image characteristic.

5. COORDINATION

If the modulation function is known, this point-image characteristic is easily obtained because the Fourier transform of a variable slit object is simple and well known. When multiplied by the modulation function it yields the transform of the image, and because only the value at the center of the image is needed, the desired value for a given width is obtained by a direct numerical or graphic integration of the product of the two curves, the oscillatory component of the inverse transformation dropping out.

An interesting result of this relationship is that the area under the curve representing the modulation function itself is exactly proportional to the slope at the center of the knife-edge image. This is also the Strehl definition for the image of a fine line.

The modulation function can also be obtained from the point-image characteristic. For a given frequency the response of the system to a square-wave bar pattern can be predicted from the point-image characteristic by a simple addition and subtraction of selected ordinates. Once a square-wave response curve is obtained, the sine-wave response curve (modulation function) can be obtained by subtracting out the harmonic components.

6. MERITS

The modulation function has a number of features to recommend it for image evaluation. It is a bounded function, relatively simple in form, relatively easy to measure, and easily manipulated. It allows the convenient analysis of cascaded systems. Its use opens the possibility of applying to optical image formation many of the concepts of communication theory and allied fields.

The point-image characteristic has the advantage of presenting graphically the image characteristics of a number of important types of objects. It is also easy to obtain. However, it is an indirect measure of the properties of a point image and is relatively difficult to manipulate.

Journal of Research of the National Bureau of Standards Vol. 61, No. 6, December 1958 Research Paper 2920

Evaluation of Lens Distortion by Visual and Photographic Methods[1]

Francis E. Washer, William P. Tayman, and Walter R. Darling

The evaluation of lens distortion by photographic and visual methods is discussed. Measurements made on a single lens using the two methods are reported. The precision of measurement of each method is determined which shows that the observed differences must be attributed to systematic error. Various sources of systematic error are considered. Uncompensated differential plate tipping is identified as the most probable cause of the observed differences. A method of correction is developed. It is concluded that when work is done with extreme care and with due account taken of various insidious sources of error, it is possible to achieve comparable results with either method.

1. Introduction

The measurement of radial distortion in the focal plane of photographic objectives has been the subject of intensive study since the advent of aerial mapping from photographs. This particular aberration is of prime interest as its magnitude determines the accuracy with which the final negative maintains the correct relationships among the array of point images making up the photograph of the corresponding array of points in the area photographed. It was early realized that the reliability of the quantitative information obtained from a photograph increased as the distortion in the taking lens decreased. Consequently, the development of improved lenses was encouraged with the result that succeeding series of new lenses were characterized by ever lower values of distortion.

During this period of change, diverse methods of evaluating the distortion of lenses came into use at various laboratories. The reason for such diversity was primarily the availability of given types of measuring instruments in various laboratories. There are now three principal methods of measuring distortion plus numerous additional methods that are either the inverse of one of the principal methods or a variation thereof. The nodal slide bench is one of the oldest methods; this is a visual method capable of high accuracy and is perhaps the most widely used. The photographic method is more recent and arose out of the desire to make measurements under conditions approximating that of use. The third principal method is the goniometric method which is used to considerable extent in Europe.

Because of the diversity of methods being used in the evaluation of distortion, it seems worthwhile to investigate the results of measurement made in a single laboratory on a number of lenses by a variety of methods to determine whether or not the values so obtained varied appreciably with method. This is being done for several different methods and some preliminary results are reported herein for two methods.

The two methods are as follows:

A. Photographic, precision lens testing camera,
B. Visual, nodal slide optical bench.

[1] This work was performed in connection with the research project sponsored by the U. S. Air Force.

2. Methods of Measurement

Methods A and B have been described at some length in the literature. However, for purposes of clarity, in the following section, a brief description of these methods is given.

2.1. Precision Lens Testing Camera. Method A

The precision lens testing camera [1][2] shown in figure 1 was developed at the Bureau by I. C. Gardner and F. A. Case. It is one of the earliest successful devices developed to measure the performance of lenses by photographic means. It consists of a bank of collimators spaced at 5-deg intervals having resolution test charts as reticles. The lens under test is mounted at the center of convergence of the collimator fan and can be aimed at any one of them by rotation of a carriage which carries the lens holder and camera back whereon the photographic recording plate is mounted. As presently constituted, the lens testing camera has 10 collimators covering a total angle of 45 deg. When used to test wide angle lenses, the camera is aimed at one of the extreme collimators (position I) and the test made. In order to cover a complete diameter, a second test is made with the camera aimed at the collimator at the opposite ex-

[2] Figures in brackets indicate the literature references at the end of this paper.

FIGURE 1. *Precision lens testing camera (Method A).*

This photograph shows the camera back with a light metering device in place on the plate holder. The values of illumination in the focal plane are read from the large meter above the collimator. Use of this device enables the operator to adjust for uniform exposure in making the test negatives from which the values of distortion, D_A, are obtained.

treme (position II). It has been found by experience that the results obtained from two negatives made in this manner are quite reliable.

To determine the value of the distortion, the best row of images is selected on the negative made in position I. The separation of all images on this row from the image formed by light from the 0° collimator is measured. The equivalent focal length (*EFL*) is then determined from the relations

$$f_1 = d_5 \cot \beta_5, \tag{1}$$

$$f_2 = d_{10} \cot \beta_{10}, \tag{2}$$

$$EFL = \frac{f_1 + f_2}{2}. \tag{3}$$

Where d_5 and d_{10} are the measured distances on the negative separating the images formed by the lens for the target in the 0° collimator from the corresponding images formed by the lens for the targets in the collimators located at $\beta = 5°$ and 10° respectively from the 0° collimator. One then determines the distances from the 0° image to the other images using the relation

$$d_\beta = EFL \tan \beta. \tag{4}$$

The difference between the distances found with the aid of eq (4) and the corresponding measured distance for a given value of β is the value of the distortion at that point. It is positive if the measured location of the image is farther from the central image than the computed location. Similar computations are made for the negative obtained for position II and the results averaged and accepted as final.

It has always been maintained that the probable error of values of distortion obtained by this method did not exceed ±0.020 mm. The sources of error arising from error in angle in plate measurement are discussed in an earlier publication [2] where it was established that errors from this source should not exceed ±0.013 mm.

In making the negatives, the collimator targets are illuminated by light from a tungsten source after passing through a K–3 filter. Eastman Kodak Spectroscopic plates, emulsion Type V–F, are used to record the image formed by the lens under test. The exposed plates are processed in trays containing Eastman Developer D–19 and 68° F for 3 min with continuous agitation.

2.2. Visual Optical Bench, Direct Nodal Slide. Method B

The visual optical bench has long been the basic tool for evaluating the optical constants of lenses. The one used at the Bureau has been in existence for approximately 30 yr and can still be regarded as a precision instrument [3]. For measuring distortion, one uses a collimator, nodal slide lens holder shown in figure 2, and micrometer microscope. The lens is carefully alined in the holder and the axial image formed by the lens under test of the illuminated

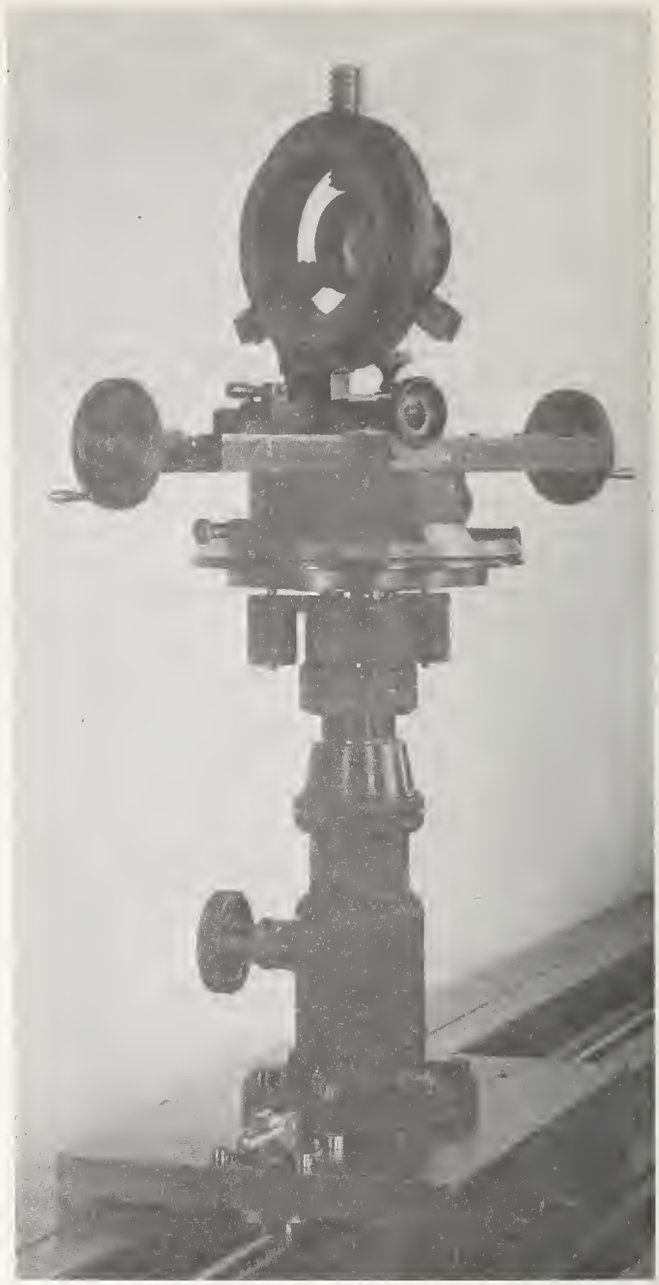


FIGURE 2. *Visual nodal slide (Method B).*
This photograph shows the nodal slide in which the lens under test is mounted.

reticle of the target is brought into coincidence with the object plane of the viewing microscope. The reticle is illuminated by filtered light from a tungsten source. The effective wavelength is approximately 575 mμ. By a series of successive adjustments, a condition is found for which a small rotation of the lens about a vertical axis does not produce a displacement of the axial image viewed. The rear nodal point of the lens is then considered to coincide with the center of vertical rotation of the nodal slide.

Assuming the equivalent focal length, f, to be known, the nodal slide is rotated by amount β about the vertical axis using the calibrated circle of the nodal slide to position it exactly. The entire saddle carrying the nodal slide and lens is then moved away from the microscope toward the collimator by an amount, $f(\sec \beta - 1)$. The viewing microscope is shifted laterally to the new position of the image

and its dial read and recorded as reading R. The nodal slide is then rotated to position $-\beta$ and a second setting of the microscope, L, is made. The distortion, D_β, is then obtained from the relation

$$D_\beta = \frac{(R-L)}{2} \sec \beta. \quad (5)$$

3. Results of Measurement

At the time this study was initiated, it was planned to make measurements of distortion on a series of seven or more lenses by different methods and to compare the results. Measurements have been made on seven lenses by two different methods and comparative tables of results prepared. All values of distortion are referred to the calibrated focal length [4] for ease of comparison. In general the results are comparable and all values fall within the range of ± 0.02 mm which has been established here as the accepted maximum value of the probable error for distortion measurements. However, the agreement fell somewhat short of that expected in that systematic differences were observed. Because of these systematic differences, it was deemed wise to concentrate on the analysis of the results of measurement made on a single lens by the two methods, in the belief that more information on the causes of lack of agreement could be gained thereby. This has proved to be the case, and in the following pages some of the steps in this investigation and analysis are reported.

3.1. Values of the Distortion by Methods A and B for Wide Angle Lens No. 3

The values of the distortion referred to the calibrated focal length obtained photographically with the precision lens testing camera (D_A) and visually with the nodal slide optical bench (D_B) are shown in table 1. These values are based on a single run by each method. For method A, a single run provides two negatives, one each for positions I and II; each negative is measured 5 times and the results are averaged. For method B, a single run consists of two independent sets of measurements; the accepted values of distortion are based on the average of two sets. It is clear from table 1, that the values of D_A and D_B do not depart from their average by more than 12 μ at most and accordingly either set of values can be regarded as accurate within the usual tolerance of ± 20 μ. However, the differences in the values obtained by the two methods $D_B - D_A$ which are also shown in table 1 do not appear to be random but rather appear to indicate the presence of systematic error. This is shown more strikingly in figure 3 where $D_B - D_A$ is plotted as a function of β.

TABLE 1. *Measured values of the distortion versus angular separation β from the axis for wide angle lens No. 3*

The values obtained with the precision lens testing camera are designated D_A while those obtained on the nodal slide optical bench are designated D_B. All values are referred to the calibrated focal length and are given in microns.

β	D_A	D_B	$D_B - D_A$
deg			
0	0	0	0
5	−11	−6	5
10	−18	−8	10
15	−17	1	18
20	−2	19	21
25	32	50	18
30	65	89	24
35	97	104	7
40	76	80	4
45	−97	−104	−7

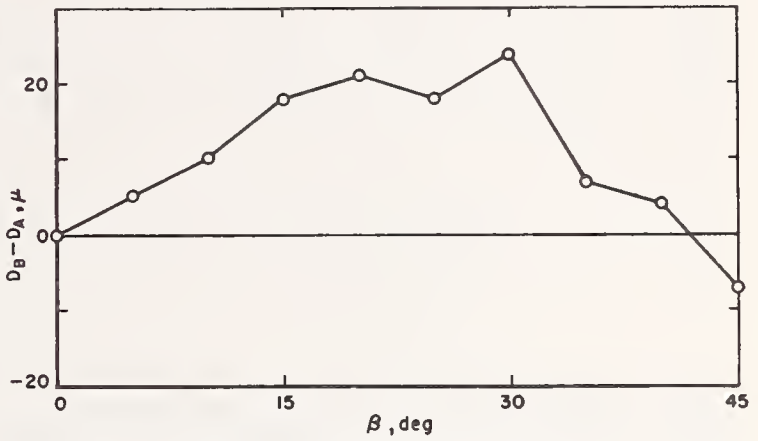


FIGURE 3. *Variation of distortion difference, $D_B - D_A$, with angular separation from the axis β.*

3.2. Precision of Measurement for Visual Nodal Slide

One's first thought when confronted with discrepancies in the results of measurement obtained by two different methods is to make additional measurements using the method most likely to be suspected. The visual nodal slide optical bench in use at the time of these measurements was somewhat antiquated so it seemed proper to check its performance first. Five additional independent sets of measurements were made. A pinhole reticle was used initially, but a transparent crossline was substituted for it during two sets of measurements. Some time elapsed between the first two sets of measurements and the 3d, 4th, 5th, and 6th calibration. During this time, the lens cells were removed from the test barrel and replaced. A similar operation occurred between calibrations 6 and 7. Additional care was taken in checking alinement in the course of calibration 3 and 7. Each set of data was carefully processed and the values of distortion adjusted to a calibrated focal length with maximum plus at 35° and maximum minus at 45°. The results are shown in table 2. Cursory examination of these values indicates quite good agreement, so an average for all seven including the first two calibrations was made. The table also shows the probable error of a single determination, and the value of PE_s. The values of PE_s are so low that it seems there is no cause to question the precision of method B. While there may be a systematic error present that impairs the accuracy, none comes to mind at the present writing.

TABLE 2. *Values of the distortion, D_B, versus β for seven sets of measurement using method B*

These values of D_B were obtained for wide angle lens No. 3 with the visual nodal slide bench. All values are referred to the calibrated focal length and are given in microns. The values of PE_s are also shown.

β	Values of D_B obtained from calibration							Average value of D_B	PE_s
	1	2	3	4	5	6	7		
deg									
0	0	0	0	0	0	0	0	0	±0
5	−6	−6	−7	−7	−7	−8	−8	−7	1
10	−7	−8	−9	−8	−6	−6	−11	−8	1
15	1	1	1	4	3	3	−7	1	2
20	19	19	21	24	23	27	13	21	3
25	48	52	52	53	60	53	45	52	3
30	88	89	85	88	85	87	84	87	1
35	103	104	105	107	104	105	111	106	2
40	83	78	86	82	80	69	77	79	4
45	−103	−104	−105	−107	−104	−105	−111	−106	2

3.3. Precision of Measurement for Lens Testing Camera

a. First Trial

Comparison of the average values of distortion based on the seven separate determinations shown in table 2 with the values of D_B for the first two determinations shown in table 1 indicates that the difference between D_B and D_A can not be explained by lack of precision in method B. The average value of D_B for the seven calibrations differ at most by 2 μ from the average of the first two calibrations. Accordingly further tests by method A were made. Five new negatives were made in position I on the precision lens testing camera and five new negatives were made in position II. Test negatives 1 to 4 were made under identical conditions in position I so also were test negatives 5 to 8 in position II. The lens was removed, replaced, and realined for test before making negatives 9 and 10. Each negative was measured and the distortion evaluated separately and the results are shown in table 3. Part (a) of table 3 shows the results for position I, and part (b) shows the results for position II. In part (c), the results shown in parts (a) and (b) are combined by pairs and results of this combination are shown. The values of the precision index, PE_s, are somewhat higher in parts (a) and (b) than those shown in part (c) where the results are averaged in the usual manner. The differences in the average value of D_A for positions I and II are not regarded as serious as it is presumed to arise from a small amount of plate tipping of a nature that becomes negligible on averaging. The final values of the average, D_A, shown in part (c) of table 3, differ by a few microns from the values of D_A shown in table 1, but have not changed in a manner that appreciably reduces the values of $D_B - D_A$. The values of the precision index PE_s shown in part (c) compare favorably with the similar values shown in table 2 for method B. However, the magnitudes of PE_s for both methods are too small to justify the existence of differences in the values of the distortion as great as are found.

TABLE 3. *Values of the distortion, D_A, versus β for the first trial using method A*

These values of distortion were obtained for Wide Angle Lens No. 3 with the precision lens testing camera. All values are referred to the calibrated focal length and are given in microns. In part (a) values are given for position I for five negatives, together with the average value of D_A and the value of PE_s. In part (b), similar results are given for position II, and in part (c) average values for paired negatives for position I and II are given.

β	(a) Values of D_A obtained in position I from negatives					Average value of D_A	PE_s
	1	2	3	4	10		
deg							
0	0	0	0	0	0	0	±0
5	−13	−13	−13	−11	−9	−12	1
10	−21	−18	−22	−18	−15	−19	2
15	−16	−12	−15	−9	−5	−11	3
20	2	1	1	3	7	3	2
25	20	24	28	26	32	26	3
30	53	60	63	58	64	60	3
35	87	91	96	86	83	89	4
40	81	75	84	77	83	80	3
45	−87	−91	−96	−86	−83	−89	4

	(b) Values of D_A obtained in position II for negatives					Average value of D_A	PE_s
	5	6	7	8	9		
deg							
0	0	0	0	0	0	0	±0
5	−10	−12	−16	−12	−12	−12	1
10	−15	−22	−22	−19	−18	−19	2
15	−10	−23	−20	−16	−18	−17	3
20	6	−9	−6	−4	2	−2	5
25	34	22	22	22	33	26	5
30	72	50	63	56	75	63	8
35	108	102	100	101	117	106	5
40	88	86	82	84	95	87	3
45	−107	−103	−101	−100	−117	−106	5

	(c) Values of D_A obtained in position I and II for negative pairs					Average value of D_A	PE_s
	1 and 5	2 and 6	3 and 7	4 and 8	10 and 9		
deg							
0	0	0	0	0	0	0	±0
5	−12	−12	−14	−12	−10	−12	1
10	−18	−20	−22	−18	−16	−19	2
15	−13	−18	−18	−12	−12	−15	3
20	4	−4	−4	0	4	0	3
25	27	23	25	24	32	26	2
30	62	55	63	57	68	61	4
35	98	96	98	94	100	97	2
40	84	80	83	80	89	83	2
45	−97	−97	−98	−93	−100	−97	2

b. Second Trial

While the first values of the distortion obtained in section 3.3a compare favorably with the values shown in table 1 for method A, it is disquieting to see the large discrepancies that exist between values obtained in positions I and II. While it is probable that these discrepancies arise from actual differences in performance along the opposing radii, nonetheless it seemed worthwhile to determine whether or not any maladjustment of the lens testing camera could produce this effect. Careful analysis showed that when the camera was properly alined

so as to point properly at the center of the collimator reticle for position I that it did not so point when swung into position II. Instead there was an error of 36 sec in the pointing for position II. Computation indicated that this defect in alinement in position II could produce small errors in the distortion although the indicated magnitude was too small to change the values of D_A by more than a few microns. This alinement error was corrected, the angles were recalibrated, 10 additional negatives were made, and the measurements described in section 3.3a were repeated. The new results are shown in table 4.

TABLE 4. *Values of the distortion, D_A, versus β for the second trial using method A*

These values of the distortion were obtained for Wide Angle Lens No. 3 with precision lens testing camera following its recalibration. All values are referred to the calibrated focal length and are given in microns. In part (a) values are given for position 1 for five negatives together with the average value of D_A and the value of PE_s. In part (b) similar results are given for position 11, and in part (c), averaged values for paired negatives for positions 1 and 11 are given.

β	(a) Values of D_A obtained in position 1 for negatives					Average value of D_A	PE_s
	11	12	13	14	20		
deg							
0	0	0	0	0	0	0	±0
5	−10	−12	−12	−16	−15	−13	2
10	−18	−23	−21	−19	−19	−20	2
15	−12	−20	−18	−15	−16	−16	2
20	1	−9	−7	0	−2	−3	3
25	27	16	19	25	22	22	3
30	64	57	57	68	62	62	3
35	95	91	94	94	98	94	2
40	86	82	86	83	87	85	2
45	−95	−91	−94	−94	−98	−94	2

β	(b) Values of D_A obtained in position II for negatives					Average value of D_A	PE_s
	15	16	17	18	19		
deg							
0	0	0	0	0	0	0	±0
5	−19	−19	−19	−19	−17	−19	0
10	−28	−26	−25	−26	−26	−26	1
15	−27	−23	−23	−21	−23	−23	1
20	−11	−8	−9	−5	−9	−8	2
25	17	18	21	25	21	20	2
30	58	56	55	62	54	57	2
35	99	102	100	104	96	100	2
40	77	80	84	85	78	81	3
45	−99	−102	−100	−104	−96	−100	2

β	(c) Values of D_A obtained in positions 1 and 11 for negative pairs					Average values of D_A	PE_s
	11 and 15	12 and 16	13 and 17	14 and 18	20 and 19		
deg							
0	0	0	0	0	0	0	±0
5	−14	−16	−16	−18	−16	−16	1
10	−23	−24	−23	−22	−22	−23	1
15	−20	−22	−20	−18	−20	−20	1
20	−5	−8	−8	−2	−6	−6	2
25	22	17	20	25	22	21	2
30	61	56	56	65	58	59	3
35	97	96	97	99	97	97	1
40	82	81	85	84	82	83	1
45	−97	−96	−97	−99	−97	−97	1

When the final results obtained from negatives 11 to 20 are compared with those obtained from negatives 1 to 10, it is clear that while there is slightly better agreement at the large angles than before, the agreement is somewhat worsened at the small angles. Finally, the average values for the second trial, shown in part (c) of table 4, are substantially the same as those obtained from the first trial, shown in table 3.

C. Summation

The final accepted values obtained by methods A and B are brought together for comparison in table 5. The values of D_A and D_B differ slightly from the values shown in table 1 but are believed to be more reliable as each value is the average of many more determinations. It is noteworthy that the magnitude of ΔD has increased slightly. This table also shows the probable error of the mean for the values of D_A, D_B, and ΔD. Consideration of these various values indicates that the systematic error still exists and has not been reduced by the multiplication of measurements.

TABLE 5. *Comparison of the average values of distortion, D_A, derived from 20 negatives using method A and the average values of distortion, D_B, derived from sevenc alibrations using method B.*

The difference is shown as $\Delta D = D_B - D_A$. Values of the probable error of the mean for each set of determinations are also shown. All values are expressed in microns.

β	D_A	D_B	$\Delta D = D_B - D_A$	PE_m for		
				D_A	D_B	ΔD
0	0	0	0	±0.0	±0.0	±0.0
5	−14	−7	7	.6	.4	.7
10	−21	−8	13	.6	.4	.7
15	−17	1	18	.9	.8	1.2
20	−3	21	24	1.0	1.1	1.5
25	24	52	28	0.8	1.1	1.4
30	60	87	27	1.0	0.4	1.1
35	97	106	9	0.3	.8	0.9
40	83	79	−4	.6	1.5	1.6
45	−97	−106	−9	.3	0.8	0.9

3.4. Effect of Plate Curvature

In view of the known effects of plate curvature on the values of distortion, [2] it seemed worthwhile to examine the emulsion surfaces of the negatives used in method A. This was done and a small amount of plate curvature was found. However, in no instance was curvature present in sufficient amount to produce more than one-fourth of the measured differences in distortion values. The average departure from flatness for the 20 negatives could not produce differences in distortion in excess of one-tenth of that found. It may therefore be stated that plate curvature is not a prime cause of the differences in values of distortion found by the two methods.

3.5. Effect of Plate Tipping

The plate holder in the precision lens testing camera is so constructed that the emulsion surfaces are coplanar for positions I and II. For the coplanar

518-743 O - 73 - 35

or parallel plane condition, small departures of the plane from true normality to the optical axis of the system under test would not produce variations in the average value of distortion even though the measured distortion on either side of center would be different. However, if a slight warpage of the holder has occurred, the plate in position II would not be coplanar with that in position I and a small amount of asymmetric distortion would persist and adversely affect the final average. Accordingly a procedure was developed for checking this possibility based on the plate tipping analysis described in an earlier paper [5]. The final product of this analysis is shown in table 6. When the values of ΔD in table 5 are divided by the appropriate multiplier M, then the magnitude of $f\epsilon_t$ can be obtained for each value of ΔD, where ϵ_t is the angle between the plane of the plate in position I and that in position II. This was done and the results are shown in table 7.

TABLE 6. *Values of $\tan^2\beta$ and the multiplier, M, for a value of $f\epsilon_t = 1.00$ micron*

The multiplier, M, is so calculated that when it is multiplied by a given value of $f\epsilon_t$, it yields the distortion correction referred to the calibrated focal length.

Angular separation, β, from axis	$\tan^2\beta$	M
deg		
0	0.0000	0.0000
5	.0077	−.0690
10	.0311	−.1234
15	.0718	−.1630
20	.1325	−.1865
25	.2174	−.1913
30	.3334	−.1727
35	.4903	−.1234
40	.7041	−.0314
45	1.0000	−.1235

TABLE 7. *Computation of $\overline{f\epsilon_t}$ to determine effective amount of uncompensated plate tipping that may be present in method A*

Values of $f\epsilon_t$ are obtained for each value of ΔD given in table 5; the average $\overline{f\epsilon_t}$ for β ranging from 10° to 30° is −129 microns; this value is used to determine the probable contribution to distortion $M\overline{f\epsilon_t}$. The last column shows the magnitude of difference that remains. All values are given in microns.

β	ΔD	$\overline{f\epsilon_t} = D/M$	$M\overline{f\epsilon_t}$	$\Delta D + M\overline{f\epsilon_t}$
deg				
0	0	-------	-------	-------
5	7	−101	−9	−2
10	13	−105	−16	−3
15	18	−110	−21	−3
20	24	−129	−24	0
25	28	−146	−25	3
30	27	−156	−22	5
35	9	−73	−16	−7
40	−4	127	4	0
45	−9	−73	16	7

The values of $f\epsilon_t$ that are shown in table 7 indicate that some effect of uncompensated plate tipping may be present in the values of distortion obtained by method A. All values of $f\epsilon_t$ are of the same sign and are of the same order of magnitude except that computed for $\beta = 40°$. The value of ΔD for this point is so small (4μ) that little weight can be given to the values of $f\epsilon_t$, derived from it. While there is appreciable variation in the remaining values of $f\epsilon_t$, it must be remembered that the warping that introduces this effect probably bends the plaet slightly as well as tipping it. An average $f\epsilon_t$, for the range 10° to 30° was determined and found to be $f\epsilon_t = -129$ microns which indicates that the effective angle between the plates in positions I and II is approximately 3.0 min of arc. The quantity $\overline{Mf\epsilon_t}$ was then evaluated for each value of β. The degree of compensation achieved by assumption of this amount of uncompensated plate tipping is shown graphically in figure 4. The solid curve shows the variation of the quantity $-\overline{Mf\epsilon_t}$ with β while the circles show the corresponding values of ΔD. The quantity, $D+\overline{Mf\epsilon_t}$, would equal zero if $f\epsilon_t$ was invariant for all values of β. However, the departure from zero is sufficiently small that it is evident that uncompensated differential plate tipping makes an important contribution to the systematic error whose cause is being sought.

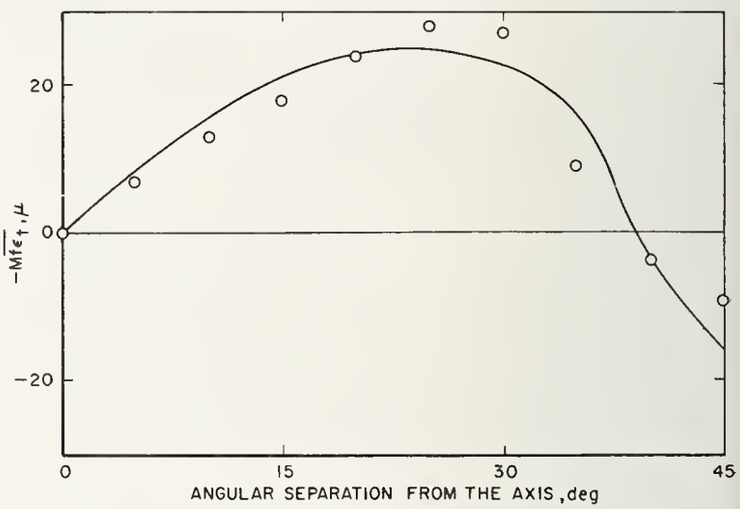


FIGURE 4. *Variation of the quantity $-M\overline{f\epsilon_t}$ with angular separation from the axis β.*

Compensation of the differences in distortion values ($\Delta D = D_B - D_A$) by an assumed differential plate tipping ϵ of approximately 3 min of arc. The solid line is a plot of the function, $-M\overline{f\epsilon_t}$; the circles are corresponding values of ΔD

Following this work careful measurements on the plate holder were made to determine if this warpage actually existed. While measurable values of warpage were found that were somewhat smaller than might be inferred from the results shown in table 7, they appeared large enough to warrant correction. Accordingly a new heavier plate holder has been constructed and has replaced the one in use during this experiment. Recent measurements show that the values of $\Delta D = D_B - D_A$ now obtained are less than one-fourth the values of ΔD obtained prior to the change.

In table 8, the correction $M\overline{f\epsilon_t}$ is applied directly to the values of distortion D_A. The corrected value $D_A' = D_A - M\overline{f\epsilon_t}$ is given in the table and agrees closely with the corresponding values of D_B. The last column shows the differences still remaining between values of D_A' and D_B. It is clear that the remaining discrepancies are so small in comparison with the values of D_A' and D_B that they can be neglected.

TABLE 8. *Comparison of D_A' (the value of D_A after correction for plate tipping) and D_B*

The values of D_A and D_B are taken from table 5; values of $Mf_{\epsilon t}$ are taken from table 7. The last column $\Delta D'$ shows that the discrepancies between the two methods have been reduced to tolerable values. All values are given in microns.

Angular separation from axis	D_A	$\overline{Mf_{\epsilon t}}$	$D_A'=D_A-\overline{Mf_{\epsilon t}}$	D_B	$\Delta D'=D_B-D_A'$
deg					
0	0	0	0	0	0
5	−14	−9	−5	−7	−2
10	−21	−16	−5	−8	−3
15	−17	−21	4	1	−3
20	−3	−24	21	21	0
25	24	−25	49	52	3
30	60	−22	82	87	5
35	97	−16	113	106	−7
40	83	4	79	79	0
45	−97	16	−113	−106	7

4. Conclusion

It is evident from the foregoing study that highly accurate values of distortion can be obtained by either the photographic method using the precision lens testing camera or the visual method using the nodal slide bench. For each method, values can be obtained that are precise to within ± 4 microns. When proper care is taken, the values obtained by either of the methods do not depart from the common average by amounts exceeding ± 5 microns.

It is however clear that to obtain and maintain such high accurarcy it is necessary to be constantly on the alert for various insidious sources of error such as plate curvature, differential plate tipping, and incorrect prime calibration of angles used in the determinations.

5. References

[1] I. C. Gardner and F. A. Case, Precision camera for testing lenses, J. Research NBS **18**, 449 (1937) RP984.

[2] F. E. Washer, Sources of error in various methods of airplane camera calibration, Photogrammetric Eng. XXII, p. 727 (1956).

[3] A. H. Bennett, The distortion of some typical photographic objestives, J. Opt. Soc. Am. and Rev. Sci. Instr. 14, 245 (1927).

[4] F. E. Washer and F. A. Case, Calibration of precision airplane mapping cameras, Photogrammetric Eng. XVI, 502 (1950); J. Research NBS **45**, 1 (1950) RP2108.

[5] F. E. Washer, Effect of camera tipping on the location of the principal point, J. Research NBS **57**, 31 (1956) RP2691.

WASHINGTON, April 12, 1958.

Reprinted from JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, Vol. 49, No. 4, 413, April, 1959
Printed in U. S. A.

Note on Measurement of Sine-Wave Response of Lenses

ROBERT E. STEPHENS
National Bureau of Standards, Washington 25, D. C.
(Received November 26, 1958)

ON reading of the ingenious method described by Lamberts[1] for evaluating both the magnitude and the phase shift of the sine-wave response of a lens it occurred to me that it should be possible to evaluate the sine of the phase angle as well as the cosine. It is of course desirable to be able to do this because for small angles up to 45° the sine is a more precise measure of an angle than the cosine. Considerable thought stimulated by the idea of measuring the difference $F(x)-F(-x)$, which of course does not work, led to the following.

The procedure of Lamberts, where two equal sine-wave object patterns are moved simultaneously in opposite directions to produce the scanning effect at the image slit, actually produces what amounts to a standing wave of luminous flux at the image (as collected by a long narrow slit). This wave can be represented by the sum of two equal traveling waves moving in opposite directions, thus;

$$F(x,t)=bo+\frac{b_1}{2}\cos\frac{2\pi}{\lambda}(x-\delta-vt)+\frac{b_1}{2}\cos\frac{2\pi}{\lambda}(x-\delta+vt).$$

By means of trigonometric identities this may be reduced to a form more convenient for our purpose, namely

$$F(x,t)=b_0+b_1\cos\frac{2\pi}{\lambda}(x-\delta)\cos\frac{2\pi}{\lambda}vt. \tag{1}$$

The variation resulting from the motions of the object patterns with the velocity v, is $\cos(2\pi/\lambda)vt$ and its amplitude is a function of position (x) in the image,

$$A(x)=b_1\cos\frac{2\pi}{\lambda}(x-\delta), \tag{2}$$

which varies from b_1 at $x=\delta$ to zero at $x=\delta\pm(\lambda/4)$. The term δ has been included so that x may be measured from an arbitrary origin (in lieu of carrying along a sine term).

On the basis of general mathematical principles we should expect to be able to evaluate b_1 and the phase angle $2\pi(x-0)/\lambda$ by evaluation of $A(x)$ at any two positions, $x=x_1$ and $x=x_1+\Delta x$, where Δx is a known quantity, and subsequent solution of simultaneous equations. However, solution is greatly simplified by choosing $\Delta x=\pm\lambda/4$.

Let the amplitude be measured at an arbitrary position of the slit designated by the coordinate $x=x_1$; it is

$$A(x_1)=b_1\cos\frac{2\pi}{\lambda}(x_1-\delta). \tag{3}$$

Then the receiving slit is displaced a distance $\lambda/4$ in either direction. The amplitude measured at the second position is

$$A\left(x_1\pm\frac{\lambda}{4}\right)=b_1\cos\frac{2\pi}{\lambda}\left(x_1-\delta\pm\frac{\lambda}{4}\right);$$

again applying trigonomentric identities this is converted to

$$A\left(x_1\pm\frac{\lambda}{4}\right)=\mp b_1\sin\frac{2\pi}{\lambda}(x_1-\delta). \tag{4}$$

Solving Eqs. (3) and (4) simultaneously we obtain

$$b_1=\left\{[A(x_1)]^2+\left[A\left(x_1\pm\frac{\lambda}{4}\right)\right]^2\right\}^{\frac{1}{2}}$$
$$\tan\frac{2\pi}{\lambda}(x_1-\delta)=\frac{A\left(x_1\pm\frac{\lambda}{4}\right)}{A(x_1)}. \tag{5}$$

The quantity b_1/b_0, called the "normalized amplitude" by Lamberts is also called the "modulation." The value of b_0 is easily measured as the average value of $F(x_1,t)$ [Eq. (1)]. The sine-wave response is the modulation in the image divided by the modulation in the object. The angle $2\pi(x_1-\delta)/\lambda$ is, of course, the phase angle φ for the slit position $x=x_1$. Of greater interest perhaps is the actual displacement of the pattern which is $\varphi\lambda/2\pi$.

This is as far as the discussion would need to go if it were practicable to determine the signs of $A(x_1)$ and $A[x_1\pm(\lambda/4)]$. However, measurements at the two positions ordinarily will give only the magnitudes of these amplitudes; consequently the sign of the phase angle is indeterminate. This problem can be resolved if the direction of the change in amplitude for small displacements of the slit can be determined. By differentiation of Eq. (2) the following is obtained:

$$\frac{dA}{dx}=\frac{-2\pi b_1}{\lambda}\sin\frac{2\pi}{\lambda}(x-\delta),$$

thus

$$\Delta A=\frac{-2\pi b_1}{\lambda}\sin\frac{2\pi}{\lambda}(x-\delta)\Delta x. \tag{6}$$

Referring to Eq. (2) again, b_1 and λ are positive quantities. Also, if the object patterns and slit are initially positioned as described by Lamberts the phase angle will be small for coarse patterns and the cosine will be positive. Thus $A(x_1)$ as well as $|A(x_1)|$ is positive. The sign of the sine thus determines the sign of the phase angle. Equation (6) shows that $\sin(2\pi/\lambda)(x_1-\delta)$ is negative if Δx and ΔA are both positive or both negative, and is positive if Δx and ΔA have different signs. On the assumption that the phase angle is a slowly and smoothly varying function of frequency (or wavelength) the progression of φ can be inferred if the steps in frequency are sufficiently small.

Equation (6) also shows that this test is insensitive if φ is small. This does not matter because one may ignore the phase shift until it can be measured with some accuracy at a higher frequency.

[1] Robert L. Lamberts, J. Opt. Soc. Am. 48, 490 (1958).

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 64C, No. 1, January–March 1960

Equipment and Method for Photoelectric Determination of Image Contrast Suitable for Using Square Wave Targets

Fred W. Rosberry

(September 1, 1959)

Conventional measurements of the resolving power of lenses employ measuring photographs of test charts containing an array of accurately spaced parallel lines. This method has limited precision because of the variability of photographic emulsions and is time-consuming in operation. This paper extends previous work by others in obviating these differences by using a direct photoelectric scanning of a line-pattern image formed by the lens under test. Square wave high contrast resolving power targets with two different line pattern arrangements were used as test objects. The image was moved across a stationary slit and photomultiplier tube. The output was recorded as relative transmission.

1. Introduction

The evaluation of the performance or efficiency of a lens involves, to some degree, the comparison of the output with the input. The respective outputs and inputs are essentially areas of information with input represented by the object or target and the output by the resultant image. A visual comparison of the two represents a typical subjective type of test. An effort was made in the construction of this equipment to minimize the subjective type of test and emphasize the objective. An objective approach includes making some sort of measurements in the object area in order that a numerical comparison might be made with similar measurements in the image. There are several quantities which can be measured in these areas to determine the preservation of information content. It is known that there is a relation between information content and resolving power, the latter of which relates to the smallest distinguishable separation of separate objects and is probably affected by the contrast of the adjacent objects to be distinguished.

Many types of line pattern charts have been developed for testing resolving power. These charts consist of line patterns of varying spacings which proceed in some form of an orderly progression. For example, the ratio of line widths existing between adjacent patterns may be $\sqrt{2}$, $\sqrt[3]{2}$, $\sqrt[4]{2}$, or $\sqrt[6]{2}$. In one chart developed at NBS the values of resolving power, indicated by successive line patterns, vary continuously in a linear manner [1].[1] Each of these charts has some desirable features. Charts having large steps between successive patterns can cover a longer range of values of resolving power with a relatively small target area as compared to charts of a linear nature or those having small steps between successive patterns. Coarse charts permit a higher degree of repeatability, but a lower degree of precision with respect to the actual limit of resolution [2].

[1] Figures in brackets indicate the literature references at the end of this paper.

The actual determination of the focal location of this resolution limit can be avoided and yet similar results produced by a method which is based on the measurement of contrast in each of the line-space patterns in the image as it approaches the limit. Earlier work along these lines consisted of recording the image on a photographic plate and then scanning or measuring the line-space density with a microdensitometer which recorded the measurements in the form of a curve on graph paper, from which density values could be read [3]. This procedure soon led to the discovery that different types of photographic plates gave different results. The slower, fine-grain emulsions gave higher limits of resolution than did the faster coarse-grain emulsions. Different development procedures applied to the same type of plate would also produce a variation in results. It naturally followed that a procedure to eliminate this photographic variable was necessary. A different type of receiver was needed to supplement the photographic emulsion.

The photoelectric tube had already proven itself a suitable and reliable receiver. The linearity of response for increasing quantities of light is superior in the phototube to that of a photographic emulsion. The work of Schade [4] demonstrated one method in which the photographic process could be eliminated completely: Scanning the image with a narrow slit followed by a phototube measuring the flux which entered the slit. Another modification incorporated in this method is the use of a target which has its line-space transmission varying along a curve which is essentially sine wave. Targets used previously were square wave in transmission with the dividing edge between line and space a sharp line. The Schade apparatus, on the other hand, uses a rapidly moving target causing a rapid sinusoidal fluctuation in intensity at the receiving slit. The phototube and accessory equipment received and converted this signal, producing on a meter or recorder, a measurement of the amplitude of the

a-c component. As the test signal emitted from the target is a sine wave and reaches the receiving slit in the same waveform, the effect of the lens under investigation in the system was only to reduce the amplitude of the fluctuations. A comparison of the amplitudes leaving the target with those leaving the lens was handled in a mathematical manner to produce a term called "sine wave response." This equipment is now at the National Bureau of Standards and has been used in a number of investigations. One objection to the system has been the short length of line which it uses. The length of the lines is approximately 2 mm in the target plane as they are located on the sound track of 35-mm film. Rather than try to make a sine wave target with greater line length, an effort has been made to utilize the square wave target and scan it at a slower speed in order to register each individual line and space. Square wave targets are readily available with line lengths of 8 in. and, by using a long target line and relatively short scanning slit, the end effects of the line image are screened out by the scanning slit.

In brief, a method has been developed here involving the use of a microphotometer to study the aerial image of long line target transparencies. This method and a few representative results are described in this paper.

2. Description of Apparatus

The direct scanning of the image with a phototube was accomplished by arranging the microphotometer as shown in figure 1. A schematic layout is shown in figure 2. The source was a ribbon-filament lamp operated at 6 v and 18 amp by a suitable transformer, with a control for varying lamp intensity. Nine inches in front of the lamp were located the filter holder and condenser system. The target was

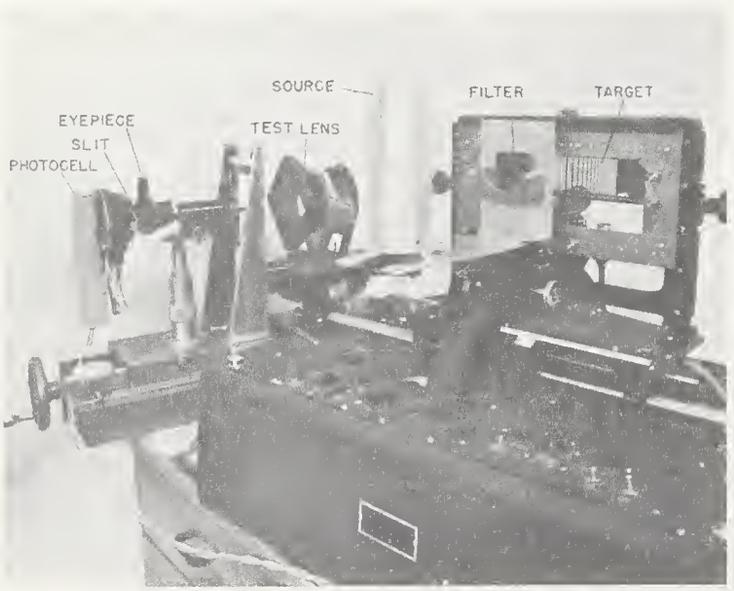


FIGURE 1. *Microdensitometer modified to accommodate lamp house and filter holder.*

The 3-line target is in place on the instrument cross slide. Test lens and receiving unit can be seen at left.

FIGURE 2. *Schematic layout of arrangement of equipment*

The arrow indicates the component in the system which was moved during a test.

mounted on the traveling carriage of the microphotometer and driven by a motor which was synchronized with the paper drive of the recorder. The plane in which the target moved was perpendicular to the optic axis of the test system. The target was so designed that, to cover the required line-space frequency range in the image plane, it was necessary that the target be reduced in size by a factor of 25, or that it be placed a distance of 26 focal lengths from the lens under investigation. The test lens having a focal length of 6 in. thus required a target-to-lens separation of 13 ft. Available space in this laboratory did not allow such a spread of equipment, so the system was folded almost back upon itself by the use of a high quality front-surfaced mirror. The next component along the optical axis was the lens under test, which in this investigation was a typical distortion-free wide angle mapping lens including anti-vignetting filter. A microscope was focused on the focal plane of this lens which reimaged the target at 10 × magnification on a slit. Immediately behind the slit, in a light-tight enclosure was a phototube, the output of which was amplified and used as the input to a recorder which traced the transmission of the target image lines and spaces on a chart.

2.1. Description of Targets

The two targets used in this work were of the square wave type made at this laboratory. Each is on a 2- by 10-in. photographic plate used as a transparency. The geometry of the scanning system requires that only a small portion of the target be illuminated at one time. The only requirement was that the image of the illuminating aperture be larger than the receiving slit and this was relatively easy to accomplish. Another advantage of this system was that the distribution of light flux need not be exactly uniform over the illuminating aperture, as there was no movement of the illuminating aperture with respect to the slit. The targets were both high contrast, with the lines having a density of approximately 3.0 and the clear spaces a density of 0.05. The target was mounted on the scanning carriage of the microphotometer and moved through the light beam. The target speeds were constant for a given pattern but for coarse patterns the scanning speed was higher than for fine patterns. An effort was made to maintain a constant lines-per-minute scanning rate for the entire target by changing target speed with a gear transmission.

One of the two targets used was a 2-in.-wide strip of the multi-line, as shown in figure 3a. The range of this target is from 0.2 to 2.0 lines/mm which gave a range, in the image plane of the lens under test, of from 5 to 50 lines/mm at 25 times reduction. The second target was a 2-in.-wide strip of the 75-line target (fig. 3b) which is in the form of 25 groups of 3 equally spaced lines. The line and space widths were the same in a given group; however, the spatial frequency values varied from one group to the next by a ratio equal to $\sqrt[6]{2}$. The range of resolving power of this target was from 0.2 to 3.2 lines/mm and gave a corresponding range in the image plane of 5 to 79.6 lines/mm.

2.2. Mirror Assembly

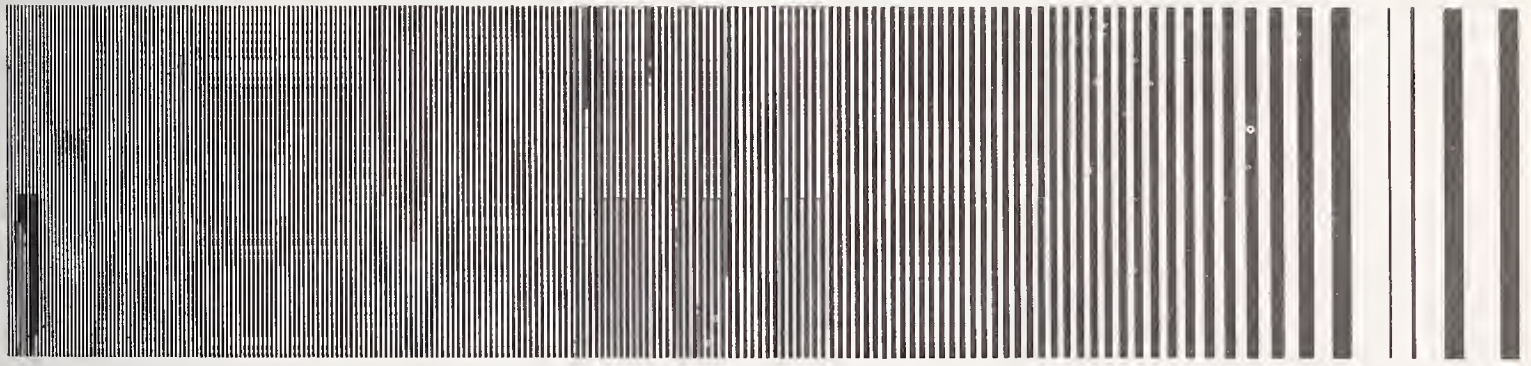
The front-surface mirror used for folding the optical axis to satisfy space limitations does not show in the schematic diagram or the photographs. It was actually an end mirror borrowed from a 4-in. aperture interferometer. The front surface was of very high quality and flatness, aluminized to enhance the reflecting properties. This mirror was in turn supported on a rigid mount, with provision for minute rotational adjustments about horizontal and vertical axes. Adjustments of this unit were necessary only in the initial alinement of the equipment. The distance from the target to the mirror was roughly the same as the distance of mirror to lens under test, making both legs of the folded system approximately equal. A poor quality mirror at this point in the system would have had a disastrous effect on the results.

2.3. Lens Assembly

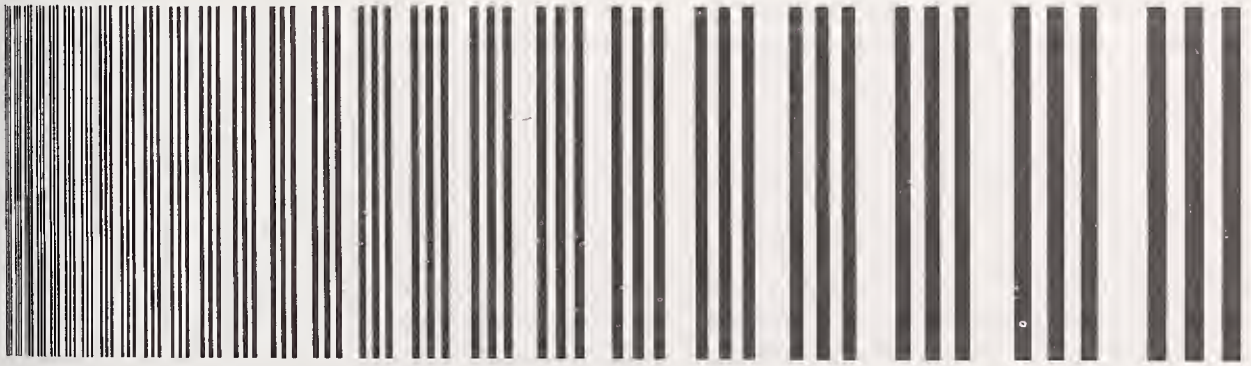
The lens to be investigated was mounted in a nodal slide as shown in figure 4 and this in turn was mounted on a saddle which slides parallel to the optic axis of the system. When the lens was correctly mounted in the slide, it could be moved about its vertical axis of rotation without displacing the image an appreciable amount to right or left of the receiving slit. This adjustment was not important for data taken only in axial positions as reported in this discussion, but a few trials were made at extra-axial positions. It was found that the equipment functioned equally well in off-axis positions, but results of these extra-axial positions are not reported here.

2.4. Receiving Unit

The receiving unit, shown in figure 1, was composed of the microscope, slit, and phototube with associated equipment such as amplifier and recorder. The microscope was of standard design, with the incorporation of an eyepiece at right angles to the line of sight and focused on the slit by means of a mirror, providing an opportunity to examine the image as it was focused on the slit. This feature was also necessary for making adjustments to render the line pattern in the image and the slit parallel. Adjustments could be made by rotating the slit with respect to the image. The microscope objective used in this work was a 10-power, 16-mm apochromat. The slit size was 35-μ wide by 10-mm long, which was somewhat shorter than the image of the



a



b

FIGURE 3. *Multiline (a) and 3-line (b) targets used in this investigation.*

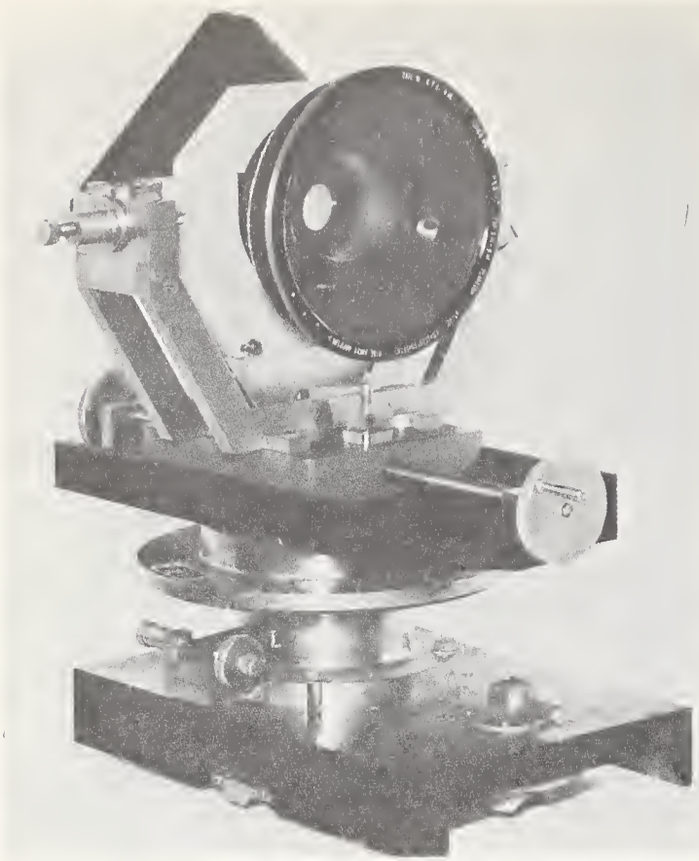


FIGURE 4. *Lens mounted in nodal slide.*

target lines, thereby virtually eliminating any of the end effects in the line patterns.

The slit support, phototube enclosure, and microscope were the standard receiving unit on the microphotometer. The phototube enclosure was originally made to accommodate a 921 type of phototube, but this did not have the necessary sensitivity for successfully recording the small flux densities which sometimes passed the somewhat narrow slit. A 1P21 photomultiplier tube was found to give much better results but would not fit into the standard housing. A new rear cover plate was made for the standard housing and the photomultiplier probe unit was attached to this cover. This separated the photomultiplier and slit about 2 in. which was more than a good design should allow. This was corrected by building a tunnel to collect all the flux emerging from the slit and pipe it to the tube. A high reflective coating on the interior of the tunnel reduced losses from scattering to a minimum. The photomultiplier unit had its own amplifier with four different ranges of sensitivity and a control for adjusting the dark current of the phototube. A small adapting unit was all that was necessary to run the microphotometer recorder from this amplifier. The recorder strip chart was a 100-division, uniformly spaced chart, which made the response plotted on a percent transmission scale.

The receiving unit, consisting of microscope, slit, and phototube, was mounted on one saddle similar to the one holding the nodal slide. Movement of one saddle with respect to the other was used as the means of locating a good focal plane of the lens. The nodal slide saddle was locked in position after the initial setup, and all focusing was done with the receiver unit. A dial indicator was fixed to the ways with its indicating probe in contact with the saddle for measuring the small displacements in focal position.

2.5. Alinement

Proper alinement of some of the components in this system was very important, while small misalinements in others did not appreciably affect the reliability of the data. The nodal slide and receiving unit were mounted on a short length of ways. It was important that the axis of these ways were parallel with the optical axis established by the light source and mirror. The target was mounted on the plate carriage of the microphotometer, thus establishing the direction of travel of the plate carriage perpendicular to the optical axis. Perpendicularity was checked by substituting a mirror for the target and adjusting the direction of the ways until reflected light from the source fell back upon itself.

The lens was mounted in the nodal slide with the aid of an axis finder to assure that the optical axis of the lens was parallel with the line of travel of the saddle on the ways. The lens was also located with its nodal point over the center of rotation of the nodal slide, thus assuring that the image would not move laterally from the receiving slit when the lens was rotated to simulate off-axis positions.

The microscope and slit combination were next alined along the axis established by the lens in the nodal slide. Alinement of these two components was important, as they were mounted on two different saddles, and focusing of the test lens was accomplished by varying the distance separating them. Misalinement here resulted in a lateral shift of the image when the lens-to-receiver distance was changed. The components on the ways were now in alinement with each other and it was relatively easy to make this optical axis coincident with the one for the whole system by moving the ways and saddles as a unit.

2.6. Calibration

The linearity of the amplifier, phototube, and the recorder was checked as a unit by substituting an adjustable slit for the fixed slit in front of the phototube unit. The space between slit and phototube was suitably masked to exclude all extraneous light, thus allowing only the slit to pass the light registering on the phototube. The recorder was then zeroed for dark current conditions. The slit was opened by small but uniform increments to its maximum opening. The amplifier and recorder thus produced a trace which was read and plotted against slit area to indicate the response of the system. The curve produced by these data verified the expected linearity of the system.

3. Description of a Typical Run

A typical run consisted of determining the response of 5 known spatial frequencies at 11 different focal positions on axis of a common type photographic objective. The response in the image was expressed in terms of contrast between lines and spaces. The target, figure 3a, provided a choice of 200 different spatial frequencies but to shorten the scanning and data handling time only 5 preselected frequencies were accurately measured, skipping over all intervening patterns. During a run the target was the only component in the system which moved. It could be moved slowly during measurement of each of the 5 frequencies and then moved rapidly through those between. The 5 spatial frequencies selected for measurement were 10, 20.5, 30.5, 40, and 63.5 lines/mm, but any or all others could have been used. A visual determination was made of the position of best focus and the first scan of the 5 frequencies was made at a focal position 1 mm back from this position. Starting at this long focal position a scan of the 5 frequencies was made at steps of 0.2 mm, to the other extreme of 1.0 mm inside of focus. This procedure produced a scan at 11 different focal positions. Measurement of the changes in focal position was accomplished with the dial indicator attached to the ways and bearing against the receiving unit saddle.

The traces produced by the recorder at each focal position were treated the same in that the maximum and minimum light transmission of each line and space of the predetermined frequency patterns was recorded. The contrast of each frequency was determined from the transmission data by using the formula

$$C=\frac{T_s-T_1}{T_s},$$

where T_s is the transmission of a space and T_1 of the adjoining line. These contrast data are plotted against spatial frequency in figure 5, which includes the 10 most informative of the 11 focal positions. There is a curve for each focal position, showing the manner in which the contrast decays toward zero. The larger numbers along the ordinate axis indicate the focal position and the smaller positive and negative numbers indicate the contrast scale. The abscissae represent the frequency in lines per millimeter. It will be noted that portions of the curves drop below the zero contrast value and have negative values. The negative values indicate the areas in which spurious resolution [5] is present and are only on the extreme short focus positions for this particular lens.

This equipment was found to lend itself very well to producing evidence showing what is taking place in these areas of spurious resolution. Two microphotometer traces of two aerial images are shown in figure 6. The trace which has the larger amplitudes represents a focal position close to best focus,

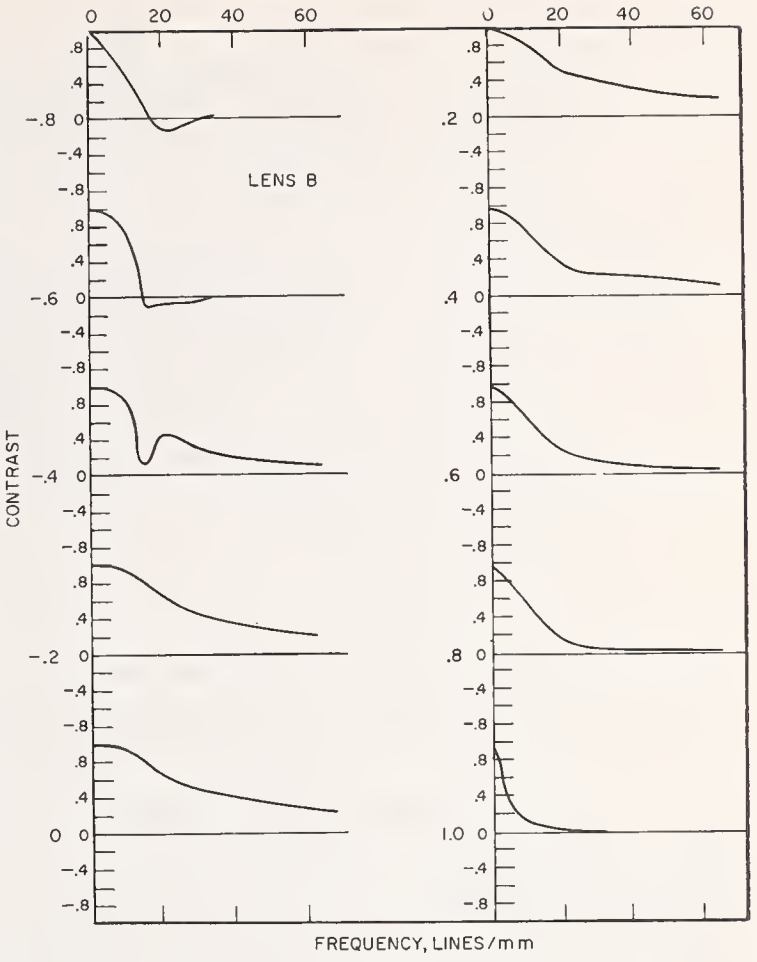


FIGURE 5. *A set of curves showing the contrast at various spatial frequencies of the direct scanned image as recorded at 10 different focal positions.*

Large numbers on the ordinate represent focal positions in millimeters.

and forms a good representation of the actual target in use as each line and space is registered in its relative position. The inner or low-amplitude trade represents a focal position 1.2-mm closer to the lens than the other trace. The two traces have been put in proper register with the maximum of the first line for the two traces in coincidence. It can be seen by following the inner trace that, for parts of the range of spatial frequency, the lines and spaces are registered out of phase and a line is indicated where a space should be. This indicates that, in this first zero point or node, one-half line has been lost, and the indicated resolution in this area to the next node is 180° out of phase. This is spurious resolution. At node two, another one-half cycle is lost. The area extending between nodes two and three shows the lines and spaces back in proper register but 360° out of phase, as one whole cycle has been lost.

The curves in figure 5 show the contrast as negative when the line-space register is out and positive when it is in. All of the contrast curve to the right of the first zero point represents spurious resolution.

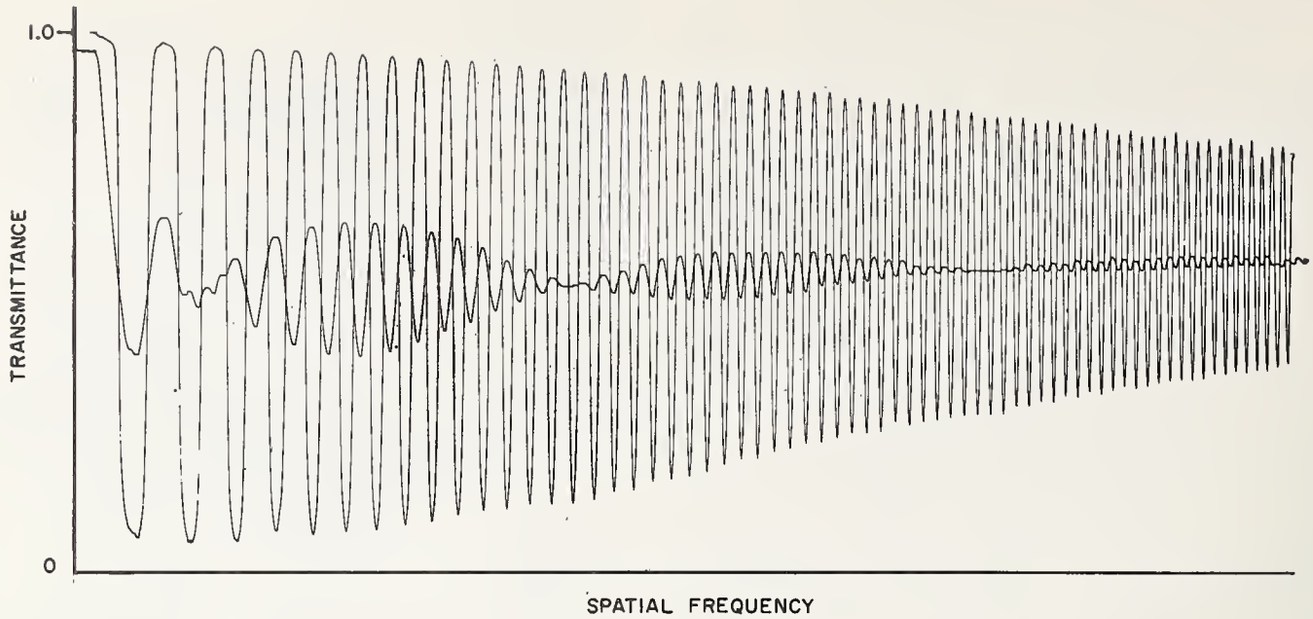


FIGURE 6. *Two microdensitometer traces made by direct scanning of the image.*
The trace with greatest magnitude was made at a position near best focus. Inside trace was made at a focal position 1.2-mm nearer lens. Note changes in phase.

4. Other Uses of Equipment

It is possible to insert color filters in the light path between the light source and target. The effective area of the receiving slit when projected back to the target is small enough to allow the use of standard 2-in. square filters. With the use of a suitable filter, investigative work can be carried out in a narrow spectral region. The use of several different narrow band filters permits the whole range of white light to be broken into as many narrow bands, which greatly reduces the effects of chromatic aberration within each band.

If repeated measurements of the same frequency pattern are made at different focal positions using the three-line target, it can be assumed that a best-focus position could be established by locating the position of maximum contrast rendition. This procedure was followed, using a filter and recording traces at 14 established focal positions. It was noted that at the various focal positions the contrast started at a low value, increased to a maximum, and then diminished. The filter was then changed to one which transmitted light in an adjacent spectral region. Another set of traces was taken using the same focal settings and noting the resultant contrast. This procedure was repeated using 6 different spectral regions to cover the range of visible wavelengths. When the 6 traces were arranged on a common focal position scale the results appeared as shown in figure 7. The numbers along the ordinate represent the Wratten filter number used to produce that trace, otherwise the ordinate scale is in transmittance. The spectral range passed by these filter numbers were: Blue with 45; green with 16 and 60; yellow with 90; orange with 25; and red with 29. Further investigation with different f/values and at off-axis positions could have easily been performed but exploration of the versatility of the equipment was the prime interest at the time.

Another image phenomenon which shows itself rather clearly can be seen by referring to figure 8. This illustration consists of two runs through focus in which the same spatial frequency pattern was scanned for each run. The upper run was the result of repetitive scans of a relatively coarse pattern of 5 lines/mm in the image while the lower run represents the response to the 64 lines/mm pattern. The unrestricted spectral range of a tungsten-lamp source was used for these traces. Close examination of the focal location of the two areas of maximum contrast shows a separation of about 0.2 mm. The values of contrast at each focal position could have been determined for these traces and curves of contrast versus focal position drawn to illustrate the same effect. The added work is not necessary here as the raw data alone indicate the results in a satisfactory manner.

The next experiment showing the possibilities of this apparatus involved making two apertures to be used as stops in the lens under investigation. These were constructed by depositing opaque material in the shape of a circle on thin cover glasses. The diameter of the full aperture of the lens at f/6.3 was measured and an annular ring was deposited which reduced the clear aperture of one-half the original area, or an f/value of 8.9. Another cover glass was used to make a negative of this aperture, which consisted of a circular opaque center with a transparent annular ring to transmit the light. The two apertures contain the same amount of clear area and are therefore the same f/value.

A full aperture (f/6.3) trace was made by again making repetitive scans of a constant spatial frequency 3-line pattern at 10 different focal settings

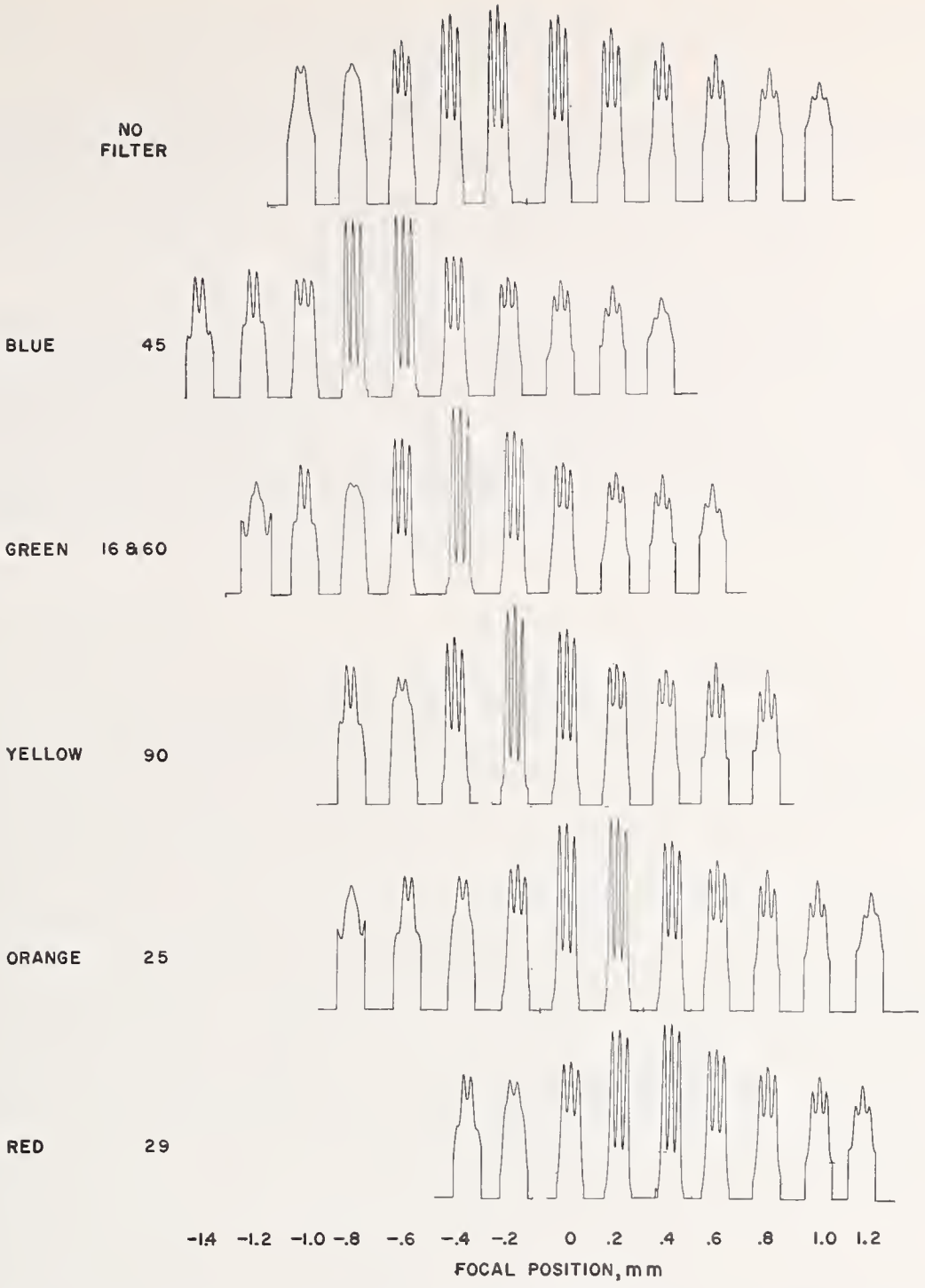


FIGURE 7. *Traces showing the response of a single 3-line pattern made at various focal positions.*
Each horizontal row was made in a different spectral region indicated by the filter number at the left. Note the shift in position of best focus.

while using a tungsten source as an illuminant. The full aperture trace was made with a clear cover glass in the same position the partially opaqued aperture stops were to occupy in the succeeding two runs. This was done to render constant the effect of the 0.1-mm glass thickness which was added to the lens system. Traces were then made with the two f/8.9 apertures, and the three traces arranged according to a common focal-position scale for comparison as seen in figure 9. A test along these lines might be used in a study of longitudinal spherical aberration in a lens system.

There are undoubtedly other areas of image evaluation which could be investigated with this equipment. One such area of study is that of star images, which was accomplished by substituting a camera for the photocell. One of the initial aims in building this apparatus was to reduce the time required for a lens test over the photographic method and this has attained that goal.

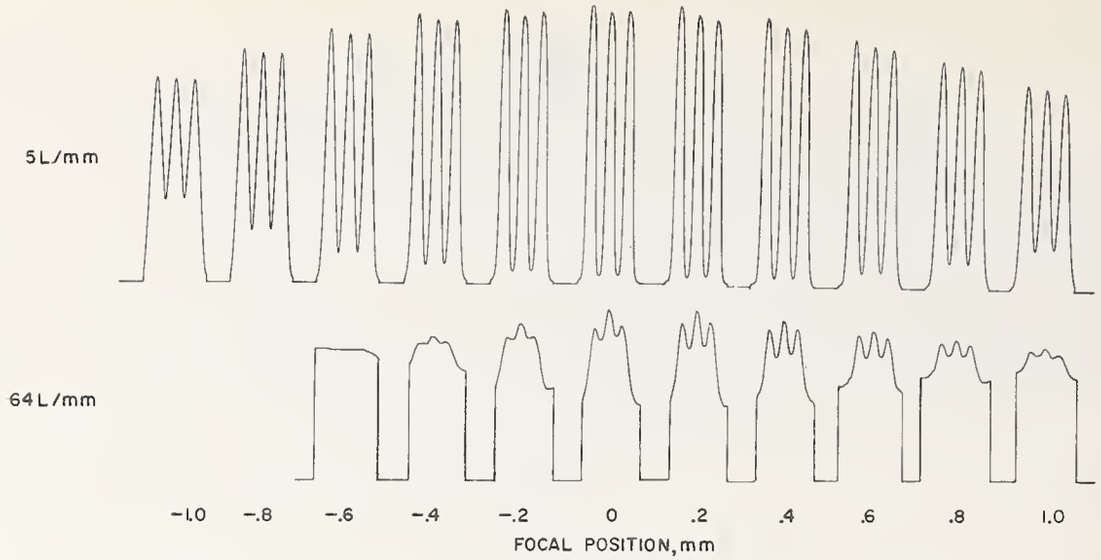


FIGURE 8. *Traces showing the response of two frequency patterns as indicated at left.*

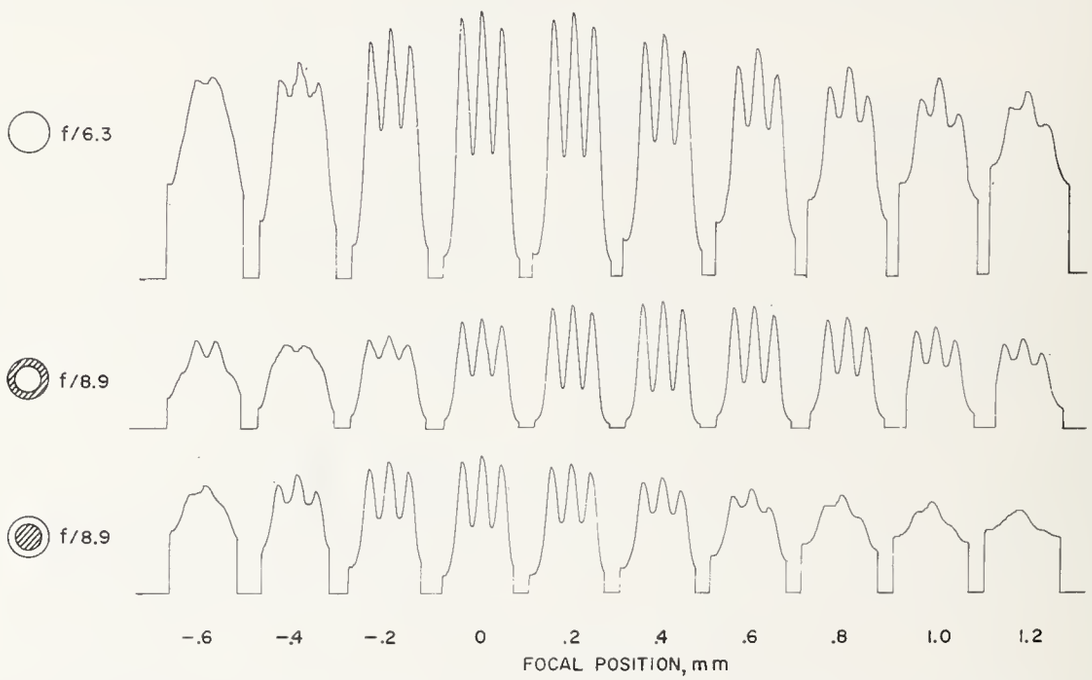


FIGURE 9. *Traces showing the response of repeated measurement of a single 3-line pattern at 10 different focal positions.*

Each horizontal row was made with a different aperture stop, the first stop was the normal, in the second the aperture was stopped down to f/8.9, and in the third the stop opening was shaped in the form of an annular ring with the same f/value as the second.

5. References

[1] F. E. Washer and F. W. Rosberry, New resolving power test chart, J. Opt. Soc. Am. **41** 9 (1951).

[2] F. E. Washer and I. C. Gardner, Method for determining the resolving power of photographic lenses, NBS Circ. 533 (1953).

[3] F. W. Rosberry, Effect of object frequency on focal position of four photographic objectives, J. Research NBS **57** 17 (1956) RP2688.

[4] O. H. Schade, A new system of measuring and specifying image definition, NBS Circ. 526 (1954).

[5] R. N. Hotchkiss, F. E. Washer, and F. W. Rosberry, Spurious resolution of photographic lenses, J. Opt. Soc. Am. **41** 9 (1951).

WASHINGTON, D.C. (Paper 64C1–25)

JOURNAL OF RESEARCH of the National Bureau of Standards—Engineering and Instrumentation
Vol. 64C, No. 3, July–September 1960

Variation of Resolving Power and Type of Test Pattern

Francis E. Washer and William P. Tayman

(April 7, 1960)

The plane of best average definition is located for a number of lenses of a type used in airplane mapping cameras. Three different types of test pattern are used for each lens. These patterns are the long-line, short-line, and annulus. Results of measurement that show the variation of resolving power throughout the region of usable imagery are given for each type of test pattern with two types of photographic emulsion. It is found that the plane of best average definition can be located equally well with each type of pattern. There are, however, pronounced differences in the numerical magnitudes of the values of resolving power determined with the various types of test pattern. In general, the highest values are attained with the long-line patterns. Values of the various rating indices $\sqrt{\overline{R_\beta T_\beta}}$, AWAR, and ADWAR are given together with a comparison of the different order of merit assigned by these indices.

1. Introduction

The quality of definition throughout the image plane of lenses used in airplane mapping cameras is of prime importance in ensuring satisfactory photography upon which to base the compilation of reliable maps. In forming an estimate of the probable quality of definition for a given lens, it is customary to use the measured values of resolving power for specified areas of the image plane as the principal criterion upon which to base the estimate.

The measured values of the resolving power have in general proved satisfactory for locating the plane of best average definition and for determining the suitability of a given lens for use in an aerial camera. Consequently, most specifications dealing with aerial cameras contain requirements prescribing the minimum acceptable values of the resolving power. Unfortunately, there is marked variance in the type of test chart used in various laboratories and in the conditions of use. This variance is a source of difficulty in any attempt to formulate an international specification dealing with the calibration of photogrammetric cameras as was done at the Washington meeting of the International Society of Photogrammetry in 1952 [1].[1] To obviate the difficulty, a specification was adopted for trial and discussion that permitted the use of several different types of resolving power target. It was noted that it was not possible at the time to standardize on a single type of resolving power target.

Three types of test chart were included in this tentative specification which were the three-line target used by the U.S. Air Force [2], the Cobb two-line chart used in Great Britain [3], and the annulus chart used in Canada [4]. All of these charts were of low contrast except the three-line chart of the Air Force which included both high- and low-contrast versions. The three-line chart used by the National Bureau of Standards was not included, as it had been in existence for only a few months prior to the conference, and its properties were not widely known.

Each of the charts covered by the specification included a series of patterns whose sizes were in a geometric progression with the $\sqrt[6]{2}$ preferred for the common ratio between adjacent pattern sizes. The NBS chart of 1952 not only differed from the others in length of line but also in the ratio between successive pattern sizes which is $\sqrt[4]{2}$.

In the seven years that have passed since the adoption of this specification, agreement on a single type of chart has not yet developed. A few publications have appeared [5, 6] that permit an estimate of the differences in resolving power likely to be found for line charts compared with annulus charts.

Because of some of the differences that exist between the procedures recommended for the calibration of photogrammetric cameras and those used at the National Bureau of Standards, it is worthwhile to describe the methods used at the Bureau and to compare the results derived therefrom with those that would be obtained by the methods approved in the international standard. In the present paper measured values of resolving power obtained with the NBS chart of 1952 [7] are compared with those obtained with the three-line chart used by the U.S. Air Force and with an annulus chart which differs from that prescribed in the international specification in that it is a high-contrast instead of a low-contrast target. In addition, the size ratio for successive patterns in each type of target is $\sqrt[4]{2}$ instead of $\sqrt[6]{2}$.

Results are presented for six lenses using two types of emulsion. Values of the several indices that might be used as rating factors such as $\sqrt{\overline{R_\beta T_\beta}}$, AWAR, and ADWAR are compared. It is found that the plane of best average definition may be located equally well with any of the three types. Values of resolving power obtained with the NBS chart are usually higher than those obtained with the other charts.

[1] Figures in brackets indicate the literature references at the end of this paper.

2. Method of Measurement

2.1. Test Camera

The National Bureau of Standards precision lens testing camera [8] was used in making the negatives from which the values of resolving power are determined. This device is well adapted for use in locating the plane of best average definition of a lens. Its construction and mode of operation simplifies the process of evaluating the resolving power of a lens in a series of focal planes spaced at definite intervals along the optical axis. Moreover, by using one type of chart in the odd-numbered collimators and another type in the even-numbered collimators, it is possible to register the imagery in the same identical series of focal planes for each type of chart for the same series of angular separations β from the axis on the same plate.

2.2. Test Charts

Three distinct types of test charts are used in the present investigation. All three are high-contrast charts, i.e., the difference in optical density between line and background exceeds 2.0. Each chart consists of a suitable range of sizes of the test object with the size proceeding in a geometric progression with $\sqrt[4]{2}$ (or 1.1892) as the common ratio. The various types of test chart are described in the following sections.

a. NBS Resolution Test Chart of 1952

This chart was introduced in 1952 and is described in detail in NBS Circular 533. The test chart, shown in figure 1, is a composite chart prepared by superimposing one test chart upon a four-times enlargement of the same chart. In this chart, all patterns have the same line count, namely three to each group. Two series of test patterns are included

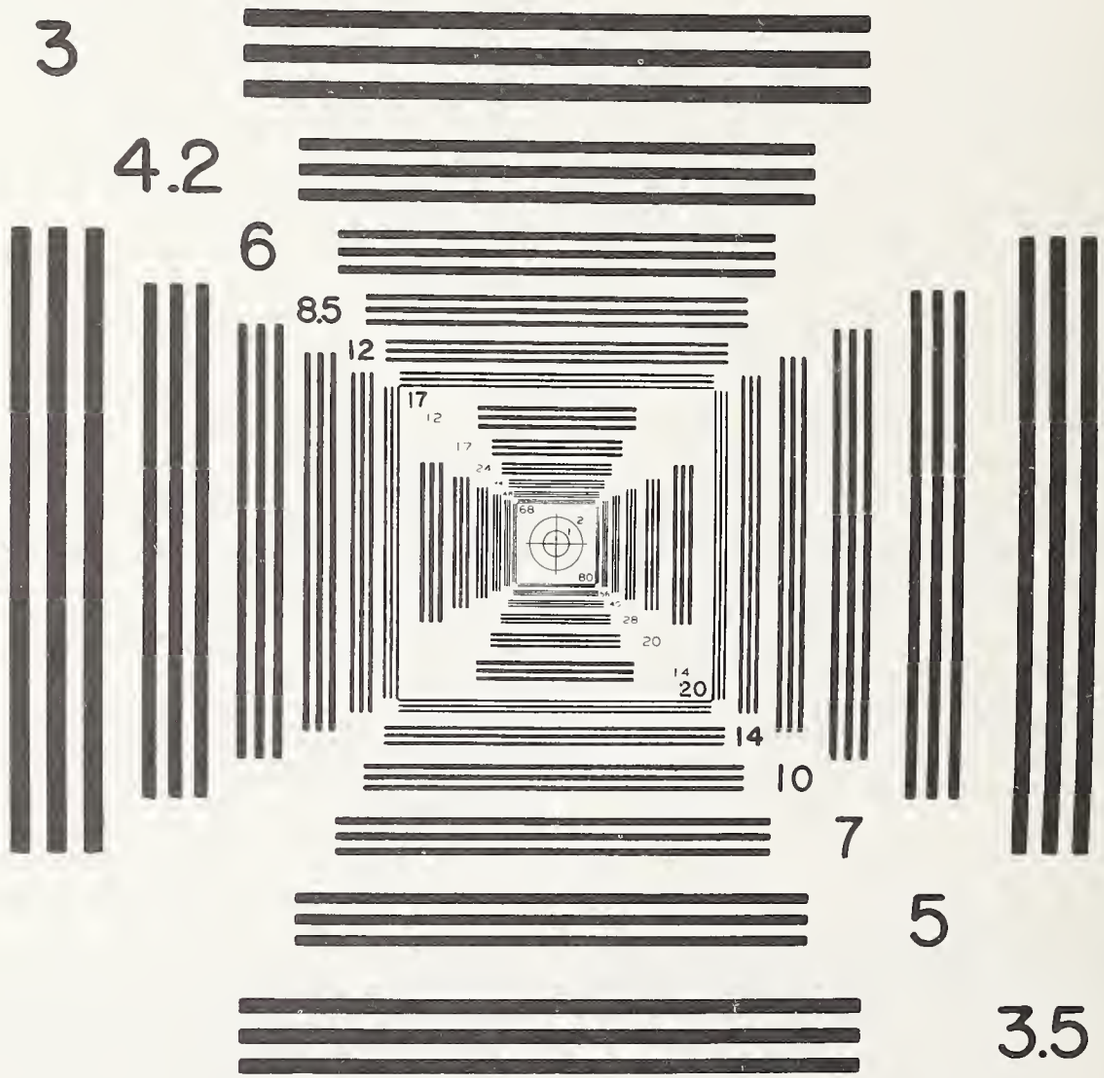


FIGURE 1. *NBS long-line resolution chart with extended range.*

This chart is a composite formed by superposing the NBS standard test chart upon a 4× enlarged copy of the master chart. The ratio of line spacings in adjacent groups of the master chart is equal to $\sqrt{2}$; however, the ratio of the line spacings of corresponding opposite patterns is $\sqrt[4]{2}$. Consequently, for this extended range chart, the values of resolving power that can be measured for a lens of 6-in. focal length range from 6.3 to 168 lines/mm. in a geometric series proceeding by $\sqrt[4]{2}$.

such that the spacings in the line pattern of a given group differ from those in the adjacent group in the ratio of $\sqrt[3]{2}$. However, the spacing in the initial pattern of one series differs from the initial pattern of the other series by $\sqrt[4]{2}$. Consequently, considered as a whole, the range of values provided by this chart proceeds by the $\sqrt[4]{2}$. The size of the chart used as a reticle in the collimators of the lens testing camera for this study was such that the range of values in the image plane of the lenses investigated extends from 6.3 to 168 lines/mm. The most striking feature of this chart is the relatively great length of the lines forming the patterns. The ratio of length to width of line is sufficiently great that the visual resolving power as read will not be subject to variations arising from end effects, and the images will continue to look like lines down to the limit of resolution. The ratio is also sufficiently great that the variation of resolving power with ratio of length to width of line can be neglected [5]. In addition, both the original test pattern and the images on a negative formed by a lens can be readily scanned with a recording microdensitometer. This property permits the user to make accurate measurements of the contrast between line and space for both object and image for studies involving degradation of contrast for a particular lens-emulsion combination.

The same size test chart is used in the various collimators so that it is necessary to apply the "cosine" and "cosine-squared" corrections to images formed off-axis.

b. Short Three-Line Charts

These charts, shown in figure 2, were adapted from that used in the camera calibrator [9]. The basic line pattern of this chart is the same as that prescribed in the specification prepared by Commission I of the International Society of Photogrammetry in 1952 [1] which is listed under paragraph 1.3.3 of the specification as the "three line high contrast target." The three-line patterns reduce in size in a geometric progression, the ratio being $\sqrt[4]{2}$. The sizes of the patterns in the off-axis targets have been modified to compensate for the "cosine" and "cosine-squared" factors affecting the resolving power values for radial and tangential lines. For lenses having a focal length of 6 in., the range of values of resolving power in the image plane of the lens under test extends from 5.9 to 264 lines/mm.

c. Annulus Test Chart

At the time the chart containing the series of short three-line patterns was made, a series of annulus patterns was included as may be seen in figure 2. The range of sizes is the same as for the short three-line patterns. The shape of the off-axis patterns was also distorted so that at specified angles they register on the test negative as circles. This pattern differs from that described in the international specification in paragraph 1.3.5 [1] in that it is high-contrast and consists of a dark circle on a light background.

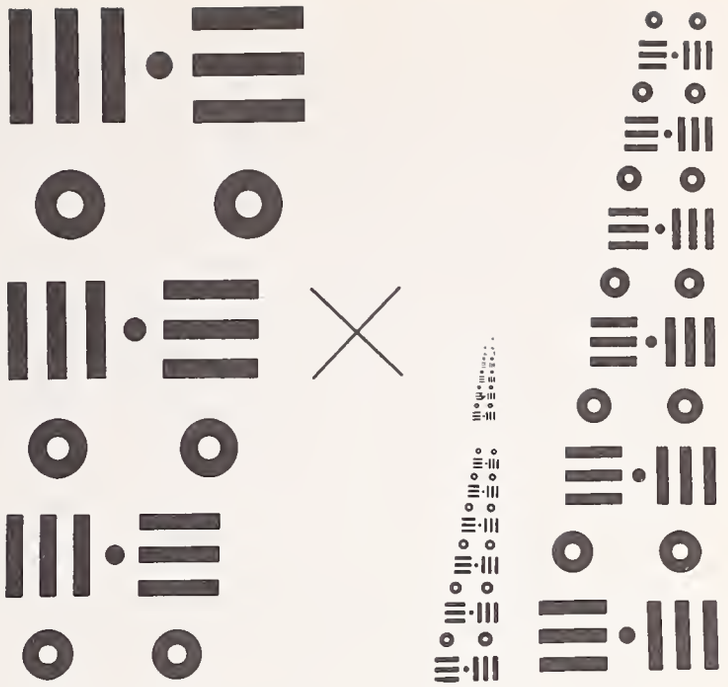


FIGURE 2. *Short three-line and annulus chart.*

The three-line chart consists of patterns of parallel lines in two orientations with spaces varying in a geometric progression by $\sqrt[4]{2}$. The circle or annulus charts change in size in a geometric progression by $\sqrt[4]{2}$. For a lens having a focal length of 6 in., the range of values in the image plane extends from 5.9 to 264 lines/mm.

2.3. Test Negative

In making the test negative, the lens under test is initially so alined that its optical axis is parallel to and approximately coincident with the axis of the collimated beam emergent from the first collimator. The lens is adjusted along the bench to a location such that the collimated beam at 40° from the axis fills the front aperture of the lens under test as viewed through the lens at an inclination of 40° from the axis. The plate holder is adjusted to a position such that the front surface of the emulsion is in the plane of best visual axial focus for the central row of images to be registered on the plate. The plate holder is then moved to a position 1.05 mm nearer to the lens, and an exposure is made by illuminating the reticles in odd-numbered collimators, which thus registers the imagery on the plate at 10° intervals for the range of angles from $\beta=0$ to $\beta=40°$. This process is repeated with the plate being moved 0.15 mm farther from the lens until 15 exposures have been made with the last for the plane 1.05 mm farther from the lens than the position of best visual axial focus. Between each exposure, the plate holder is also moved downward by an amount sufficient to avoid superposition of successive rows of images. The foregoing operation registers the imagery for the long-line patterns on the plate. The plate holder is then returned to its initial position for which the emulsion surface is 1.05 mm nearer to the lens than the plane of best visual axial focus; the plate is displaced sidewise in its holder approximately 12 mm; and the entire bench on which the camera is mounted is rotated on its pivot by 5° so

that the axis of the lens is now parallel to and nearly coincident with the axis of the collimated beam emergent from the second collimator. The foregoing procedure is then repeated with the exception that exposures are made by illuminating the reticles in the even-numbered collimators, which again register imagery on the plate at 10° intervals for the range of angles from $\beta=0°$ to $\beta=40°$. This process registers the imagery for the short three-line and annulus patterns on the plate. The exposed plate is then processed to form the finished negative from which values of resolving power are determined.

All exposures are made with the reticles illuminated by light from tungsten lamps with K–3 filters between the light source and test charts. Neutral filters are used to adjust the intensity of the illumination of the plate so that the final resulting optical density of each image on the negative is approximately the same for all values of β. The two types of photographic plates used in this work were Eastman Spectroscopic VF, which has a fine-grained panchromatic emulsion, and Eastman Super Panchro Press, Type C.

2.4. Reading the Negative

The negative images were examined with a microscope using powers ranging from 30× to 50×. The criteria for declaring a particular line pattern to be the finest resolved were that all coarser patterns were clearly resolved and that the number of lines in a given pattern was the same as that of the corresponding pattern in the object. For the annulus pattern, the prime criterion that all coarser patterns were resolved was used.

It was readily apparent that there were marked differences in the ease of reading the three types of patterns. The long-line pattern was the easiest of the three, and the feeling of confidence in the values recorded was accordingly the greatest. The short-line pattern was the next easiest, but there was an appreciably higher level of uncertainty in the reliability of the higher values of resolving power. The annulus pattern was the most difficult when attempting to determine the limit of resolution. It is noteworthy that the single circular black spot, which appears on the test chart in each cell of the target and which is identical in size with the white inner circular area of the annulus pattern, was easier to read than the annulus. It was usually possible to distinguish this single spot on the negative down to smaller sizes than the annulus; it could usually be detected down to one or two steps smaller in size. The presence of this spot was frequently used as an aid in determining the limiting resolution for the annulus.

3. Results of Measurements

The results of measurement on six wide-angle lenses are reported and analyzed in this study. Three lenses, designated Nos. 1, 2, and 5, are essentially distortion free; the other three, designated 3, 4, and 6, have moderate amounts of distortion. Measurements of resolving power were made at 10° intervals from 0° to 40°.

Negatives were made with both VF and SP emulsions for each of the six lenses. For a given lens, the plane of best visual axial focus is the same for each emulsion. The measured equivalent focal length for the plane of best visual axial focus for a given lens was the same for each emulsion and type of target pattern. The maximum range of change in measured equivalent focal length for the various conditions of test did not exceed ± 0.02 mm for a given lens.

In the presentation and analysis of the results of measurement in an investigation of this sort, it is necessary to keep in mind the primary aims of the study. First, the intercomparison of the magnitudes of the measured values of resolving power for each type of pattern and emulsion is important. Second, it is important to determine whether or not any shift of the plane of best definition occurs with type of target pattern. In addition, the manner in which the values of resolving power for the various patterns wax and wane as the image plane approaches and passes through the plane of best visual axial focus is of interest.

In an earlier paper [10], the measured values of resolving power for tangential lines (T_β) and radial lines (R_β) obtained with the short three-line chart using VF and SP emulsions for values of β ranging from 0° to 40° at 10° intervals are presented in tabular form for lens No. 1. In addition, the values of the geometric means values $\sqrt{R_\beta T_\beta}$ are also given. In this paper, a discussion of the various indices used in locating the plane of best average definition is given. These indices are as follows:

(*a*) *Root Mean Square Mean* (*RMSM*). This is defined as

$$\mathrm{RMSM}=\sqrt{\frac{\sum R_\beta T_\beta}{n}}=\sqrt{\overline{R_\beta T_\beta}}.$$

In the above expression, R_β and T_β are measured values of resolving power for radial and tangential lines for discrete values of β ranging from 0° to 40° in 10° steps. The number of steps, n, in this instance is five.

(*b*) *Area Weighted Average Resolution* (*AWAR*). This is defined as

$$\mathrm{AWAR}=\sum \frac{a_\beta}{A}\sqrt{R_\beta T_\beta},$$

where a_β is the area of a particular annular zone, R_β is the average resolving power for radial lines, and T_β is the average resolving power for tangential lines for that zone. The value of $\sum a_\beta$ is the entire area A of the picture formed.

(*c*) *Area and Depth-of-Focus Weighted Average Resolution* (*ADWAR*). This is defined as

$$\mathrm{ADWAR}=\sum \frac{a_\beta r_\beta}{A}\sqrt{R_\beta T_\beta},$$

where

$$r_\beta^2=\frac{d_0 d_0}{d_{T_\beta} d_{R_\beta}},$$

where d_{T_β} is the depth of focus for tangential lines, d_{R_β} is the depth of focus for radial lines, and d_0 is the depth of focus on axis for a given value of resolving power.

Because some of the results for one type of pattern used in this paper have already been presented in tabular form, it was decided to present the results of measurement and some of the quantities derived from the measurements in graphical form.

The results of measurement are presented in figures 3 through 9. Values of the resolving power for tangential lines (T) and radial lines (R) are shown for the two emulsions and three patterns for lenses 1 and 3 only. These results on these two lenses are given in the left and central columns of graphs in figures 3, 4, 5, and 6. The nature of the test pattern is indicated by the symbols, L for long-line, S for short-line, and A for annulus. The values obtained for the annular patterns are shown on both sets of graphs for ready comparison with the T and R values for L and S. It is readily apparent from these graphs that the higher values found for resolution of long lines are more evident for the regions inside focus (negative values of Δf) than for the regions outside focus (positive values of Δf). This is in keeping with a similar finding reported by Selwyn and Tearle [11]. This effect is present for both VF and SP emulsions for axial resloving power; it is usually more noticeable for the VF emulsion. The values found for the annulus pattern are generally substantially lower than the values for either tangential or radial lines throughout the region of usable imagery as might be inferred from similar comparisons previously reported [5, 6].

The values of $\sqrt{R_\beta T_\beta}$, the geometric mean value of resolving power for tangential and radial lines, are given for each value of β for each of the six lenses for both emulsions. For lenses No. 1 and 3, values of $\sqrt{R_\beta T_\beta}$ are shown in the right-hand column of figures 3, 4, 5, and 6. The values of $\sqrt{R_\beta T_\beta}$ for lenses No. 2, 5, 4, and 6 are presented in the five upper sets of graphs in figures 8 and 9. These curves permit a ready comparison of the average values of resolving power for the three types of patterns for the five values of β. It also seems more appropriate to use values of $\sqrt{R_\beta T_\beta}$ in a comparison of values obtained by line targets with that obtained by the annulus target. For the annulus, the value of $\sqrt{R_\beta T_\beta}$ is simply the observed value of the resolving power.

For lenses No. 1 and 3, values of the three indices $\sqrt{\overline{R_\beta T_\beta}}$, AWAR, and ADWAR for the three types of line patterns and two emulsions are given in figure 7. For lenses 2, 5, 4, and 6, values of $\sqrt{\overline{R_\beta T_\beta}}$ are shown in the lower row of graphs in figures 8 and 9. Values of AWAR and ADWAR were determined for these four lenses but are not shown.

3.1. Location of the Plane of Best Average Definition

The plane of best average definition was located for each lens both by graphical analysis and by location of the maximum value of the various indices for each type of pattern and each emulsion.

a. Graphical Analysis

The plane of best average definition for each lens for each set of conditions was located by the graphical method reported in a previous paper [10]. The separations, Δf, along the optical axis of the lens of the focal planes of best average definition for short-line and annular patterns selected by graphical analysis from the focal planes of best average definition for long-line patterns selected in the same manner, are shown in table 1. It is clear from this table that there is no significant variation in the location of the focal plane of best average definition for these three types of test pattern. Such variation, as appears is usually random and within the limits of experimental error. Table 2 shows the variation in location of the plane of best average definition for the three types of patterns using SP emulsion compared with that for long-line patterns and VF emulsion. In this instance, the optimum image plane for SP emulsion appears on the average to be slightly nearer the lens than the corresponding plane for VF emulsion.

TABLE 1. *Separations, Δf, of the focal plane of best average definition for long lines selected by graphical analysis from focal planes of best average definition for short-line and annulus patterns selected by graphical analyses*

Results are given for six lenses and two emulsions.

Lens No.	Values of Δf			
	VF emulsion		SP emulsion	
	Short line	Annulus	Short line	Annulus
	mm	*mm*	*mm*	*mm*
1	0.00	0.15	0.00	0.00
2	.00	.15	.00	−.15
3	.30	.00	.00	.00
4	.00	.15	.00	.00
5	.00	.00	.00	.00
6	.00	.00	.00	.00

TABLE 2. *Separations, Δf, of the focal plane of best average definition for long lines with VF emulsion selected by graphical analysis from the focal planes of best average definition for long line, short line, and annulus patterns obtained with SP emulsion by graphical analyses*

Lens No.	Values of Δf		
	Long line	Short line	Annulus
	mm	*mm*	*mm*
1	0.15	−0.15	−0.15
2	−.15	−.15	−.30
3	.00	.00	.00
4	.00	.00	.00
5	.00	.00	.00
6	−.15	−.15	−.15

518-743 O - 73 - 36

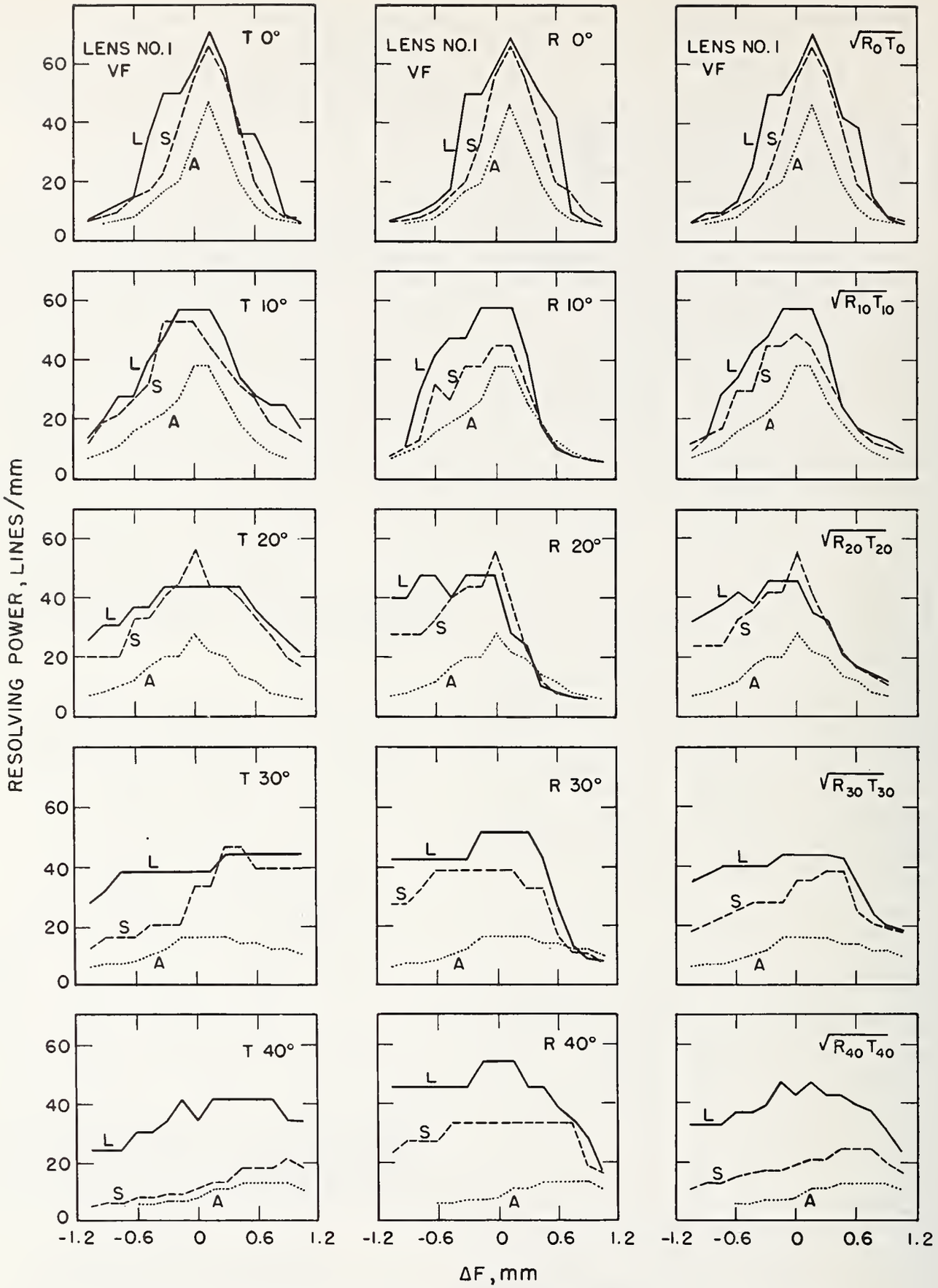


FIGURE 3. *Resolving power versus position of the image plane for lens No. 1 with VF emulsion.*

The resolving power throughout the region of usable imagery is shown for tangential (T) and radial (R) lines and for the average $\sqrt{R_\beta T_\beta}$ at 10° intervals from 0° to 40°. Values are given for long-line (L) and short-line (S) test patterns. Values for the annular (A) pattern are shown with the corresponding T, R, and $\sqrt{R_\beta T_\beta}$ values for ready comparison. The zero of abscissas (Δf) marks the position of best visual axial focus, and positive values of abscissas indicate positions farther from the lens.

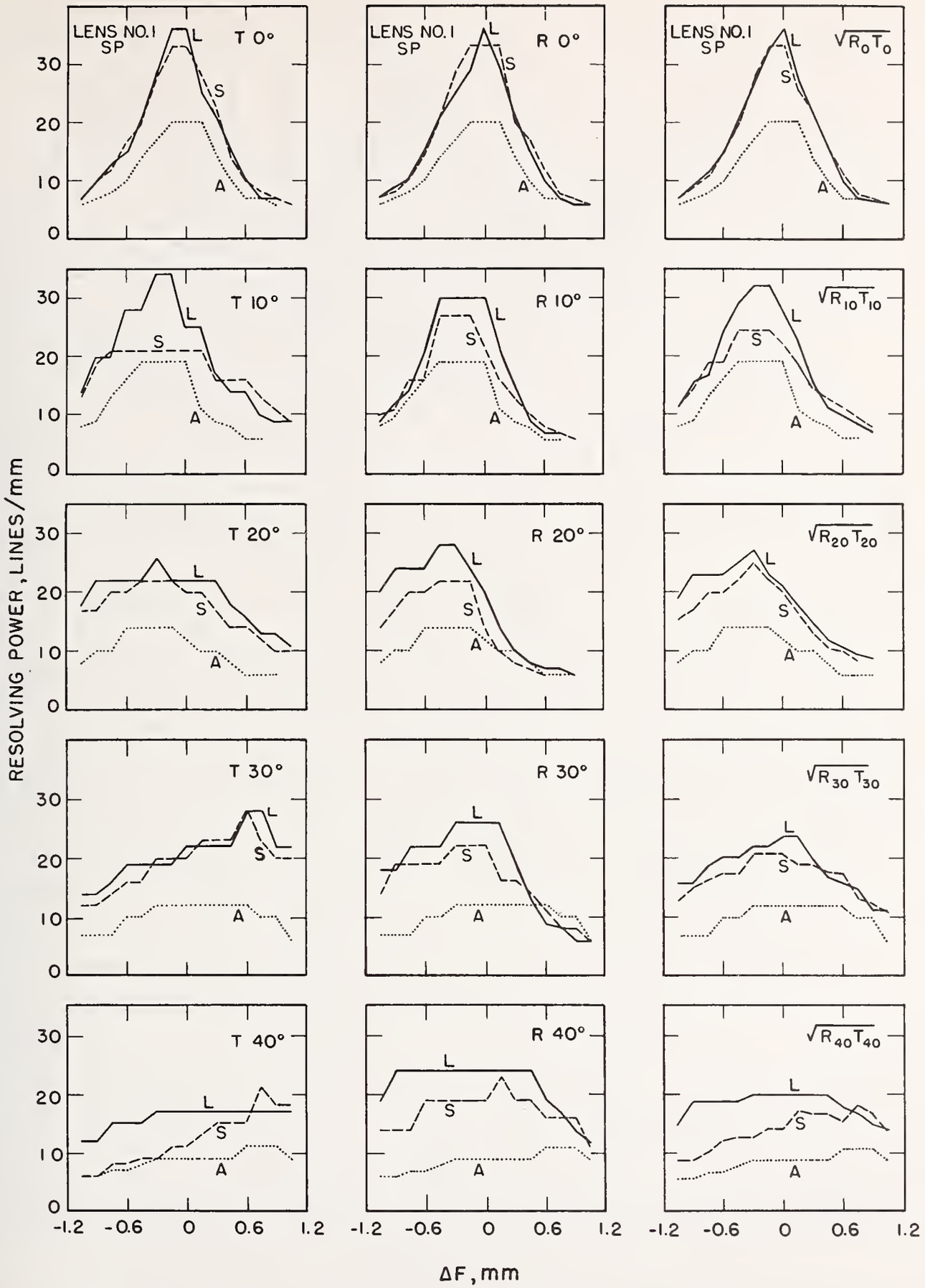


FIGURE 4. *Resolving power versus position of the image plane for lens No. 1 with SP emulsion.*

The resolving power throughout the region of usable imagery is shown for tangential (T) and radial (R) lines and for the average $\sqrt{R_\beta T_\beta}$ at 10° intervals from 0° to 40°. Values are given for long-line (L) and short-line (S) test patterns. Values for the annular (A) pattern are shown with the corresponding T, R, and $\sqrt{R_\beta T_\beta}$ values for ready comparison. The zero of abscissas (Δf) marks the position of best visual axial focus, and positive values of abscissas indicate positions farther from the lens.

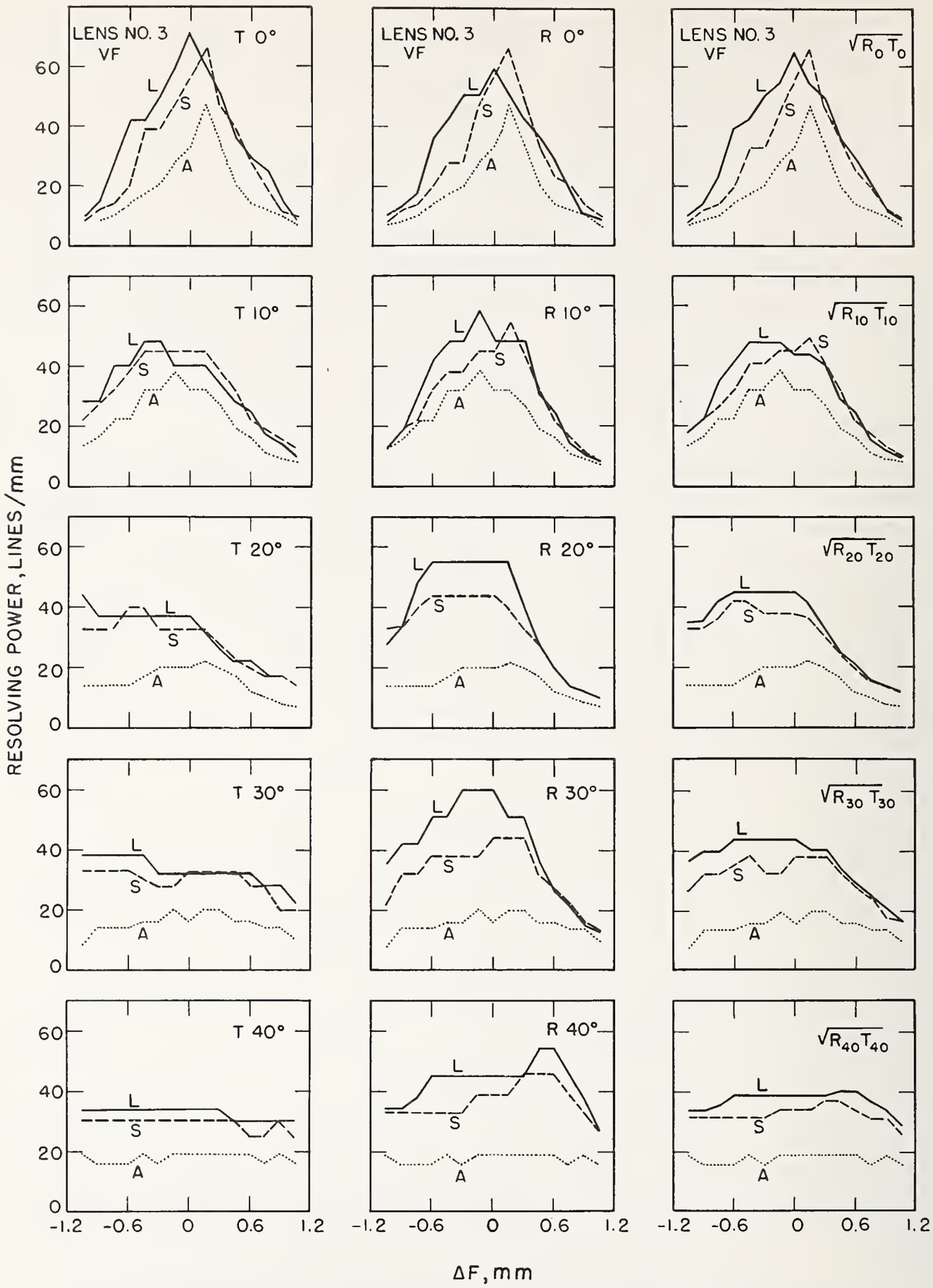


FIGURE 5. *Resolving power versus position of the image plane for lens No. 3 with VF emulsion.*

This set of curves gives information on lens No. 3 of the same type that is given in figure 3 for lens No. 1.

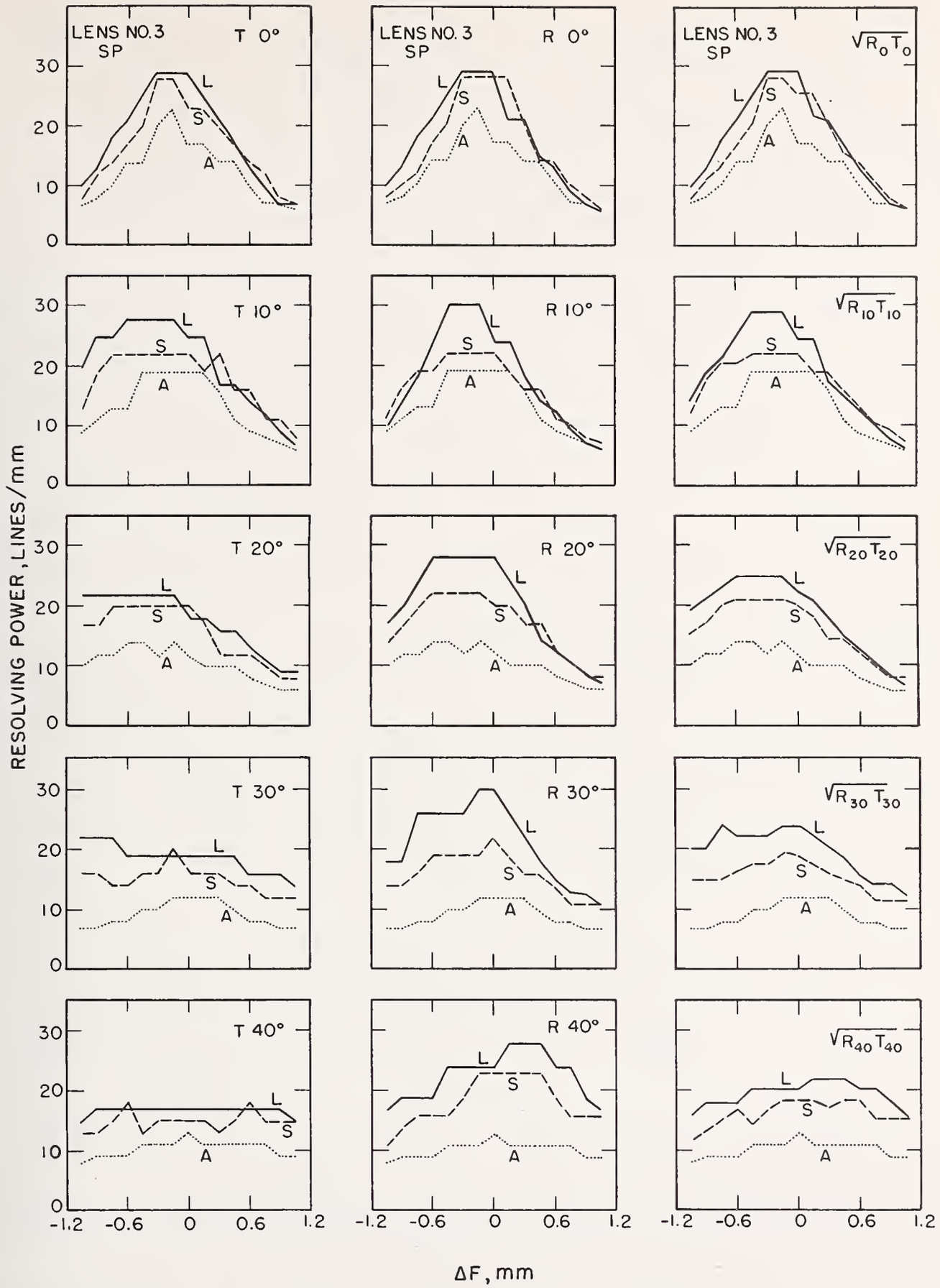


FIGURE 6. *Resolving power versus position of the image plane for lens No. 3 with SP emulsion.*

This set of curves gives information on lens No. 3 of the same type that is given in figure 4 for lens No. 1.

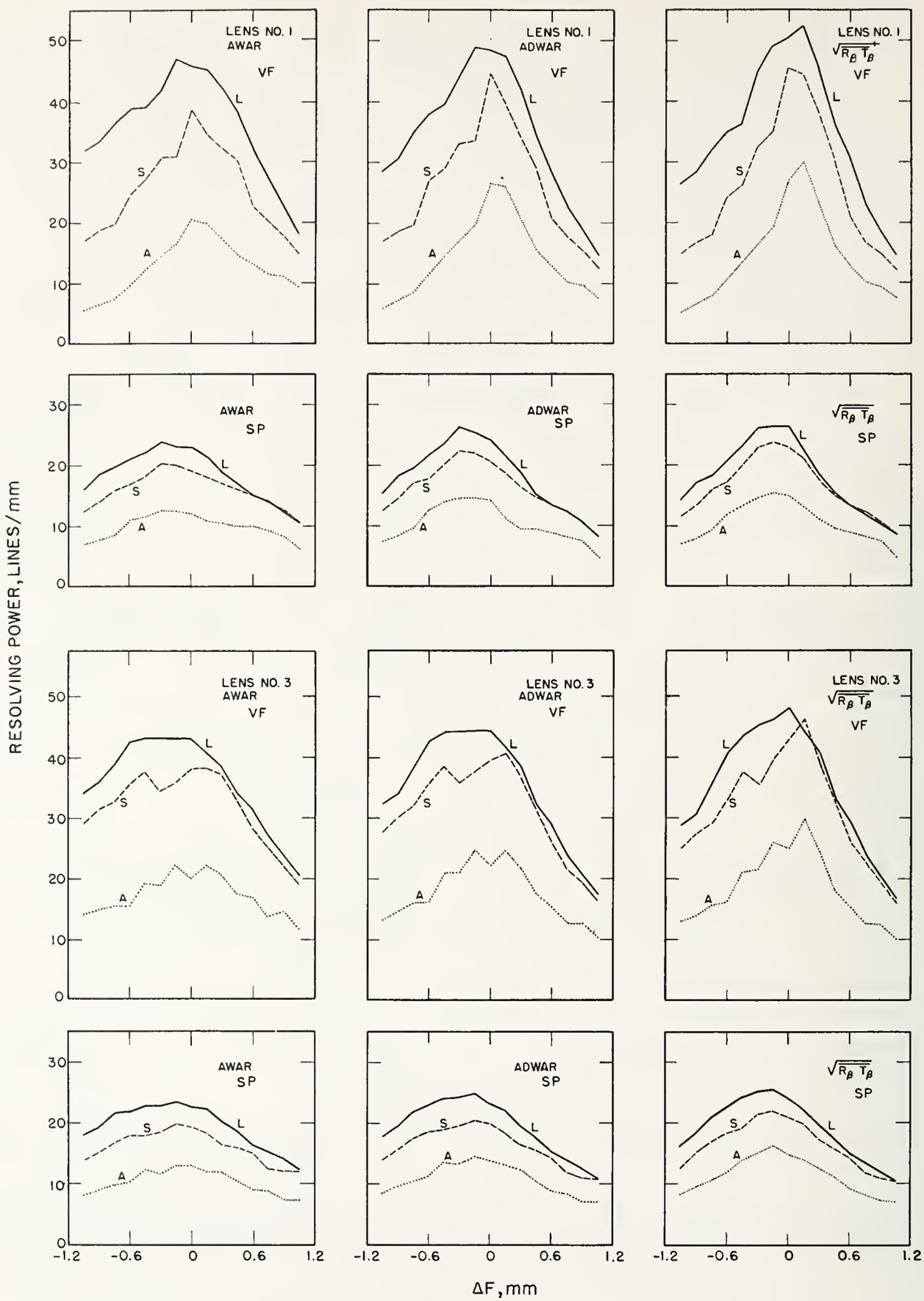


FIGURE 7. *Variation of AWAR, ADWAR, and* $\sqrt{R_\beta T_\beta}$ *with position of the image plane for lenses Nos. 1 and 3 for VF and SP emulsions.*

Values of these indices are given in lines per millimeter for long-line (L), short-line (S), and annular (A) patterns. The zero of abscissas (Δf) marks the position of best visual axial focus, and positive value of abscissas indicate positions farther from the lens.

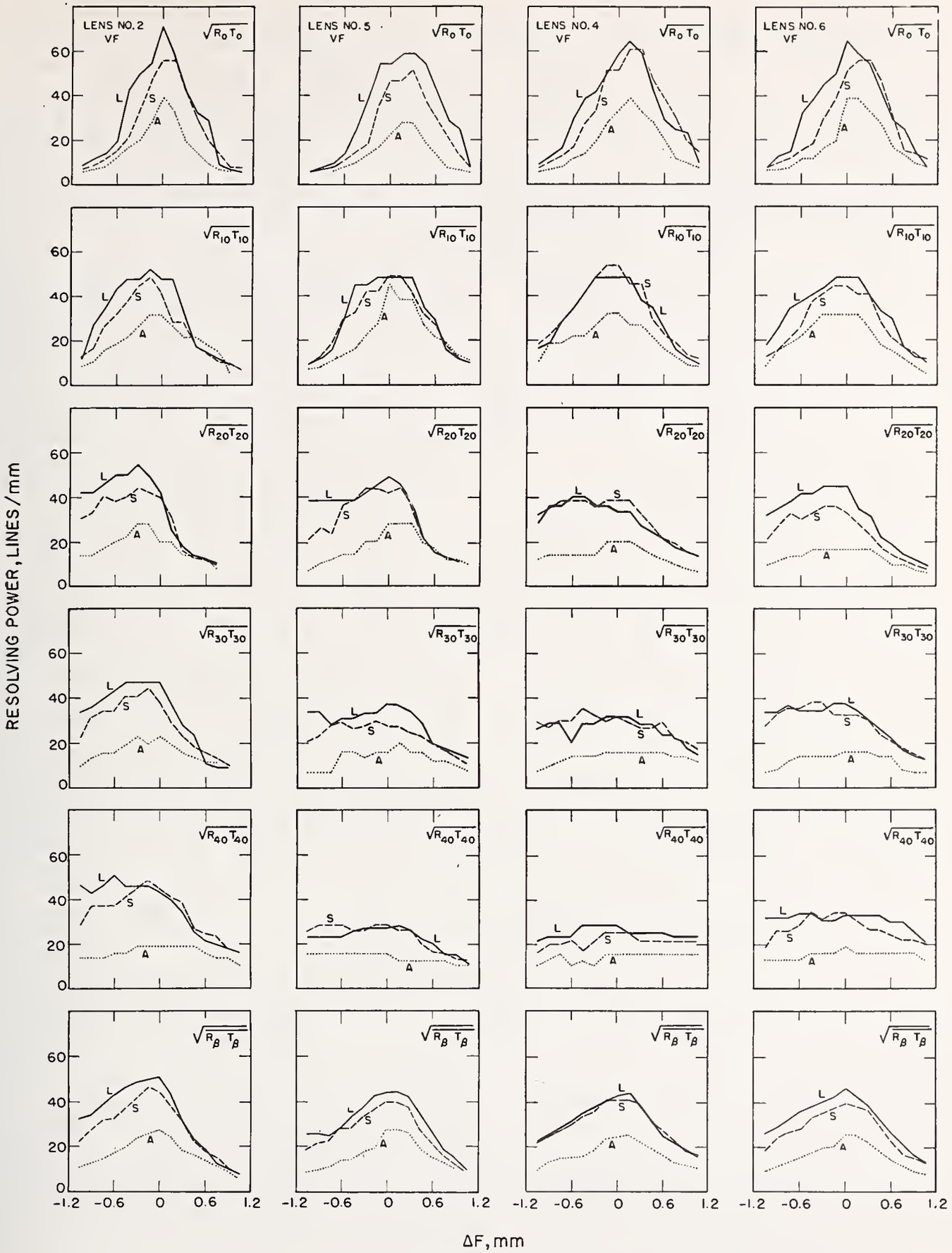


FIGURE 8. *Average resolving powers versus position of the image plane for four lenses with VF emulsions.*

Values of the average resolving power $\sqrt{R_\beta T_\beta}$ are shown at 10° intervals from 0° to 40°. The lowest box in each column shows values of the root mean square mean $\sqrt{\overline{R_\beta T_\beta}}$. Values are given for long-line (*L*), short-line (*S*), and annular (*A*) patterns. The zero of abscissas (Δf) marks the position of best visual axial focus, and positive value indicate positions farther from the lens.

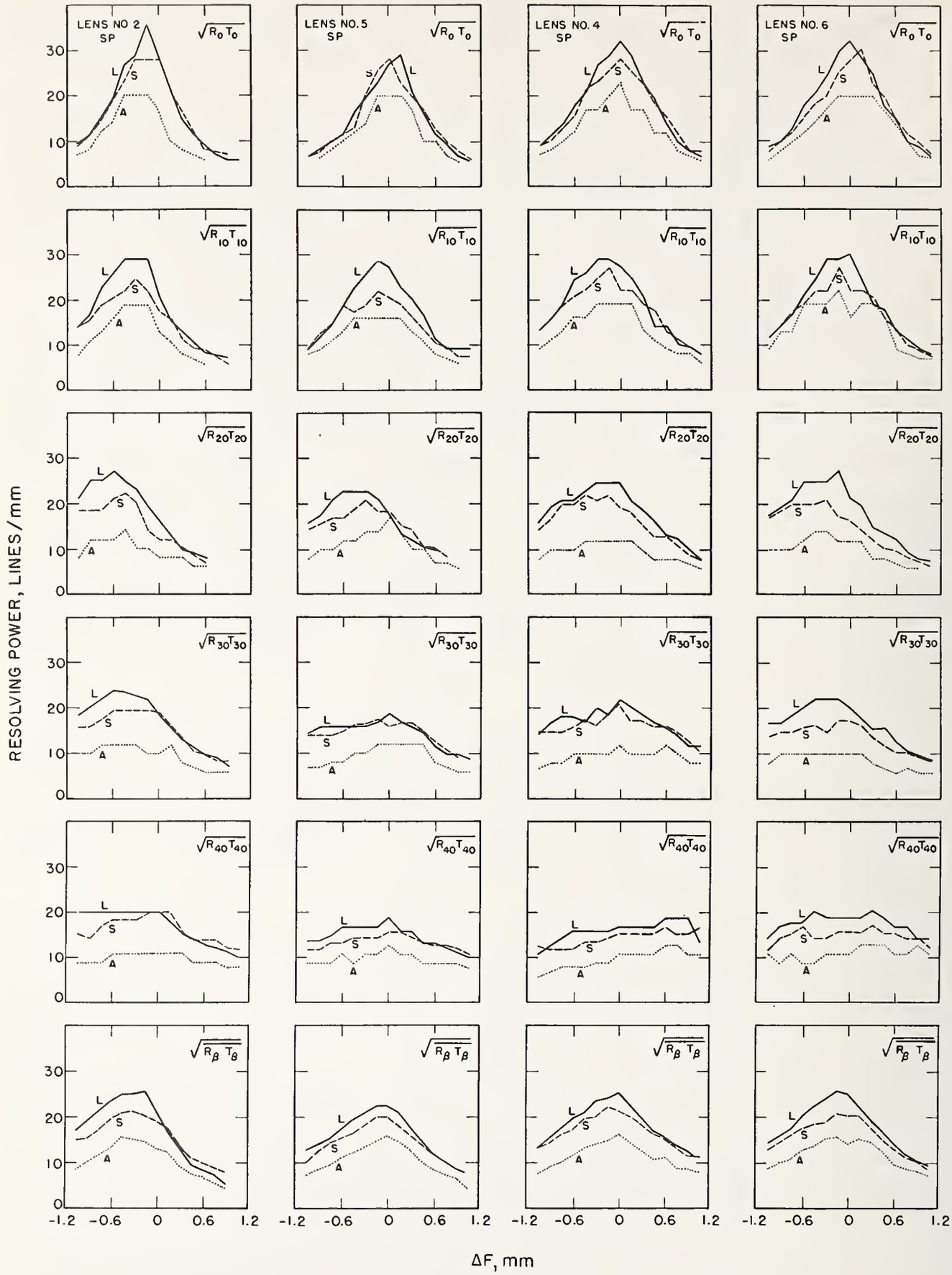


FIGURE 9. *Average resolving powers versus position of the image plane for four lenses with SP emulsions.*

Values of the average resolving power $\sqrt{R_\beta T_\beta}$ are shown at 10° intervals from 0° to 40°. The lowest box in each column shows values of $\sqrt{\overline{R_\beta T_\beta}}$. Values are given for long-line (*L*), short-line (*S*), and annular (*A*) patterns. The zero of abscissas marks the position of best visual axial focus, and positive value indicate positions farther from the lens.

b. Maximum Value of Indices

The plane of best average definition for each lens for each set of conditions was also located with the aid of each of the three indices, $\sqrt{\overline{R_\beta T_\beta}}$, AWAR, and ADWAR. In each instance, the optimum image plane was located with respect to the focal plane of best average definition for long-line patterns selected by graphical analysis. The results are shown in table 3. It is clear from this table that the variation is, in general, within the limits of experimental error.

TABLE 3. *Separations, Δf, of the focal plane of best average definition for long lines selected by graphical analysis from the focal planes indicated by the maximum value of the indices* $\sqrt{\overline{R_\beta T_\beta}}$, *AWAR, and ADWAR*

Results are shown for six lenses and two emulsions.

Lens No.	VF emulsion			SP emulsion		
	Long line	Short line	Annulus	Long line	Short line	Annulus
(a) Separations, Δf, of plane selected by graphical analysis from planes selected by maximum value of $\sqrt{\overline{R_\beta T_\beta}}$ for						
	mm	*mm*	*mm*	*mm*	*mm*	*mm*
1	0.15	0.00	0.15	0.00	0.00	0.00
2	.15	.00	.15	.15	.00	−.15
3	.15	.30	.30	.00	.00	.00
4	.15	.00	.15	.00	−.15	.00
5	.15	.15	.00	.00	.00	.00
6	.00	.00	.00	.00	.00	.00
(b) Separations, Δf, of plane selected by graphical analysis from planes selected by maximum value of AWAR for						
1	−0.15	0.00	0.00	−0.15	−0.15	0.00
2	−.15	.00	−.15	−.30	−.15	−.15
3	−.30	.30	.00	.00	.00	.15
4	.00	.00	−.15	.00	.00	.00
5	.00	.00	.00	.00	−.15	.00
6	.00	−.30	.00	.00	.00	.00
(c) Separations, Δf, of plane selected by graphical analysis from planes selected by maximum value of ADWAR for						
1	−0.15	0.00	0.00	−0.15	−0.15	0.00
2	.00	.00	.00	−.15	.00	−.15
3	−.15	.30	.00	.00	.00	.00
4	.00	−.15	−.15	.00	−.15	.00
5	.00	.00	.00	.00	−.15	.00
6	.00	−.15	.00	.00	.00	.00

4. Comparison of Results Obtained With the Three Charts

In the foregoing sections, it was shown that the focal plane of best average definition could be readily located using any one of the three types of test charts under consideration. Moreover, such minor differences in location that were noted could be regarded as negligible in comparison with experimental error.

In the present section, some attention is given to the actual magnitudes of the values of resolving power obtained with the various charts. Some attention is also given to the use of the indices as rating factors.

4.1. Values of Resolving Power in Focal Plane of Best Average Definition

Values of $\sqrt{R_\beta T_\beta}$, $\sqrt{\overline{R_\beta T_\beta}}$, AWAR, and ADWAR are shown in tables 4 and 5 for various patterns and emulsions. These values are all for the plane of best average definition for long-line patterns. In some instances, the value of a given index was higher for an adjacent plane. This is indicated in table 6 which shows the departure in percent of the maximum of a given index from the value reported in tables 4 and 5. It is noteworthy that the departures are for the most part negligible. It is clear from tables 4 and 5 that it would be difficult to determine the value of resolving power that might be obtained for a given type of pattern at a given angle β when

TABLE 4. *Comparison of values of average resolving power obtained with three types of test chart using VF emulsion in the plane of best average definition for long lines selected by graphical analysis.*

Values of $\sqrt{R_\beta T_\beta}$, $\sqrt{\overline{R_\beta T_\beta}}$, AWAR, and ADWAR are shown for six lenses for long line, short line, and annulus charts. All values of resolving power are given in lines per millimeter.

Lens No.	Target	$\sqrt{R_\beta T_\beta}$					Indices		
		0°	10°	20°	30°	40°	$\sqrt{\overline{R_\beta T_\beta}}$	AWAR	ADWAR
1	Long line	59	58	46	44	43	50	46	49
	Short line	56	49	56	35	19	45	38	45
	Annulus	33	38	28	16	8	27	20	26
2	Long line	54	52	49	47	46	50	48	49
	Short line	47	49	42	45	49	46	46	46
	Annulus	28	32	28	20	19	26	23	26
3	Long line	54	48	45	44	39	46	43	46
	Short line	66	49	36	38	34	46	38	41
	Annulus	33	32	20	16	19	25	20	22
4	Long line	59	48	39	32	29	43	35	38
	Short line	51	54	34	33	26	41	34	38
	Annulus	33	32	20	16	13	24	18	21
5	Long line	54	48	49	38	28	44	39	41
	Short line	47	49	42	28	29	40	34	38
	Annulus	23	45	28	16	16	28	23	28
6	Long line	65	48	45	38	33	47	39	42
	Short line	51	45	34	33	34	40	35	37
	Annulus	39	32	17	16	19	26	20	22

TABLE 5. *Comparison of values of average resolving power obtained with three types of test chart using SP emulsion in the plane of best average definition for long lines as selected by graphical analysis*

Values of $\sqrt{R_\beta T_\beta}$, $\sqrt{\overline{R_\beta T_\beta}}$, AWAR, and ADWAR are shown for six lenses for long line, short line, and annulus charts. All values of resolving power are given in lines per millimeter.

Lens No.	Target	$\sqrt{R_\beta T_\beta}$					Indices		
		0°	10°	20°	30°	40°	$\sqrt{\overline{R_\beta T_\beta}}$	AWAR	ADWAR
1	Long line	32	32	23	22	20	23	25	26
	Short line	33	24	22	21	14	24	20	22
	Annulus	20	19	14	13	9	15	12	15
2	Long line	29	29	23	23	20	25	23	25
	Short line	25	24	20	20	18	22	20	21
	Annulus	20	19	10	12	11	15	12	14
3	Long line	29	29	25	24	20	26	24	25
	Short line	28	22	21	20	19	22	20	21
	Annulus	23	19	14	12	11	16	13	14
4	Long line	32	27	25	22	17	25	22	24
	Short line	25	27	22	19	14	22	19	21
	Annulus	23	19	12	12	11	16	13	14
5	Long line	27	27	21	19	19	23	20	22
	Short line	28	20	18	16	16	20	17	18
	Annulus	20	16	17	12	13	16	14	15
6	Long line	30	29	27	22	19	23	25	26
	Short line	25	27	18	18	16	21	18	20
	Annulus	20	22	12	10	11	16	12	14

TABLE 6. *Departure in percent of the maximum value of a given index from that occurring in the plane of best average definition for long lines selected by graphical analysis*

Results are shown for $\sqrt{\overline{R_\beta T_\beta}}$, AWAR, and ADWAR for six lenses with each type of target for VF and SP emulsions.

Lens No.	VF			SP		
	$\sqrt{\overline{R_\beta T_\beta}}$	AWAR	ADWAR	$\sqrt{\overline{R_\beta T_\beta}}$	AWAR	ADWAR
	Long-line			Long-line		
1	4.0	2.6	0.6	3.4	3.0	0.0
2	2.5	1.0	.0	2.7	1.7	1.6
3	3.3	0.0	3.3	0.0	0.0	0.0
4	3.0	.0	0.0	.0	.0	.0
5	0.9	.0	.0	.0	.0	.0
6	.0	.0	.0	.0	.0	.0
	Short-line			Short-line		
1	0.0	0.0	0.0	0.0	1.5	1.8
2	.0	.0	.0	.0	1.0	0.0
3	.0	.0	.0	.0	0.0	.0
4	.0	.0	.2	.0	.5	.0
5	.5	.0	.0	.0	1.2	1.6
6	.0	3.6	.5	.0	0.0	0.0
	Annulus			Annulus		
1	0.0	3.9	1.9	0.0	0.0	0.0
2	6.5	0.0	0.0	3.8	6.1	6.7
3	13.0	.0	.0	0.0	0.8	0.0
4	3.6	.5	.5	.0	.0	.0
5	0.0	.0	.0	.0	.0	.0
6	.0	.0	.0	.0	.0	.0

the value of resolving power for a different type pattern is known. The relative magnitudes of $\sqrt{R_\beta T_\beta}$ at a given value of β for the various type patterns are significantly different for the different lenses. Certain generalizations can, however, be drawn. The values of $\sqrt{R_\beta T_\beta}$ in general are highest for the long-line and lowest for the annular patterns with the values for the short-line patterns being usually slightly lower than those for the long-line patterns.

4.2. Use of Indices as Rating Factor

In table 7, the relative ratings of lenses 1 through 6 are shown for each of the various conditions of test. It is, at once, apparent that the order of merit is not the same for the various patterns. This variation in the order of merit emphasizes the fact that indices such as these ought not to be used to differentiate between lenses unless the differences are of the order of 10 percent or more. Thus, in considering the relative merits of these lenses, 1, 2, and 3 are usually more favorably rated than 4, 5, and 6. However, it would be difficult to arrange 1, 2, and 3 in order of decreasing performance. Likewise, it would be difficult to arrange 4, 5, and 6 in proper order. Most of this difficulty arises from the relatively small differences in the particular index.

4.3. Relative Magnitudes of Values of Resolving Power

a. For Three Types of Test Pattern

It was evident at a fairly early stage in the investigation that there is no clear-cut ratio existing between

TABLE 7. *Rating of six lenses in order of merit for VF and SP emulsions*

The ratings are made for each of three targets on the basis of the relative magnitude of the three indices $\sqrt{\overline{R_\beta T_\beta}}$, AWAR, and ADWAR.

Lens No.	VF emulsion			SP emulsion		
	Rating based on			Rating based on		
	$\sqrt{\overline{R_\beta T_\beta}}$	AWAR	ADWAR	$\sqrt{\overline{R_\beta T_\beta}}$	AWAR	ADWAR
	(a) Long-line			(a) Long-line		
1	1	2	1	4	1	1
2	1	1	1	2	4	3
3	4	3	3	1	3	3
4	6	6	6	2	5	5
5	5	4	5	4	6	6
6	3	4	4	4	1	1
	(b) Short-line			(b) Short-line		
1	3	2	2	1	1	1
2	1	1	1	2	1	2
3	1	2	3	2	1	2
4	4	5	4	2	4	2
5	5	5	4	6	6	6
6	5	4	6	5	5	5
	(c) Annulus			(c) Annulus		
1	2	3	2	5	4	1
2	3	1	2	5	4	3
3	5	3	4	1	2	3
4	6	6	6	1	2	3
5	1	1	1	1	1	1
6	3	3	4	1	4	3

values of resolving power obtained for one type of pattern and those obtained for another type of pattern. This is at once apparent from consideration of the various curves in figures 3 through 9. Even on axis, there are variations which might be in part a consequence of the large interval between successive patterns but are more likely a consequence of the inherent uncertainty in locating the limit of resolution. However, some idea of the approximate magnitude of the values of resolving power for the various patterns relative to one another can be gained from consideration of the relative magnitudes of the various indices. Ratios have been computed for long-line versus short-line, short-line versus annulus, and long-line versus annulus for the various indices for VF and SP emulsions. These ratios are shown in table 8. From these values, one may infer that the values of resolving power obtained with long-line patterns using SP emulsion are likely to be 18 percent higher than values obtained with short-line patterns. In a similar manner, values for short-line patterns may be expected to be 45 percent higher than for annular patterns. Long-line patterns may be expected to yield values approximately 70 percent higher than annular patterns.

b. For Two Types of Emulsion

Of the two emulsions for which values are reported in this study, the VF emulsion has been used at the Bureau over a long period of years in making the negatives used in evaluating equivalent focal length, distortion, and resolving power of lenses. It has been eminently satisfactory for this purpose, and

TABLE 8. *Ratio of values of given indices as a function of pattern type*

Values are given (a) for the average ratio (L/S) of the values of a given index for long lines and the same index for short lines. Values are also given (b) of the average ratio (S/A) of the values of a given index for short lines and the same index for annulus type patterns. Values are also shown of the average ratio (L/A) of the values of a given index for long lines and the same index for annulus type patterns. Values are shown for both VF and SP emulsions together with the probable error of a single observation (PE_s) for each ratio.

Index	(a) Values of L/S and PE_s for emulsion			
	VF		SP	
	L/S	PE_s	L/S	PE_s
$\overline{R_\beta T_\beta}$	1.09	±0.04	1.11	±0.05
AWAR	1.11	.05	1.22	.06
ADWAR	1.08	.03	1.20	.03
	(b) Values of S/A and PE_s for emulsion			
	S/A	PE_s	S/A	PE_s
$\sqrt{\overline{R_\beta T_\beta}}$	1.66	±0.11	1.40	±0.09
AWAR	1.82	.12	1.51	.11
ADWAR	1.70	.11	1.43	.07
	(c) Values of L/A and PE_s for emulsion			
	L/A	PE_s	L/A	PE_s
$\overline{R_\beta T_\beta}$	1.80	±0.07	1.55	±0.07
WAR	2.02	.14	1.84	.18
DWAR	1.84	.12	1.72	.08

numerous specifications involving resolving power of lenses are based upon determinations made therewith. In recent years, there has been a tendency on the part of many engaged in evaluating lens performance to use emulsions similar to the SP emulsion used in this study. This practice is justified on the grounds that with this emulsion, the test conditions more nearly approximate the conditions of use.

Because the two types of emulsion are used for much the same purpose, it is of interest to compare the values of resolving power under like conditions for the two emulsions. It is clear from the curves shown in figures 3 through 9, that there is no clear-cut ratio that is valid for all conditions of focus and for all angular separations from the axis. Accordingly, the ratios existing between the values of the various indices for the two emulsions were determined and are shown in table 9. From this table, one may infer that the values of resolving power obtained with VF emulsion are likely to be 84 percent higher for long-line patterns, 98 percent higher for short-line patterns, and 66 percent higher for annular patterns than are obtained for SP emulsion.

TABLE 9. *Ratio* q *of the value of a given index in plane of best average definition for VF emulsion to the value of the same index in the plane of best average definition for SP emulsion for three types of test patterns*

Values of the ratio are shown $\sqrt{\overline{R_\beta T_\beta}}$, AWAR, and ADWAR obtained with long-line, short-line, and annular patterns. These are average values for six lenses, and the probable error of a single determination of each ratio is also shown.

Index	Ratio for patterns comprised of					
	Long lines		Short lines		Annuli	
	q	PE_s	q	PE_s	q	PE_s
$\sqrt{\overline{R_\beta T_\beta}}$	1.94	±0.12	1.97	±0.09	1.66	±0.09
AWAR	1.80	.14	1.97	.11	1.64	.11
ADWAR	1.79	.11	1.99	.11	1.68	.13

5. Conclusion

In this study, the values of the resolving power of six lenses were measured using three types of target pattern and two emulsions. Analysis of the results of measurements leads to the following conclusions.

(1) For either of the two emulsions used, the same plane of best average definition is usually selected by graphical analysis for long-line, short-line, and annulus target patterns.

(2) The plane of best average definition located with the VF emulsion is generally identical with that located with SP emulsion.

(3) While small variations in the location of the focal plane occur with type of index used, these variations are generally sufficiently small that the plane selected by anyone of the three indices is likely to prove satisfactory.

The authors express their appreciation to other members of the staff of the National Bureau of Standards for assistance during this work and in particular to E. C. Watts who prepared the illustrations.

6. References

[1] Rept. of Comm. I, Intern. Soc. Photogrammetry, Specification of methods of calibrating photogrammetric cameras and measuring their resolution, image illumination, and veiling glare, Photogrammetria **X,** 85 (1953–1954).

[2] Military Standard MIL–STD–150–A; Photographic lenses.

[3] British Standard 1613 : 1949, Resolving power of lenses for cameras and enlargers.

[4] L. E. Howlett, Photographic resolving power, Can. J. Research **24** [A], 15–40 (1946).

[5] F. H. Perrin and J. H. Altman, Studies in the resolving power of photographic emulsions, VI. The effect of the type of test pattern and luminance ratio in the test object, J. Opt. Soc. Am. **43,** 780 (1953).

[6] P. Hariharan, Resolving power of photographic emulsions **46,** 315 (1956).

[7] F. E. Washer and I. C. Gardner, Method for determining the resolving power of photographic lenses, NBS Circ. 533 (1953).

[8] I. C. Gardner and F. A. Case, Precision camera for testing lenses, J. Research NBS **18,** 449 (1937) RP984.

[9] F. E. Washer and F. A. Case, Calibration of precision airplane mapping cameras, J. Research NBS **45,** 1 (1950) RP2108; Photogrammetric Eng. **XVI,** 502 (1950).

[10] F. E. Washer and W. P. Tayman, Location of the plane of best average definition for airplane camera lenses, Photogrammetric Eng. (June 1960).

[11] E. W. H. Selwyn and J. L. Tearle, The performance of aircraft camera lenses, Proc. Phys. Soc. **LVIII** (493) 1946.

(Paper 64C3–39)

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 50, NUMBER 10 OCTOBER, 1960

Four-Color Achromats and Superchromats

ROBERT E. STEPHENS
National Bureau of Standards, Washington 25, D. C.
(Received May 31, 1960)

Materials suitable for the construction of achromatic lenses corrected for four wavelengths are selected by the use of three graphs which are the three orthographic projections of a three-dimensional graph in P, Q, and V; where $V=(n_d-1)/(n_F-n_C)$, $P=(n_h-n_C)/(n_F-n_C)$, and $Q=(n_{1.014}-n_C)/(n_F-n_C)$. One set of three glasses, Schott PKS-1, F-1, and KzFS-4, has been selected, from which a four-color achromat has been designed. Analysis at 12 wavelengths shows this design to have negligible residual chromatic aberration. It is consequently a *superchromat* in accordance with Herzberger's prediction.

THE preparation of this paper was originally projected as a continuation and extension of the principles presented in the author's earlier paper[1] on the design of three-color achromats. It was considerably accelerated, however, by Herzberger's suggestion[2] that a lens would have the same focal length for all wavelengths in the spectral range approximately from 0.365 μ to 1.014 μ if it were designed to be chromatically corrected for four wavelengths in that range. Herzberger stated some of the conditions necessary for materials to be suitable for the construction of such a system. However, he did not state all the conditions and did not show that a suitable set of materials exists.

Herzberger has found that the index of refraction of all glasses and other optical materials so far examined can be described by an equation of the form

$$n_\lambda = A + B\lambda^2 + \frac{C}{(\lambda^2-0.028)} + \frac{D}{(\lambda^2-0.028)^2}.$$

If this were rigorously true, then it would follow that a lens system so corrected as to have the same focal length at four wavelengths would be automatically corrected at all wavelengths. Such a lens is called a *superchromat* by Herzberger. This author would prefer to call it *super achromat* in the belief that its properties are more nearly connoted by the latter name.

The formal statement of the conditions to be fulfilled for a lens of four elements to be chromatically corrected for four colors is

$$\begin{aligned} \phi_1+\phi_2+\phi_3+\phi_4&=\Phi \\ (\phi_1/V_1)+(\phi_2/V_2)+(\phi_3/V_3)+(\phi_4/V_4)&=0 \\ (\phi_1/U_1)+(\phi_2/U_2)+(\phi_3/U_3)+(\phi_4/U_4)&=0 \\ (\phi_1/W_1)+(\phi_2/W_2)+(\phi_3/W_3)+(\phi_4/W_4)&=0, \end{aligned} \tag{1}$$

where ϕ_j is the dioptric power of the jth lens element, $V_j=[(n_d-1)/(n_F-n_C)]_j$. U and W are similar to V but with n_F replaced by the indices for other wavelengths. For a solution to Eqs. (1) to exist, it is necessary that the determinant of the coefficients not be zero. Mathematically, that is a sufficient condition but practically the value of the determinant must be of considerable magnitude. I am not prepared to set a definite limit, but a solution in which $(\phi_1+\phi_2)=100\phi$, $(\phi_3+\phi_4)=-99\phi$, certainly cannot be called a practical solution as the curvatures would constitute a much more serious restriction to the focal length and aperture than a modest residual of secondary or tertiary chromatic aberration.

It is also possible that a three-element system be chromatically corrected for four wavelengths. The equations governing such a system are

$$\begin{aligned} \phi_1+\phi_2+\phi_3&=\Phi \\ (\phi_1/V_1)+(\phi_2/V_2)+(\phi_3/V_3)&=0 \\ (\phi_1/U_1)+(\phi_2/U_2)+(\phi_3/U_3)&=0 \\ (\phi_1/W_1)+(\phi_2/W_2)+(\phi_3/W_3)&=0. \end{aligned} \tag{2}$$

It is not possible to satisfy these for arbitrary values

[1] R. E. Stephens, J. Opt. Soc. Am. 49, 398–401 (1959).
[2] Max Herzberger, Optica Acta (Paris) 6, 197–215 (1959).

of the U's, V's, and W's, but if

$$\begin{vmatrix} (1/V_1) & (1/V_2) & (1/V_3) \\ (1/U_1) & (1/U_2) & (1/U_3) \\ (1/W_1) & (1/W_2) & (1/W_3) \end{vmatrix} = 0, \tag{3}$$

it may be possible. A further condition is that

$$\begin{vmatrix} 1 & 1 & 1 \\ (1/V_1) & (1/V_2) & (1/V_3) \\ (1/U_1) & (1/U_2) & (1/U_3) \end{vmatrix} \neq 0. \tag{4}$$

These two conditions are mathematically necessary and sufficient, but again the value of determinant (4) must be of considerable magnitude for a practical solution to exist.

Likewise, it is conceivable that two elements be sufficient. The equations for such a system are

$$\begin{aligned} \phi_1 + \phi_2 &= \Phi \\ (\phi_1/V_1) + (\phi_2/V_2) &= 0 \\ (\phi_1/U_1) + (\phi_2/U_2) &= 0 \\ (\phi_1/W_1) + (\phi_2/W_2) &= 0. \end{aligned} \tag{5}$$

These could be satisfied if

$$(V_1/V_2) = (U_1/U_2) = (W_1/W_2),$$

and $|V_1 - V_2| \neq 0$. Again, a practical solution demands that $|V_1 - V_2|$ be sufficiently large (say greater than five).

To attempt to determine the suitability of various combinations of glasses for four-color achromats by means of the determinants would require the evaluation and comparison of some 10 million or more of such determinants even for just the glasses contained in the Schott catalog. Obviously a more expeditious procedure is desirable. One such procedure, which is used by this author, consists of a logical extension of the methods presented in the earlier paper.[1] The third and fourth equations of (1), (2), and (5) are modified by the following substitutions

$$1/U = P/V \quad \text{and} \quad 1/W = Q/V.$$

It has been shown in the earlier paper that the suitability of glasses for construction of achromats corrected at three wavelengths can be determined from a graph on which the characteristics P and V of a glass are plotted as the coordinates of a point. On such a plot, the straight line through the points for two glasses is the locus of the effective values, $\bar{P}$ and $\bar{V}$, of all two-element lens systems made up of these two glasses.

If two glasses can be combined in such proportions that the value of $\bar{P}$ matches the P of a third glass and the difference $(\bar{V} - V)$ is of practical magnitude (say greater than five), a three-color achromat for wavelengths C, F, and λ_P can be constructed. Now Q is the same kind of quantity as P except that the third wavelengths are different. On a graph of Q and V, one could determine the suitability of glasses for achromatism at wavelengths C, F, and λ_Q. If for three glasses $\bar{P}_{12} = P_3$, $\bar{Q}_{12} = Q_3$, and $\bar{V}_{12} \neq V_3$, we would have glasses with which a three-element lens corrected for four wavelengths could be made (if $|\bar{V} - V_3|$ were large enough).

The two graphs PV, and QV, are the orthographic projections of a three-dimensional array in which the properties, P, Q, and V, of a glass are the coordinates of a point. The third orthographic projection of this array gives a third graph on which the coordinates are P and Q. This corresponds to one of Herzberger's preferred graphs (P^*,P^{**}).

The line through two points on the P, V graph and the line through the points for the same two glasses on the Q, V graph are obviously the projections of a line in PQV-space which is the locus of effective values $\bar{P}$, $\bar{Q}$, and $\bar{V}$, for all possible binary combinations of the two glasses in question. Consequently, the third or PQ projection of this line is the locus of all possible values of $\bar{P}$ and $\bar{Q}$ for the same binary combinations. If the point for a third glass were found to lie on the line of two other glasses, then presumably a four-color achromat of three elements could be designed. This condition and its analytical equivalent,

$$\begin{vmatrix} 1 & 1 & 1 \\ P_1 & P_2 & P_3 \\ Q_1 & Q_2 & Q_3 \end{vmatrix} = 0,$$

are those stated by Herzberger. Obviously, this condition is a necessary one, but it is not sufficient. A further necessary condition is that $|\bar{V}_{12} - V_3|$ be sufficiently large (say greater than five). The two conditions together are both necessary and sufficient to guarantee that a four-color achromat can be made of three glasses.

The PQV-characteristics of seventeen glasses from Schott catalog number 350 and of fluorite are plotted in Fig. 1. The quantities are defined as given here:
For P, the wavelength of the mercury line h (0.4047 μ) was chosen rather than the line at 0.365 μ, because, some of the glasses are evidently too absorbent to be measured at the latter wavelength and in particular, the index for F-16 is not given for that wavelength. It was desired to include F-16 because of its very exceptional refractive characteristics.

The glass F-16 which has such exceptional characteristics for three-color correction when the third color is in the violet region of the spectrum is seen to be not as good as an ordinary flint when correction in the red and infrared are considered. The glasses KzFS-4 and PKS-1 are seen by their positions on the PV and QV graphs to be really exceptional. Consequently, a search was made for a flint glass such that the three points on the PQ graph for KzFS-4, PKS-1 and a flint would lie on a straight line. The flint F-1 was found to be satisfactory.

On the PV and QV graphs, straight lines were drawn

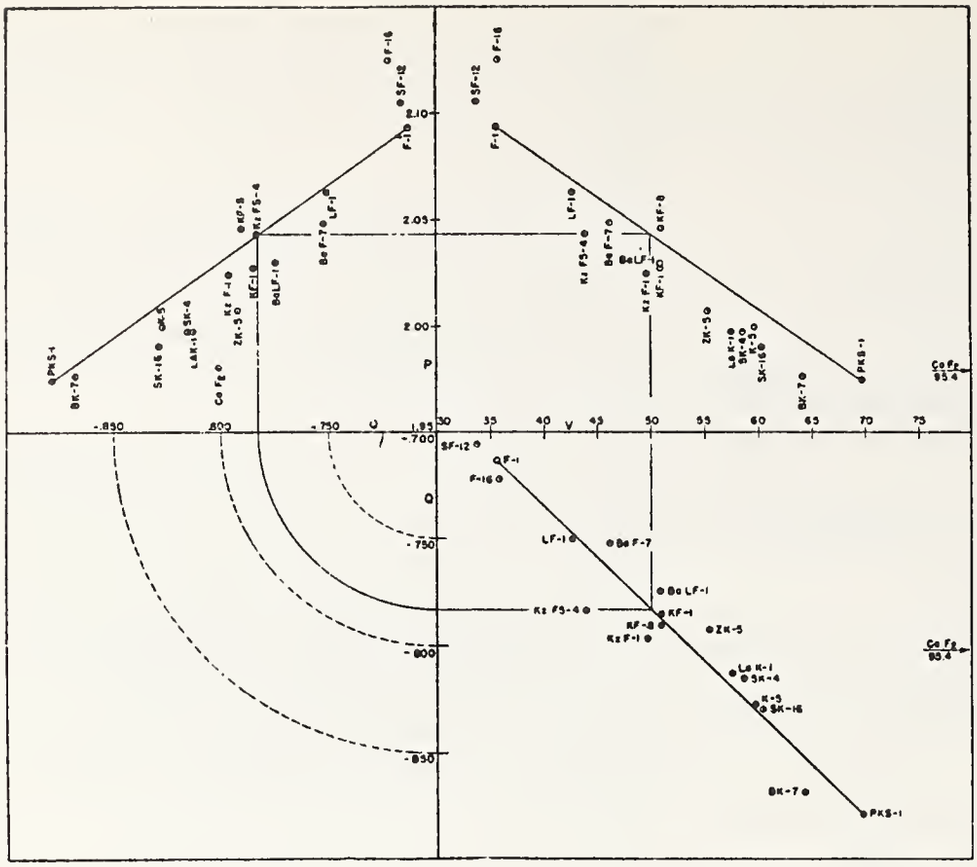


FIG. 1. $V=(n_d-1)/(n_F-n_C)$, $P=(n_\lambda-n_C)/(n_F-n_C)$, $Q=(n_{1.014}-n_C)/(n_F-n_C)$.

connecting the F-1 and PKS-1 points. On the PV line, the point having the same value of P as that of KzFS-4 was located and the corresponding value of V, which will be called $\bar{V}_{12}$, was found; thus $\bar{V}_{12}=50.0$. Approximately the same value would be found by the QV graph. The value of V_3 for KzFS-4 is 43.91. Thus, $(\bar{V}_{12}-V_3)=6.09$; this is not as large as we would like but is not completely impractical. By the procedures of the earlier paper, the dioptric powers of the three elements, for a total power of one diopter, were computed and found to be (in diopters)

$$\phi_1=4.83 \text{ (PKS-1)},$$
$$\phi_2=3.38 \text{ (F-1)},$$
$$\phi_3=-7.21 \text{ (KzFS-4)}.$$

The corresponding curvatures, $(1/R_1-1/R_2)$, were found to be (also in diopters)

$$K_1=9.337,$$
$$K_2=5.400,$$
$$K_3=-11.754.$$

The chromatic differences of dioptric power $(\phi_\lambda-\phi_C)$ were computed for all eleven wavelengths (other than C) for which indices are given in the Schott catalog. The residual for $\lambda=0.365\,\mu$ is -0.00022; the others are so small and unsystematic that they are not significant. The system is thus shown to be almost perfectly achromatic from $\lambda=0.4047\,\mu$ to $\lambda=1.014\,\mu$. This design verifies Herzberger's prediction and, in accordance with his terminology, is a *superchromat*.

If no suitable combination of three glasses could be found, one might expect to find a set of four glasses that could be combined to produce a *superchromat*. If one examines the points on the PQ graph, he will see that the diagonals of the quadrilateral formed by any four points intersect unless three of the points are on a straight line. This intersection represents the common effective values of P and Q for combinations in the proper proportions of glasses 1 and 2, and 3 and 4. If these values of P and Q are read, we can then go to either the PV, or the QV graph to determine suitability.

For example, on the PV graph draw a straight line through the points for glasses 1 and 2; draw another straight line through the points for glasses 3 and 4. At the common value of P_{12} and P_{34} found above, the values $\bar{V}_{12}$ and $\bar{V}_{34}$ are found. If $|\bar{V}_{12}-\bar{V}_{34}|$ is large enough (say five or larger) a four-element *superchromat* could be designed from these glasses. Since the three-glass set used above gave a difference $\bar{V}_{12}-V_3=6.09$, a set of four for which $|\bar{V}_{12}-\bar{V}_{34}|$ exceeds 6.09 would have to be found in order to get a better result than was obtained above with three. The author has not found any such set although his cursory search by no means precludes the possibility that such exists.

Study of the PV and QV graphs leads to the belief that pairs of glasses may be selected from which binary achromats with much reduced secondary chromatic aberration can be designed. For example, KF-8 and KzFS-4 have a V-difference of 7.00, which is a little

better than was obtained above with three glasses, and the P- and Q-differences are both small. Since the slopes of the lines between the two glasses on these graphs indicate the residual chromatic aberrations at the respective wavelengths of P and Q, this binary combination would be almost as good as the *superchromat*.

It is evident that superior achromats (binary) can be achieved by the use of fluorite as one of the components and a zinc or lanthanum glass as the other. It may also be that a glass exists, or could be made, which, with fluorite, would make a two-element *superchromat*.

CONCLUSIONS

It has been shown that, for the purpose of selecting glasses for achromatic lenses of all types, the representation of the characteristics of glasses on three graphs (PV, QV, and PQ), drawn in projective relation, is most expeditious. These graphs facilitate the design of achromats, three-color achromats, and superchromats. The accuracy with which the proportions of the various glasses can be determined by the use of these graphs, each in ordinary letter size, is all that is justified by the accuracy of index data in catalogs. The fourth characteristic of glass, the index of refraction for some wavelength in the region, although not irrelevant, is almost taken care of in the preliminary selection by its inclusion implicitly in V. No separate graphical representation of it is necessary. However, the customary graph of n_d vs V is useful to a designer, and is considered by this author to be of more significance in design work than one of n_d vs $(n_F - n_C)$.

Herzberger's prediction of the possibility of the *superchromat* has been confirmed by analysis through the spectrum of one design consisting of glasses selected by the methods presented in this paper.

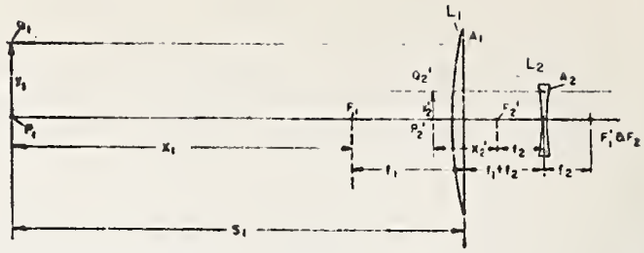


FIG. 1.

Magnifications of a Telescope

ROBERT E. STEPHENS
*National Bureau of Standards, Washington, D. C.**
(Received December 2, 1960)

GEOMETRICAL optics is a subject that seems to give students more difficulty in real comprehension than most of the other divisions of classical physics. The full explanation of this is not known to this writer, but it is suspected that several generations of textbook writers have, in the laudable attempt at simplification, left out so much that what remains is more a subject for committing to memory than for understanding on the basis of fundamentals.

The effect of oversimplification upon comprehension is bad even if the presentation is correct and results in valid conclusions; but if a derivation is made incorrectly by the use of faulty reasoning the student's failure to comprehend can hardly be surprising.

A well-known text on optics contains an example of oversimplification and faulty mathematical procedure which gives an incorrect result. The problem is also interesting in itself because of an apparent paradox. The problem is given in the book in question as an example of the successive application of the thin-lens equations, which relate image and object distances and magnification to a system of two thin lenses in sequence. The chosen example consists of a converging lens followed by a diverging lens in the arrangement of the Galilean-type telescope. It certainly is an unfortunate choice as an example because of the peculiarities of that system. In addition, the mathematical treatment is faulty because of an incorrect estimate of the limiting value of an indeterminate ratio, and results in an incorrect answer. The procedure used in the book would, if completed correctly, give the lateral linear magnification. The fact that the faulty procedure gives the correct expression for the angular magnification is purely accidental. It certainly is unsatisfactory pedagogically.

The following rigorous analysis of this problem, which is simple and free from indeterminate forms, brings out the somewhat surprising facts about the position, size, and apparent size of the image formed by a telescopic system.

Analysis. Let two thin lenses, whose focal lengths are f_1 and f_2, be axially aligned and separated by a space $t=f_1+f_2$. This spacing places the first focus of the second lens in coincidence with the second focus of the first lens. The focal length of the first lens (f_1) is positive; that of the second (f_2) may be either positive or negative. (Two negative lenses cannot be placed so that $t=f_1+f_2$, because t must be positive to give a physically realizable system.) A typical telescopic system of the Galilean type, in which the second lens is a diverging one, is shown schematically in Fig. 1. The problem is to find the position of the image, the linear magnification, and the angular magnification of the system. In general these are functions of the object distance. In this discussion we are especially concerned with their limiting values as the object distance approaches infinity.

The formulas used in solving the problem are[1] the usual ones

$$(1/s)+(1/s')=1/f \tag{1}$$

and

$$m=-s'/s, \tag{2}$$

where s and s' are the axial object and image distances, respectively, measured from the vertex of the lens, f is the focal length, and m is the lateral linear magnification; and the Newtonian forms

$$xx'=f^2 \tag{3}$$

and

$$m=-f/x=-x'/f, \tag{4}$$

where x and x' are the object and image distances, measured from the first and second foci, respectively, instead of from the vertex.

Applying Eqs. (3) and (4), for the first lens we obtain

$$x_1=f_1^2/x_1' \tag{5}$$

and

$$m_1=-x_1'/f_1. \tag{6}$$

For the second lens we obtain

$$x_2'=f_2^2/x_2 \tag{7}$$

and

$$m_2=-f_2/x_2. \tag{8}$$

To get the magnification of the system we combine Eqs. (6) and (8) obtaining

$$m=m_1m_2=(-x_1'/f_1)(-f_2/x_2);$$

but since the spacing is such that the second focus of the first lens is coincident with the first focus of the second lens, $x_2=-x_1'$. Thus we obtain

$$m=-f_2/f_1. \tag{9}$$

This is the correct expression for the lateral linear magnification of a telescopic system. It is a constant, independent of the object and image distances. Thus there is no indeterminacy as the object distance approaches infinity. This may be seen by reference to Fig. 1. Since the ray $\langle Q_2'A_2\rangle$ extended is the image of ray $\langle Q_1A_1\rangle$ extended, both of which are parallel to the axis, the image of any point on $\langle Q_1A_1\rangle$ or its projection must lie somewhere along $\langle Q_2'A_2\rangle$ or its projection. An object whose tip is on $\langle Q_1A_1\rangle$ has the length y_1, and its image has the length $-y_1f_2/f_1$.

From Eqs. (5) and (7) we find that

$$x_2'/x_1=(f_2^2/x_2)(x_1'/f_1^2),$$

but, as before $x_2=-x_1'$; consequently,

$$x_2'/x_1=-(f_2/f_1)^2. \tag{10}$$

Thus the ratio of the image distance to the object distance (each measured from the appropriate focus) is a constant, not equal to -1 unless $f_2=f_1$.

A telescopic system cannot be described by the Gaussian thick lens constants, because the focal lengths are infinite and there are no planes of unit magnification. Thus the positions of the object and the image cannot be given with reference to principal planes. Equation (10) is probably the simplest and most satisfactory relation between the object and image distances for the two-component lens system.

Since $s=x+f$ we can see that the ratio s_2'/s_1 approaches (10) as the x's and s's approach infinity and that in the limit

$$\left[\frac{s_2'}{s_1}\right]_{s_1=\infty}=-\left[\frac{f_2}{f_1}\right]^2. \tag{11}$$

It may be more satisfying to work this out in another way, in which no indeterminacy arises. Using Eq. (1) we obtain for the first lens

$$1/s_1=(s_1'-f_1)/s_1'f_1$$

and for the second lens

$$s_2'=s_2f_2/(s_2-f_2).$$

Then

$$\frac{s_2'}{s_1}=\frac{(s_1'-f_1)}{s_1'f_1}\frac{s_2f_2}{(s_2-f_2)}.$$

But $s_2=t-s_1'=f_1+f_2-s_1'$; consequently

$$\frac{s_2'}{s_1}=\frac{(s_1'-f_1)f_2-f_2^2}{s_1'f_1}.$$

When $s_1=\infty$, $s_2'=\infty$ also, and $s_1'=f_1$; the ratio s_2'/s, remains determinate and is

$$\left[\frac{s_2'}{s_1}\right]_{s=\infty}=-\left(\frac{f_2}{f_1}\right)^2.$$

To analyze the performance of a telescopic system, applied in the usual way for observing distant objects, the customary textbook treatment defines the magnification as the ratio of the angular subtense of the image to the angular subtense of the object. It may alternatively be defined as the ratio of the size of the image on the retina of the eye when observing through the telescope to the size of the retinal image when observing the distant object with the unaided eye. The fact that one eye position is used to compute the angular subtense of the image and a different one is chosen for computing the subtense of the object when viewed unaided causes the textbook writer no logical difficulty. Neither is it likely to cause the instructor any trouble. But it can and often does confuse the student. Neglecting this difference in eye positions does lead, in this case, to the correct limiting value for the magnification.

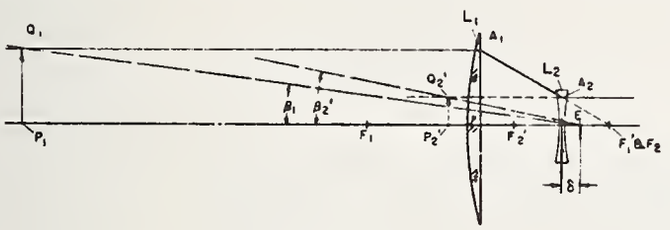


FIG. 2.

It is possible, with only slight additional complication of the algebra, to eliminate completely the logical problem arising from different viewpoints and at the same time eliminate the necessity for evaluating an indeterminate expression. We choose a suitable fixed eye position from which to observe, both with and without the telescope; such a point is E in Fig. 2. This point is located at any convenient position behind the eyepiece when the telescope is present. Let its position be specified by the distance δ, measured from the second lens. The object distance measured from this point (as observed with the telescope removed) is

$$l_1=x_1+2f_1+f_2+\delta.$$

The image distance from this same point (telescope present) is

$$l_2'=x_2'+f_2-\delta.$$

The angle subtended by the object is given by

$$\tan\beta_1=-\frac{y_1}{l_1}=-\frac{y_1}{x_1+2f_1+f_2+\delta};$$

and the angle subtended by the image is given by

$$\tan\beta_2'=\frac{y_2'}{l_2'}=-\frac{f_2}{f_1}\frac{y_1}{(x_2'+f_2-\delta)}.$$

The angular magnification, defined by

$$M=\tan\beta_2'/\tan\beta_1,$$

is

$$M=\frac{f_2}{f_1}\frac{(x_1+2f_1+f_2+\delta)}{(x_2'+f_2-\delta)}.$$

This expression is indeterminate for $l_1=l_2'=x_1=x_2'=\infty$; so we modify it by using the values for x_1 and x_2' given by Eqs. (5) and (7); thus

$$M=-\frac{f_2}{f_1}\frac{[(f_1'/x_1')+2f_1+f_2+\delta]}{[(f_2^2/x_2)+f_2-\delta]}.$$

As before, $x_2=-x_1'$; consequently

$$M=\frac{f_2}{f_1}\frac{[f_1^2+x_1'(2f_1+f_2+\delta)]}{[-f_2^2+x_1'(f_2-\delta)]}. \tag{12}$$

This expression remains determinate at $x_1=\infty$, as $x_1'=0$; therefore

$$M_\infty=(f_2/f_1)(-f_1^2/f_2^2)=-f_1/f_2. \tag{13}$$

We can analyze the peculiar action of a telescope by noting that although the image is smaller than the object it is much nearer the observer and thus appears magnified. If we remember the derivation we realize that the ratio in the brackets in Eq. (12) is equal to the ratio of the observation distances for object and image,

$$-\frac{[f_1^2+x_1'(2f_1+f_2+\delta)]}{[f_2^2-x_1'(f_2-\delta)]}=\frac{l_1}{l_2'}.$$

A telescope focused so that $t=f_1+f_2$ is normally used only for object distances that are very large compared to the length of the instrument, for which x_1' is very small. Consequently, the distance ratio is always

$$l_1/l_2'=-(f_1/f_2)^2, \text{ approximately.}$$

The limiting value given by Eq. (11) and the angular magnification [Eq. (13)] are quite independent of the value of δ. We could even have chosen different values of δ in the numerator and denominator of Eq. (12). This latter condition is usually the case in textbook analysis; also usually without adequate explanatory text in justification.

As a numerical example the image formed by a telescope with a power of 10× has linear dimensions only 1/10 of those in the object; however, the image is only 1/100 as far away and thus appears 10 times as large.

It is to be noted that for a remote object the two magnifications of a telescope, linear and angular, are reciprocals. This is made use of in evaluating the power of a telescope (the angular magnification) by measuring its linear magnification. Since any object at any distance may be used we choose the most convenient one, the rim of the objective, which is the entrance pupil. The image of this is the exit pupil. The diameter of the exit pupil divided by the diameter of the rim of the objective is the linear magnification and the reciprocal is the angular magnification. It is, of course, necessary to make sure that the rim of the objective is the actual entrance pupil. It is, in all well-designed telescopes.

Although the longitudinal magnification is of negligible importance, it is included for completeness. If two axial points are separated by a distance Δx_1, their images will be on the axis, separated by a distance $\Delta x_2'$. The ratio, $\Delta x_2':\Delta x_1$, is by definition the longitudinal magnification. It is easily deducible from Eq. (10) that

$$\Delta x_2'/\Delta x_1=-(f_2/f_1)^2.$$

The results deduced in this paper have been known for a long time; they seem, however, to have been appreciated only by a few specialists. This manuscript was stimulated by the error found in the well-known textbook, and was submitted for publication in the hope that it might help a little toward the understanding of optics by physics students in the future.

* Consultant in optics at the National Bureau of Standards and instructor, NBS Graduate School.

[1] The notation and sign conventions agree with those of A. C. Hardy and F. H. Perrin in *The Principles of Optics* (McGraw-Hill Book Company, Inc., New York, 1932).

518-743 O - 73 - 37

Reprinted from
PHOTOGRAMMETRIC ENGINEERING
March
1961

*Measurement of Contrast in the Aerial Image**

FRED W. ROSBERRY,
Physicist, National Bureau of Standards

ABSTRACT: *The quality of a lens is evaluated by examining the image it produces and comparing it with the object or target of which the image is made. An instrument has been developed which scans the aerial image directly with a slit and photomultiplier tube. The resultant variations in luminosity in the image are recorded on a moving chart paper. This process substitutes a linear receiver or phototube for a non-linear receiver such as an emulsion. Results are shown in which contrast is plotted against image frequency for several f/values of a lens. A comparison is also made between image contrast and star images at several focal positions. A second comparison is made between measured contrast and best definition at several focal positions.*

ANALYZING the image of a lens as a means of judging its quality or capabilities might be compared to determining the response of an electronic amplifier. In each case, a known quantity is presented as input and the resultant output is analyzed and compared with the input. Both amplifier and lens are limited in response to the input signal. Beyond this limit in each case is only noise, either electrical from the amplifier, or light in the optical image.

In setting up a lens for test, an illuminated target is placed in front of the lens, within limits of distance and field angle, to produce an image.[1] This image exists only in space, to be seen and examined, it is usually registered by one of the following methods:

1. A diffuser is employed such as a ground glass which allows immediate examination of the image at the expense of considerable degradation of quality;
2. A photographic emulsion such as a film or plate is used in which a secondary chemical process is necessary to bring out the latent image for further examination;
3. The primary image is used as the object of a projection lens system such as a microscope objective. This approach, or the one permitting the image to fall directly on a slit, is referred to as the "aerial image."

To make the third method as objective as possible and to reduce the subjective or human element, a procedure for instantaneous read-out must be established to minimize the use of the eye. The phototube is an excellent tool and is used in this method since it responds to the amount of luminous energy falling on it. The response is the product of the area, the energy density, and the spectral response of the tube. This response varies as

* Presented at the Society's 26th Annual Meeting, Hotel Shoreham, Washington, D. C., March 23–26, 1960.

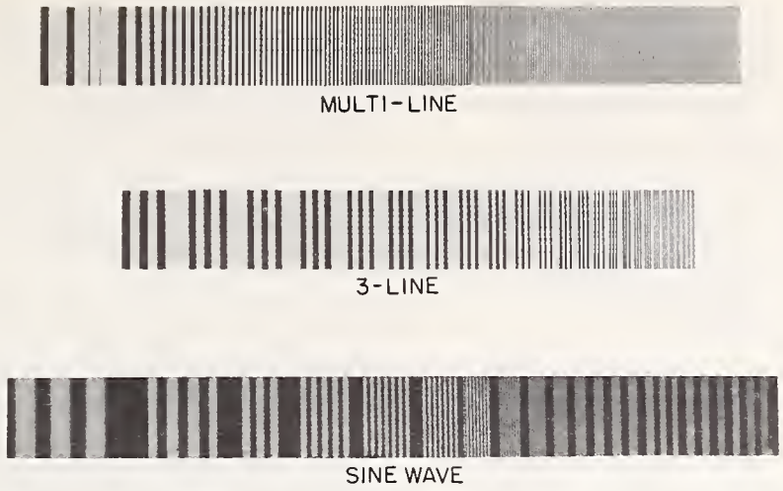


FIG. 1. Three types of resolution targets which have been used to determine image contrast.

the area of the light beam striking the tube is changed. The conventional way for controlling this variable is by use of a slit which renders constant the illuminated area, the source of illumination being constant. The tube is then required to read out only the variations of illuminance from one very small area of the image to another.

The entire image is scanned by movement of the image of the target relative to the phototube-slit combination. This movement, accomplished by lateral movement of the target, allows the receiving slit to be always in the same axial position of the lens under investigation.

The signal from the target must function as a known input to the system and the detail of the signal should have a geometric resemblance to the receiving slit shape. A target was therefore developed with detail in the form of parallel lines and spaces, commonly called the bar type of target.[2] The lines and spaces can be varied in width to produce coarse and fine patterns. The fine pattern in the target is in the range of high spatial frequency. The spacing from bar to bar generally varied between $\sqrt{2}$ and $\sqrt[6]{2}$.

A high progression rate target provides for a longer range of frequencies but gives less accuracy at locating a given frequency than a low progression rate target of the same overall length. The latter has a shorter range of frequencies but it provides more steps. In these two targets an opportunity also exists for variation in the arrangement of the line patterns, which are usually composed of groups of three lines with a constant line space width for each group.

Another type of target is one in which each line and space differs in width in a continuous manner from the adjacent lines and spaces. Target transparencies that are positive or negative in regard to line patterns and background can also be used. These transparencies may have clear lines with an opaque background or vice versa. In general the limits of resolution observed with positive and negative targets will not be the same. A target with a clear background will usually give a lower resolution limit due to the increased amount of flare light contributed by the background.

In choosing a target, other factors to be considered are the length of the line the target produces in the image plane, whether the line is longer or shorter than the receiving slit, and target contrast. Contrast is the difference in density between dark and light parts of the target detail. Three types of targets in general use are shown in Figure 1.

When the target is moved in a plane parallel to the image plane along a line at right-angles to the line pattern on the target, the image will correspondingly move in its plane. With the slit in the test instrument fixed, the target image passes over the slit. As the phototube and slit measures illuminance, a measure of the respective transmission of each line and space is made. If the transmissions of the lines and spaces are termed T_L and T_S respectively, the contrast is expressed as:

$$C = \frac{T_S - T_L}{T_S}.$$

If the results are measured in terms of density, the contrast is determined by the

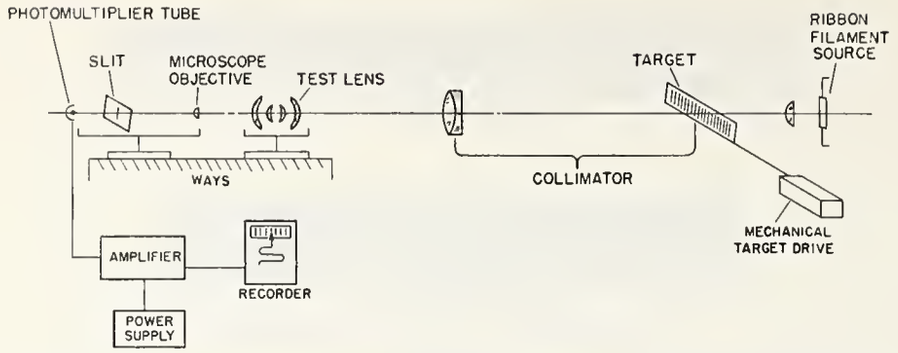


FIG. 2. A schematic layout showing arrangement of apparatus as used to determine image contrast.

difference of the line and space densities. Thus:

$$C = \Delta D = D_L - D_S.$$

In the newer types of rapid scanning electronic benches, which use a target whose transmittance fluctuates sinusoidally while scanning lines and spaces, a term called contrast transmission function or modulation is employed. This modulation is expressed as:

$$M = \frac{T_S - T_L}{T_S + T_L}.$$

Determinations are made of target modulation (M_t), and of the image produced by the lens under test (M_i). A simple ratio of the two yields a value called response:

$$R = \frac{M_i}{M_t}.$$

The equipment currently in use at the National Bureau of Standards generally satisfies the conditions required for analyzing an aerial image of a target. A schematic diagram of this equipment is shown in Figure 2. The source is a ribbon filament tungsten lamp focused by a condensing lens on the target located at the focus of a collimator. The lens under investigation is located on a nodal slide in the collimated beam. The nodal slide and pickup unit are mounted on separate saddles that ride on a length of ways which has been aligned with the optical axis established by the collimator.

Immediately behind the test lens is the pickup unit consisting of microscope, slit, and phototube. Associated equipment is used to amplify and record the output of the tube. As the receiving slit and source remain stationary on the axis of the system, and as only a small portion of the target is illuminated at one time, the entire target is scanned by one movement across the optical axis. A short length of ways guides this movement and a synchronous motor provides the power. The instrument is operated in such a manner that the image of the entire target is allowed to pass over the receiving slit.

Tangential and radial line patterns are produced from one target by rotating the lens in the nodal slide about two different axes: a vertical axis for tangential patterns, and a transverse horizontal axis for radial patterns.

Figure 3 shows two traces made by scanning the aerial image of the target. The trace forming the outer envelope—the one with the largest magnitudes of oscillations—was made at a focal position near best focus. The inner trace was made at a focal position 1.2 mm. inside of the first focal position. The lens under test was of about 150 mm. focal length.

The two traces have been started in phase at the low frequency end of the target. By following the inner trace it can be seen that resolution ends at the first null point where one half of a line is lost, and where apparent or spurious resolution starts. By using the outer trace as a control, it can be seen how in this area of spurious resolution, transmission is out of phase. A line is indicated where a space should be. The next null point indicates the loss of another one half of a line. This loss puts the lines and spaces back in phase, but a count shows one whole line and space missing.

The only measurable quantity in this type of investigation is the light flux. There are many parameters which can be given different values, resulting in a corresponding variation in light flux. Some of these parameters are associated with the lens, for example: location of focal-plane, off-axis location of receiving slit in the focal-plane, and f/value. Typical parameters associated with the target are: contrast, type (such as square wave or sine wave, and positive or negative), spatial frequency, and line orientation (radial or tangential). A source parameter is spectral range.

Figure 4 illustrates a plot in which contrast

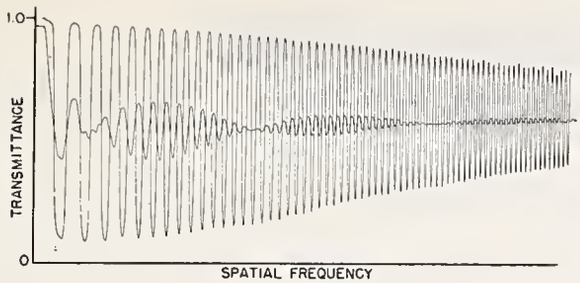


FIG. 3. Recorder trace showing the response in the aerial image at two focal positions, The outer trace was made near the best focus while inner trace was at a location 1.2 mm. nearer the lens.

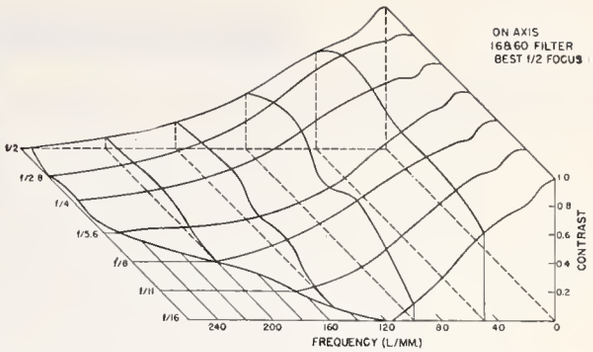


FIG. 4. A 3-dimensional plot illustrating the contrast at various spatial frequencies for seven lens stop openings.

has been plotted against two parameters, *f*/ value and spatial frequency. These data were obtained in testing a lens of 58-mm focal length, on axis, at the best *f*/2 focus, in a spectral range passed by Wratten filters 16 and 60 in combination. Figure 5 shows the results of two tests through focus. The first test measured image contrast of two spatial frequency line patterns at both radial and tangential line orientations, illustrated by the four curves. Results of the second test are shown at the top. In this test, the target was replaced by a pin hole, and photographs were made at the same focal positions of the resultant star images. These data were obtained from the image of a lens of 32 mm. focal-length at a stop opening of *f*/4.5, and at 15 degrees off axis.

The intersections of the radial (solid lines) and tangential (dotted lines) curves for each spatial frequency locate the approximate position of best focus. The star image in the top strip at focal position 0.5—this is almost at the intersection of the radial and tangential curves for the finer lines—is of such shape that a circle surrounding the entire pattern would be smaller than one around any of the other star images, thus showing the highest concentration of light. This result agrees with findings made in studies of spot diagrams.[3]

Figure 6 is a comparison between contrast, as measured in the aerial image, and best definition, as determined by photographs made in the same focal-plane. The horizontal row of traces (center) shows the raw data as

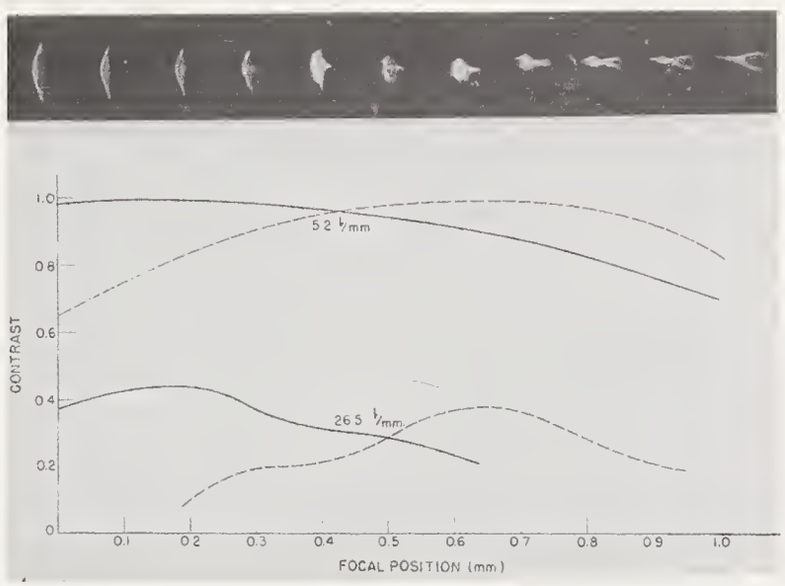


FIG. 5. Curves indicating contrast of tangential image patterns with solid lines and radial patterns with dashed lines at various focal positions. Star images are shown, for the various focal positions, at the top of the illustration.

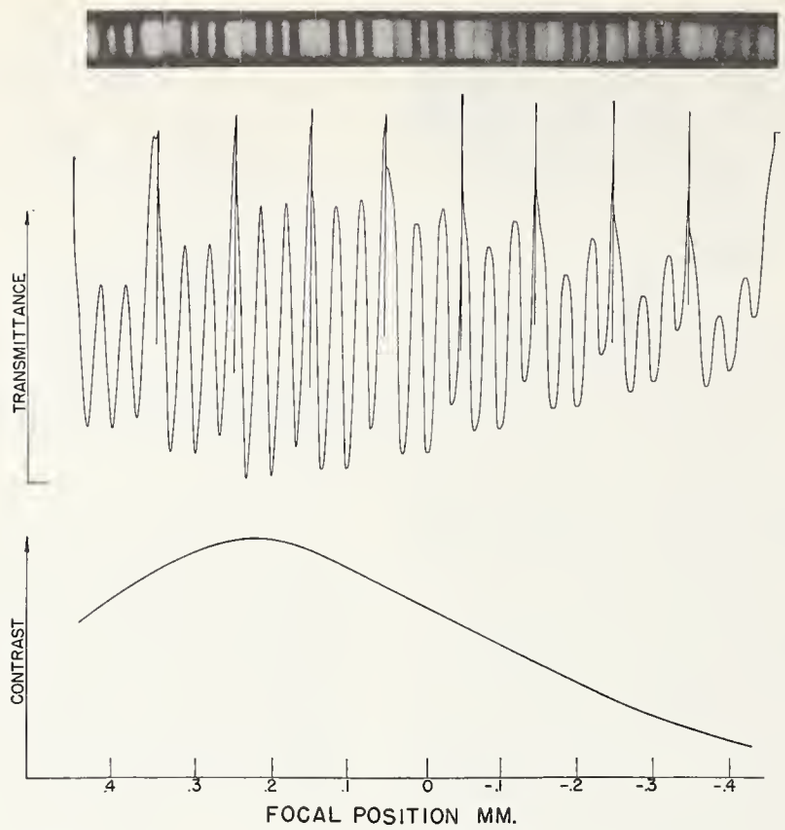


FIG. 6. A comparison on a common focal position scale of contrast (bottom curve) which was computed from recorder traces of middle row and photographs of the image of a 3-line target pattern (top row) made at the same focal positions.

traced by the recorder, and was made by repeated scans of the same three-line frequency pattern. The bottom curve indicates the contrast at various focal positions, as determined from the traces. The row of photographs arranged across the top of the illustration show how the three-line, square-wave pattern appears to the eye as it is scanned by the slit. A comparison of the contrast curve with the photographs shows that the position of maximum contrast differs from the focal position of best definition for a line spacing of 11.1 lines/mm.

This discrepancy between position of highest contrast and best definition exists only in a spatial frequency area somewhat removed from the limit of resolution; near this limit, the two focal positions are coincident. Apparently, the reason for this phenomenon is: When the square wave pattern in the image is degraded because of a small departure from best focus, it approaches in wave form the fundamental term in the series representation, the third harmonic being unresolved. In Figure 6, the position of best definition is indicated by a wave form more nearly resembling a square-wave with flat tops although of lower amplitude.

REFERENCES

1. Rosberry, F. W., "Equipment and Method for Photoelectric Determination of Image Contrast Suitable for Using Square Wave Targets," *J. Research*, Vol. 64C, No. 1 (1960).
2. Washer, F. E. and F. W. Rosberry, "New Resolving Power Test Chart," *J. Opt. Soc. Am.*, 41, 597 (1951).
3. La Bauve, R. J. and R. A. Clark, "Potentialities, for Image Evaluations of Geometric Ray Trace Focal Plots, *J. Opt. Soc. Am.*, 46, 677 (1956).

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 65C, No. 3, July–September 1961

Location of the Plane of Best Average Definition With Low Contrast Resolution Patterns

Francis E. Washer and William P. Tayman

(April 28, 1961)

The plane of best average definition is located for each of several airplane-camera lenses using two types of low contrast test pattern and two emulsions. A low contrast pattern composed of dark lines on a light background and the reverse pattern consisting of light lines on a dark background are used. The results of measurement indicate that the position of the plane of best focus and the numerical magnitudes of the root mean product mean $\sqrt{\overline{R_\beta T_\beta}}$ value of the resolving power are not significantly affected by this reversal of contrast In addition, the results obtained using low contrast targets are compared with those obtained with high contrast targets. Hence the position of the selected focal plane remains invariant although the values of the measured resolving power are substantially higher for the high contrast targets.

1. Introduction

In specifications dealing with airplane mapping cameras, great emphasis is placed upon the photographic resolving power of the lenses used in the camera. The method of measurement to be used in evaluating the resolving power is usually described in careful terms, with the type of test chart, the characteristics of the registering emulsion, the physical conditions of test, and the manner of reporting results all given in detail. A 1950 military specification, MIL–STD–150 [1] [1] attempted to establish standard practices in the evaluation of lens performance. A recent revision of this comprehensive specification, MIL–STD–150A [2] was issued in 1959. At the Washington meeting of the International Society of Photogrammetry in 1952, a specification dealing with the calibration of photogrammetric cameras was adopted for trial and discussion [3].

At the time the International Specification for the calibration of photogrammetric cameras was drawn, it was not possible to standardize on a single type of resolving power chart. Consequently several were listed as suitable for use; these include the three-line chart used by the U.S. Air Force [2], the Cobb two-line chart used in Great Britain and the annulus chart used in Canada [4]. These charts were of low contrast except that the three-line chart of the Air Force was provided in both high and low contrast versions. The three-line chart of the National Bureau of Standards [5] was not included as it has been available for only a few months prior to the drawing of the specification and its properties were not widely known.

In the eight years that have passed since the adoption of this specification, agreement on a single test chart has not yet developed. A few publications have appeared [6, 7] that permit an estimate of the differences likely to be found in comparing various types of charts.

Because of the differences that exist among the various test charts used in various laboratories, a study was initiated at the National Bureau of Standards in which the values of resolving power for the same group of lenses were determined for a variety of test charts, target contrasts, and emulsions. The results of this study should be useful in estimating the probable values of resolving power that would be obtained for a lens with a given type of chart when results are available for the same lens with a different type of chart.

This is the concluding portion of a three part investigation. The results for the first part dealing with various indices and methods used in locating the plane of best average definition have been reported in Photogrammetric Engineering [8]; the results for the second part which comprises results for lone-line, short-line, and annular charts are reported in J. Research NBS [9]. The results of the third part which deals with variations arising from differences in target contrast for two emulsions are reported in the present paper.

Measurements of photographic resolving power at the National Bureau of Standards are customarily made using a long-line three line high contrast test chart having dark lines on a light background with the images registered on a high contrast emulsion which has a higher resolving power than the emulsion commonly used in aerial photography. In many other calibration laboratories, the tendency is to use test charts having light lines on a dark background

[1] Figures in brackets indicate the literature references at the end of this paper.

with the images registered on a medium contrast emulsion with resolving power comparable to that being currently used in aerial photography.

In the present study, values are obtained for two emulsions and three targets of differing contrast. Results are given for four lenses of types commonly used in photogrammetric cameras.

2. Method of Measurement

2.1. Test Camera

The National Bureau of Standards precision lens testing camera [10] was used in making the negatives from which the values of resolving power are determined. Its construction and mode of operation simplify the process of evaluating the resolving power of a lens in a series of focal planes spaced at definite intervals along the optical axis. The precision lens testing camera is equipped with 10 collimators spaced at 5° intervals and spanning the range from 0° to 45°. Resolving power test charts placed in the focal plane of the collimators and imaged by the lens under test permit the study of quality of imagery for a series of focal planes of the test lens. By using one type of chart in 5 of the 10 collimators (designated the odd-numbered collimators) spaced at 10° intervals and another type in the remaining five collimators (designated the even-numbered collimators) it is possible by proper manipulation of the various controls to record the imagery of a lens under test on a single photographic plate for each of two types of chart in the same series of focal planes for the same angular separations, β, from the axis of the lens under test.

2.2. Test Charts

Three types of test chart are used in the present investigation. All three are based upon the NBS resolution chart of 1952 but differ in contrast between the lines and spaces. Each chart has the same range of sizes of the test object with the size proceeding in a geometric progression with $\sqrt[4]{2}$ (or 1.1892) as the common ratio. The same size test chart is used in the various collimators so that it is necessary to apply the "cos" and "cos^2" corrections to images formed off-axis. The various types are described in the following section.

a. High Contrast Test Chart (HDL)

This is a composite chart based upon the NBS resolution test chart of 1952 which is described in detail in NBS Circular 533 [5]. It is a long-line three line chart with dark lines on a light background. The difference of log luminance between the dark lines and the light background exceeds 2.0.

b. Low Contrast Test Chart (LDL)

This chart is identical in all respects to that described in (a) except that the contrast is low between the dark lines and the light background. The difference of log luminance between the dark lines and the light background is 0.20 $\pm$0.02.

c. Low Contrast Test Chart (LLD)

This chart is identical to that described in (b) except that the contrast is reversed, that is, the difference in log luminance between the light lines and the dark background is 0.20 $\pm$0.02.

2.3. Test Negative

In making the test negative, the lens under test is initially so alined that its optical axis is parallel to and approximately coincident with the axis of the collimated beam emergent from the first collimator. The lens is adjusted along the bench to a location such that the collimated beam at 40° from the axis fills the front aperture of the lens under test as viewed through the lens at an inclination of 40° from the axis. The plate holder is adjusted to a position such that the front surface of the emulsion is in the plane of best visual axial focus for the central row of images to be registered on the plate. The plate holder is then moved to a position 1.05 mm nearer to the lens, where an exposure is made by illuminating the reticles in odd-numbered collimators, which thus records the imagery on the plate at 10° intervals for the range of angles from β=0° to β=40°. This process is repeated with the plate moved 0.15 mm farther from the lens until 15 exposures have been made with the last for the plane 1.05 mm farther from the lens than the position of best visual axial focus. Between each exposure, the plate holder is also moved downward by an amount sufficient to avoid superposition of successive rows of images. The foregoing operation registers the imagery for the long-line high contrast patterns on the plate. The plate holder is then returned to its initial position for which the emulsion surface is 1.05 mm nearer to the lens than the plane of best visual axial focus; the plate is displaced sidewise in its holder approximately 12 mm; and the entire bench on which the camera is mounted is rotated on its pivot by 5° so that the axis of the lens is now parallel to and nearly coincident with the axis of the collimated beam emergent from the second collimator. The foregoing procedure is then repeated with the exception that exposures are made by illuminating the reticles in the even-numbered collimators, which again registers imagery on the plate at 10° intervals for the range of angles from β=0° to β=40°. This process registers the imagery for the long-line low contrast patterns on the plate. The exposed plate is then processed to form the finished negative from which values of resolving power for tangential lines (T_β) and radial lines (R_β) are determined.

All exposures are made with the reticles illuminated by light from tungsten lamps with Wratten K–3 filters between the light source and test charts. Neutral filters are used to adjust the intensity of the light reaching the plate so that the final resulting optical density of each image on the negative is approximately the same for all values of β. The two types of photographic plates used in this work were Eastman Spectroscopic VF which has a fine grained panchromatic emulsion and Eastman Super Panchro Press, Type C. The plates were developed for 3 min. in D–19 developer at 68 °F.

2.4. Reading the Negative

The negative images were examined with a microscope using powers ranging from 30 to 50×. The criteria for considering a particular line pattern to be the finest resolved were that all coarser patterns were resolved and that the number of lines in a given pattern was the same as that of the corresponding pattern in the object.

3. Results of Measurements

The results of measurement on four wide-angle lenses are reported and analyzed in this study. Two lenses, designated Nos. 1 and 5, are essentially distortion-free; the other two, designated 7 and 8, have moderate amounts of distortion. Measurements of resolving power were made at 10° intervals from 0° to 40°.

Negatives were made with both VF and SP emulsions for each of three target contrasts for each of the four lenses. The measured equivalent focal length for the plane of best visual axial focus was the same for each emulsion and each type of target pattern. The maximum range of change in measured equivalent focal length did not exceed ±0.02 mm for a given type of lens.

The results of measurement are shown graphically in figures 1 and 2. The values of the geometric mean resolving power $\sqrt{R_\beta T_\beta}$ are shown plotted against the separation, Δf, from the plane of best visual axial focus. Positive values of Δf denote positions of the image plane farther from the lens than the plane of best visual axial focus. In figure 1, values of $\sqrt{R_\beta T_\beta}$ obtained with VF emulsion are shown for values of β ranging from 0° to 40°. The lowest frame in each column shows values of the root mean product mean [8] $\sqrt{\overline{R_\beta T_\beta}}$ versus Δf. The curves, numbered 1, show the results obtained with the high contrast target having dark lines on a light background. The curves, marked 2, show the results for the low contrast target having dark lines on a light background. The curves, marked 3, show results for the low contrast targets having light lines on a dark background. In figure 2, comparable results are shown for the same four lenses using SP emulsion.

Values of $\sqrt{R_\beta T_\beta}$ are shown in these graphs as this quantity is of primary interest in evaluating the image forming qualities of a lens intended for use in an airplane mapping camera. In two earlier papers [8, 9], values of R_β and T_β are shown for lens No. 1.

4. Location of the Plane of Best Average Definition

The plane of best average definition was located for each set of conditions by the maximum value of the index $\sqrt{\overline{R_\beta T_\beta}}$. The plane so located is likely to be slightly farther from the lens than that located by graphical analysis [8]. For the purposes of the present paper, the use of the maximum of $\sqrt{\overline{R_\beta T_\beta}}$ is satisfactory. The values of the index $\sqrt{\overline{R_\beta T_\beta}}$ for the range of Δf extending from 0.45 mm nearer to the lens to 0.30 mm farther from the lens than the plane of best visual axial definition is shown in table 1 for the four lenses for all of the various conditions. The values listed are the averages for several runs ranging from 1 to 6. The total number of runs for the high contrast target is equal to the sum of the runs for the two low contrast targets. This occurs because a single negative always contains a record of the imagery for the high contrast target and for one of the two low contrast targets. In table 1 the underlined value of the index in each column is the maximum for the indicated set of conditions.

5. Comparison of Results Obtained With the Three Charts

The principal points to be considered in the analysis of the results of measurement are possible differences in the location of the focal plane for the various conditions, the reduction in average resolving power that may result from such differences, relative magnitudes of average resolving power for the two emulsions and effects of contrast. These are discussed in the present section.

5.1. Relative Positions of Focal Plane

The relative positions of the focal plane for the various conditions are indicated for the four lenses in table 1. The displacements for each of the contrasts and emulsions from the focal plane of best average definition for the high contrast target using VF emulsion of the corresponding planes are shown in table 2. It is clear from this table, that differences in contrast of target have little or no effect upon the location of the focal plane of best average definition for either emulsion. Only in the case of the low contrast target having light lines on a dark background (LLD) used with SP emulsion is there any indication of a possible effect on the location of the focal plane and even in this instance the apparent shift may be a result of random error.

On the average, the plane of best average definition for the various contrasts appear to be 0.15 mm nearer to the lens for the SP emulsion than is the corresponding plane for VF emulsion. This difference is small and may be real. However, it would be unlikely to present serious cause for concern in the locating of the plane of best average definition. It is likely that the average plane selected using VF emulsion will be satisfactory if this plane be used for photography with SP emulsion. Likewise the plane selected using SP emulsion would be reasonably satisfactory for photography with VF emulsion.

5.2. Reduction in Resolving Power From the Maximum

If the differences in the relative location of the focal plane are regarded as genuine, it is instructive to determine the magnitude of the effect of these differences on resolving power under the various

TABLE 1. *Variation of $\sqrt{R_\beta T_\beta}$ with separation, Δf, from the plane of best visual axial focus for four lenses*

Values of $\sqrt{R_\beta T_\beta}$ in lines per millimeter are given for two emulsions and three target contrasts. The columns, designated HDL, show values obtained with a high contrast target having dark lines on a light background. The columns, designated LDL, show values for a low contrast target having dark lines on a light background. The columns, designated LLD, show values obtained with a low contrast target having light lines on a dark background. The maximum values of $\sqrt{R_\beta T_\beta}$ are underlined. The number of test negatives used in each set of determinations is shown in the row marked *n*.

	Values of $\sqrt{R_\beta T_\beta}$ for lens No. 1 obtained with emulsion					
	VF			SP		
n	2	1	1	3	1	2
Δf	HDL	LDL	LLD	HDL	LDL	LLD
mm						
−0.45	39.3	27.2	25.2	20.0	14.6	13.5
−.30	42.0	28.9	27.8	22.2	15.3	13.8
−.15	46.4	31.3	28.3	22.9	15.4	13.7
.00	48.6	32.4	28.7	21.8	14.9	13.5
.15	48.4	29.3	26.6	19.4	12.7	11.8
.30	38.6	23.2	23.2	15.6	11.1	10.4

	Values of $\sqrt{R_\beta T_\beta}$ for lens No. 5 obtained with emulsion					
	VF			SP		
n	3	2	1	6	3	3
Δf	HDL	LDL	LLD	HDL	LDL	LLD
mm						
−0.45	30.1	22.2	22.3	19.8	14.4	13.1
−.30	34.6	26.0	24.8	21.8	15.1	14.5
.15	39.8	27.6	27.2	23.4	15.6	14.1
.00	42.2	28.8	27.2	22.8	15.2	13.8
.15	42.1	26.8	28.3	21.4	13.5	12.6
.30	36.6	23.0	23.4	17.6	11.7	10.8

	Values of $\sqrt{R_\beta T_\beta}$ for lens No. 7 obtained with emulsion					
	VF			SP		
n	4	2	2	2	1	1
Δf	HDL	LDL	LLD	HDL	LDL	LLD
mm						
−0.45	42.6	31.6	29.6	24.6	16.1	14.3
−.30	46.5	33.0	29.8	24.8	16.4	14.3
−.15	48.1	33.0	30.0	24.5	16.2	14.0
.00	46.8	32.3	29.2	23.6	16.2	13.9
.15	43.4	27.2	27.4	21.5	13.9	12.4
.30	36.5	24.2	23.4	17.9	12.8	11.1

	Values of $\sqrt{R_\beta T_\beta}$ for lens No. 8 obtained with emulsion					
	VF			SP		
n	3	1	2	2	1	1
Δf	HDL	LDL	LLD	HDL	LDL	LLD
mm						
−0.45	34.5	28.0	24.2	21.4	13.3	13.2
−.30	36.7	28.3	26.6	21.5	13.3	13.2
−.15	37.4	28.8	26.4	21.7	13.7	12.9
.00	38.9	27.8	25.9	21.0	13.4	12.5
.15	37.4	25.5	24.4	18.8	11.7	11.6
.30	33.2	23.2	21.6	17.2	10.8	10.9

TABLE 2. *Location of the focal plane of best average definition with respect to that obtained for a high contrast target using VF emulsion for four lenses*

Results are given for VF and SP emulsion and for two types of low contrast target. The symbols in the colmn heading are defined in table 1. A negative value of Δf indicates a position of a selected focal plane nearer to the lens than that selected for the high contrast target using VF emulsion. The focal plane of best average definition is determined by the maximum value of $\sqrt{R_\beta T_\beta}$.

Lens	Separations, Δf, in millimeters from plane of best average definition obtained with high contrast target for emulsion					
	VF			SP		
	HDL	LDL	LLD	HDL	LDL	LLD
1	0.00	0.00	0.00	−0.15	−0.15	−0.30
5	.00	.00	.15	−.15	−.15	−.30
7	.00	.00	.00	−.15	−.15	−.15
8	.00	−.15	−.30	−.15	−.15	−.30
Average	0.00	−0.04	−0.04	−0.15	−0.15	−0.26

TABLE 3. *Effect on average resolving power, $\sqrt{R_\beta T_\beta}$ for various contrast targets produced by using focal plane of best average definition obtained with VF emulsion for the high contrast target*

Values of the reduction from the maximum are given in percent for four lenses using VF and SP emulsions for one high and two low contrast targets. The meanings of the symbols heading the columns are as given in the caption of table 1.

Lens	Reduction in % of $\sqrt{R_\beta T_\beta}$ from its maximum for					
	VF emulsion			SP emulsion		
	HDL	LDL	LLD	HDL	LDL	LLD
1	0.0	0.0	0.0	4.8	3.2	2.2
5	.0	.0	3.9	2.6	2.6	4.8
7	.0	.0	0.0	1.2	1.2	2.1
8	.0	3.5	2.6	3.2	2.2	5.3

conditions. For example, one may select the focal plane of best average definition for high contrast targets with VF emulsion and then use this focal plane for photography under all conditions of contrast for both emulsions. This has been done and the effects on resolving power for the other conditions are shown in table 3. From this table it is clear that for 3 of the 4 lenses, there is no reduction in resolving power when using VF emulsion for the low contrast target having dark lines on a light background (LDL) while there is a 3.5 percent loss in the case of lens No. 4. For the low contrast target having light lines on a dark background, the change is zero for two lenses and reductions of 3.9 and 2.6 percent are shown for the other two. When SP emulsion is used, reductions in resolving power ranging from 1.2 to 5.3 percent occur, with the reduction under 3 percent in 7 of 12 instances. One can expect a 3 percent variation in the magnitude of $\sqrt{R_\beta T_\beta}$ on the basis of observational error, hence in only five of the twelve cases is the reduction in resolving power in excess of the probable observational error.

If the plane of best average definition for the high contrast target using SP emulsion is taken as a reference plane, values of the reductions in resolving power can be determined for the other condition.

The values of these reductions are shown in table 4. For this condition, the reduction is zero for the low contrast target (LDL) with SP emulsion. For the VF emulsion, the reduction is under 3 percent in 5 of 12 cases and does not exceed 5.6 percent for any of the remainder.

5.3. Relative Magnitude of the Resolving Power

From figures 1 and 2, it is apparent that the resolving power for VF emulsion is higher than that for SP emulsion for all values of β ranging from 0° to 40°. It is moreover clear that this is true for all three conditions of contrast. It is not possible to determine a numerical value of the ratio of the resolving powers that would be invariant for all values of β and all values of Δf. Accordingly, average values of the ratios based upon the values of $\sqrt{\overline{R_\beta T_\beta}}$ shown in table 1 have been determined for each lens and are shown in table 5. In table 5, the ratios of average resolving power for the two emulsions and three contrasts with respect to the average resolving power for the high contrast target obtained with SP emulsions for four lenses. From the average values of the ratios given in this table, an estimate can be made of the probable value of other ratios that are of interest. For example, values of resolving power obtained with the low contrast target are approximately two-thirds of that obtained with the high contrast target for either of the two emulsions.

TABLE 4. *Effect on average resolving power, $\sqrt{\overline{R_\beta T_\beta}}$ for various contrast targets produced by using focal plane of best average definition obtained with SP emulsion for the high contrast target*

Values of the reduction from the maximum are given in percent for four lenses using VF and SP emulsions for one high and two low contrast targets. The meanings of the symbols heading the columns are as given in the caption of table 1

Lens	Reduction in % of $\sqrt{\overline{R_\beta T_\beta}}$ from its maximum for					
	VF emulsion			SP emulsion		
	HDL	LDL	LLD	HDL	LDL	LLD
1	4.5	3.4	1.4	0.0	0.0	0.7
5	5.6	4.2	3.9	.0	.0	2.8
7	3.3	0.0	0.7	.0	.0	0.0
8	3.8	.0	.8	.0	.0	2.3

TABLE 5. *Ratios of average resolving powers for two emulsions and three targets of various contrasts*

Ratios are given showing the relative magnitude of the average resolving power obtained under specified conditions to that obtained with a high contrast target on SP emulsion. Values are given for four lenses, using two emulsions and three contrasts. Values of the averages are given together with the average probable error $\overline{PE}_a$ of a single determination. The meanings of the symbols heading the columns are as given in the caption of table 1.

Lens	Ratios of average resolving powers for					
	VF emulsion			SP emulsion		
	HDL	LDL	LLD	HDL	LDL	LLD
1	2.17	1.42	1.32	1.00	0.69	0.63
5	1.77	1.22	1.21	1.00	.67	.62
7	1.93	1.32	1.24	1.00	.67	.59
8	1.80	1.33	1.23	1.00	.63	.63
Average	1.92	1.32	1.25	1.00	0.66	0.62
$\overline{PE}_a$ (%)	±6.9	±4.0	±2.8	----------	±2.6	±2.1

5.4. Effect of Reversed Contrast

In many specifications that require measurements of resolving power of lenses with low contrast targets, it is stipulated that the test chart shall consist of light lines on a dark background. This is so specified because of the belief that the use of a target having dark lines on a light background will result in lower measured values of resolving power than would be found with a target of opposite contrast because of the effect of veiling glare [3]. In figures 1 and 2, the curves marked 2 show values of $\sqrt{R_\beta T_\beta}$ as a function of Δf for the low contrast target having dark lines on a light background; the curves marked 3 show values of $\sqrt{R_\beta T_\beta}$ as a fraction of Δf for a low contrast target having light lines on a dark background. It is noteworthy that in most instances, the curve marked 2 lies slightly above the curve marked 3. Even if one ascribes the difference to experimental error, it is improbable that such error could reduce all of the observed values of curve 2 below those for curve 3. From these results, it is clear that the values of resolving power obtained using targets having dark lines on a light background are not lower than those obtained using targets having light lines on a dark background when the difference in log luminance between line and background is 0.2.

It seems probable, therefore, that when resolving power tests are made using transparent charts located in the focal plane of collimators, the use of light lines on a dark background will not, in general, yield higher values of the resolving power than will the use of dark lines on a light background. In this case, the values of resolving power are not noticeably reduced by the effect of veiling glare. Although veiling glare may affect the values of resolving power on the axis which may be the case for lens No. 5 in figure 2, it is not likely to do so for extra-axial imagery. The effect of veiling glare possibly would be more pronounced in using resolution targets on a range rather than in collimators which may account for the prevalence of the specification of targets having light lines on a dark background.

6. Conclusion

In this study, the effect of contrast upon choice of a focal plane has been investigated using four lenses, two target contrasts, and two emulsions. Analysis of the results of measurement leads to the following conclusions.

(1) For either of the two emulsions used, the choice of focal plane of best average definition is not affected by differences in contrast between lines and background in the target.

(2) When a lens images targets where the difference in log luminance between lines and background equals 0.20, the values of measured resolving power are not significantly affected by reversal of target contrast. In other words, low contrast targets with either dark lines on a light background or light lines on a dark background will yield substantially the same measured values of average resolving power for a given lens.

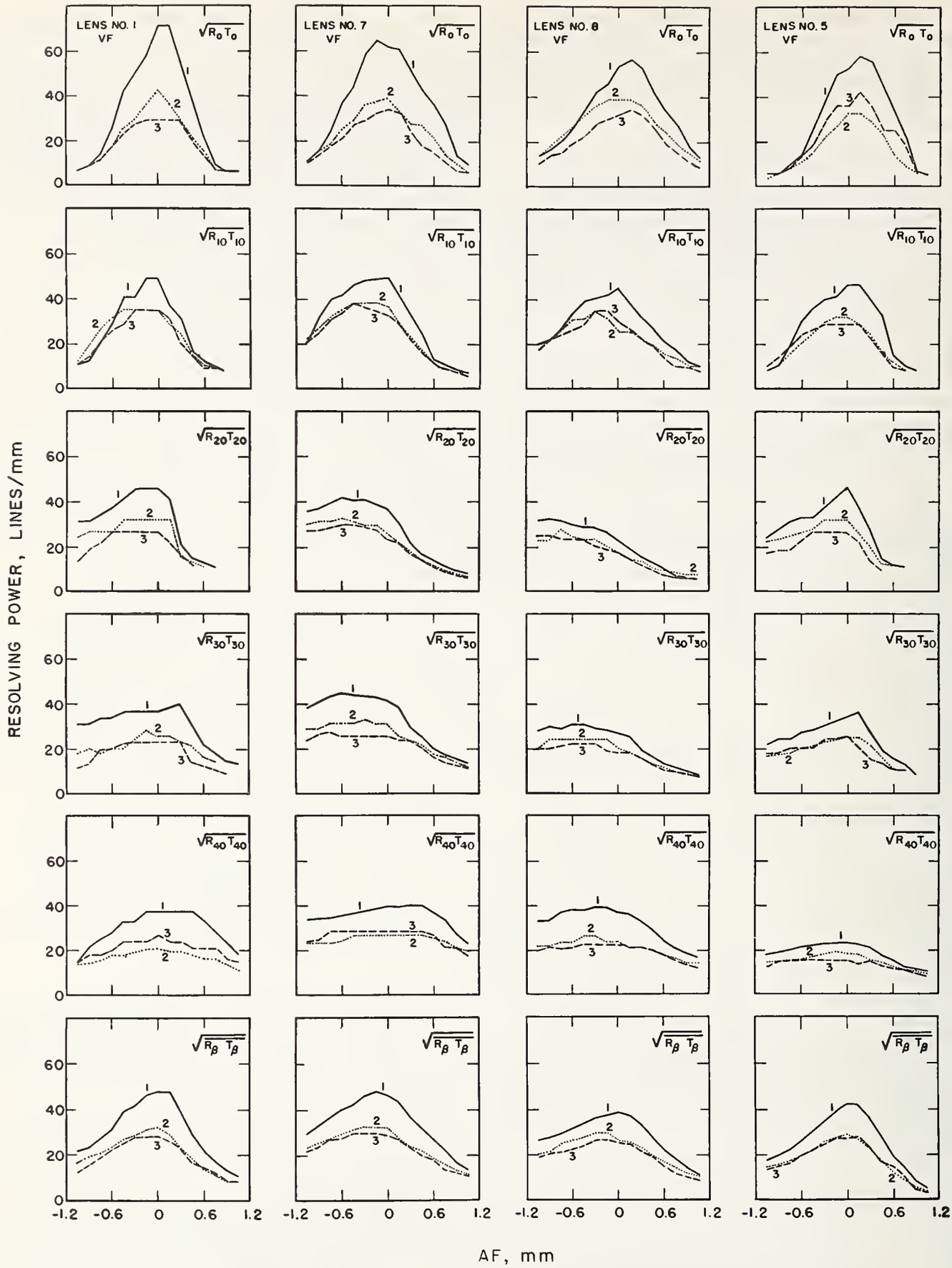


FIGURE 1. *Average resolving powers versus position of the image plane for four lenses with VF emulsion.*

Values of the average resolving power $\sqrt{R_\beta T_\beta}$ are shown at 10° intervals from 0° to 40°. The lowest box in each column shows values of $\sqrt{\overline{R_\beta T_\beta}}$. Curve 1 shows the result obtained with a high contrast target having dark lines on a light background with contrast between lines and background greater than 2.0 on a density scale. Curve 2 shows the results for a low contrast target (0.2 difference of log luminance between lines and background), having dark lines on a light background. Curve 3 shows results obtained for a low contrast target (0.2 difference of log luminance between lines and background) but having light lines on a dark background. The zero of abscissas marks the position of best visual axial focus, and positive values of Δf indicate positions farther from the lens.

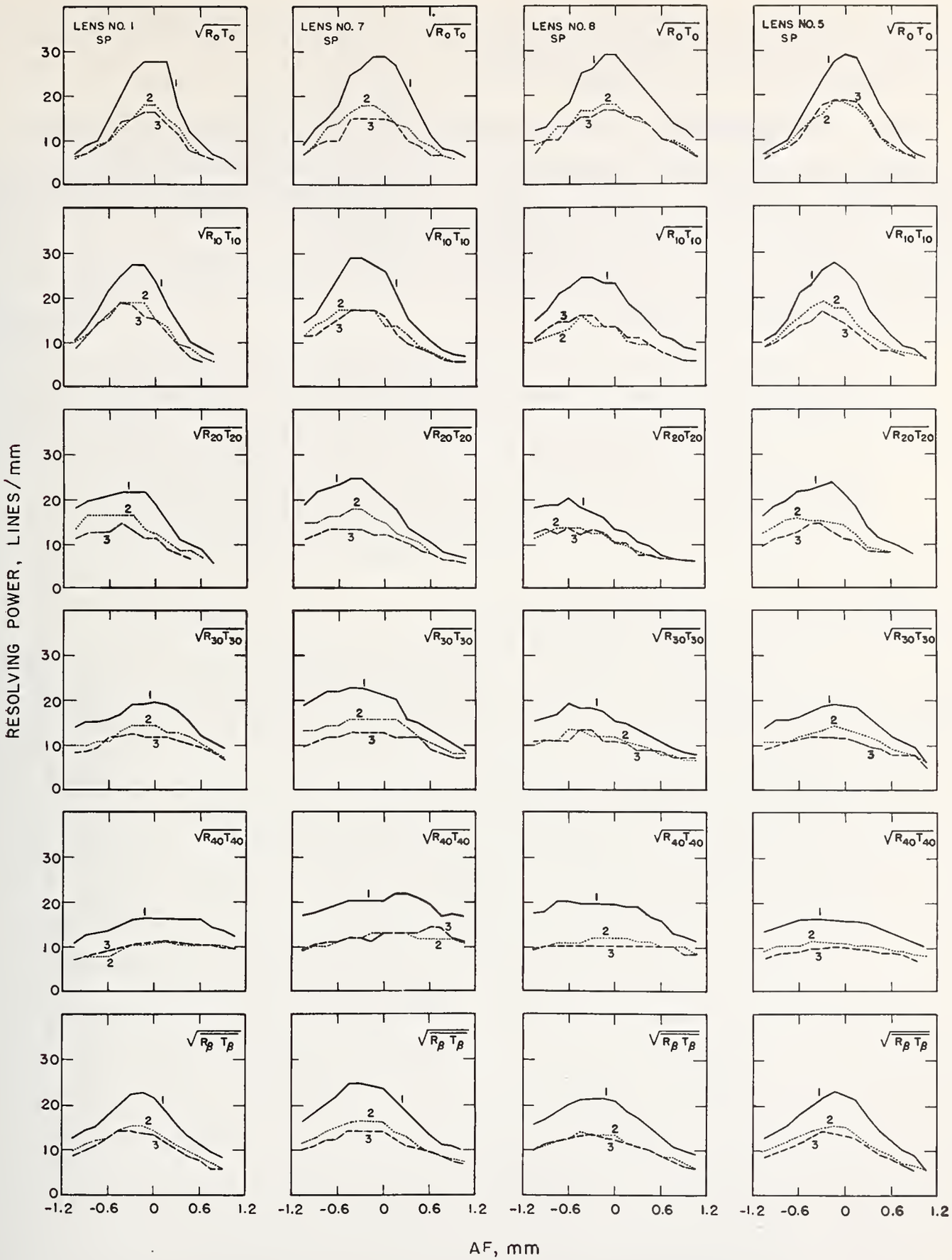


FIGURE 2. *Average resolving power versus position of the image plane for four lenses with SP emulsion.*

Values of the average resolving power $\sqrt{R_\beta T_\beta}$ are shown at 10° intervals from 0° to 40°. The lowest box in each column shows values of $\sqrt{R_\beta T_\beta}$. Curve 1 shows the results obtained with a high contrast target having dark lines on a light background with contrast between lines and background greater than 2.0 on a density scale. Curve 2 shows the results for a low contrast target (0.2 difference of log luminance between lines and background), having dark lines on a light background. Curve 3 shows results obtained for a low contrast target (0.2 difference of log luminance between lines and background) but having light lines on a dark background. The zero of abscissas marks the position of best visual axial focus, and positive values of Δf indicate positions farther from the lens.

The authors express their appreciation to other members of the staff of the National Bureau of Standards for assistance during this work and in particular to E. C. Watts who prepared the illustrations.

7. References

[1] Military Standard MIL–STD–150, Photographic Lenses, 23 Oct. 1950.
[2] Military Standard MIL–STD–150A, Photographic Lenses, 12 May 1959.
[3] Report of Commission I, International Society of Photogrammetry Specification of methods of calibrating photogrammetric cameras, Photogrammetria X, 85 (1953–1954).
[4] L. E. Howlett, Photographic resolving power, Canadian Journal of Research **24A**, 15–40 (1946).
[5] F. E. Washer and I. C. Gardner, Method for determining the resolving power of photographic lenses, NBS Circ. 533 (1953).
[6] F. H. Perrin and J. H. Altman, Studies in the resolving power of photographic emulsions, J. Opt. Soc. Am. **43,** 780 (1953).
[7] P. Hariharan, Resolving power of photographic emulsions, J. Opt. Soc. Am. **46,** 315 (1956).
[8] F. E. Washer and W. P. Tayman, Location of the plane of best average definition for airplane camera lenses, Photogrammetric Engineering XXVI, 475 (1960).
[9] F. E. Washer and W. P. Tayman, Variation of resolving power and type of test pattern, J. Research NBS **64C,** 209 (1960).
[10] I. C. Gardner and F. A. Case, Precision camera for testing lenses, J. Research NBS **18,** 449 (1937).

(Paper 65C3-71)

JOURNAL OF RESEARCH of the National Bureau of Standards—A. Physics and Chemistry
Vol. 65A, No. 6, November–December 1961

Comparison of Lens Response for Sinusoidal and Square-Wave Targets at Several Focal Positions

Sayeda H. Emara*

(June 20, 1961)

A study has been made of the sine-wave and square-wave responses of a lens at two apertures and several focal positions, both on- and off-axis. Two focal positions, one of which gives the best definition and the other the highest contrast for coarse patterns, were located precisely. At these focal positions and at several other arbitrarily chosen positions the sine-wave and square-wave responses were measured. Because of the scattering character of the photographically-made target objects, a special technique has been employed for calibrating the targets to obtain their contrast as seen by the lens under test.

The results show that for large apertures of the lens there is some frequency at which the response (either sine or square) is the same for the two focal positions (curves cross). This phenomenon has been further studied by computing the square-wave response of the lens from its sine-wave response; and it was found that there is close agreement between the computed and experimentally determined responses.

1. Introduction

During the past few years the National Bureau of Standards and a number of other laboratories have been engaged in a research program with the aim of developing a more objective method of assessing the image qualities of a lens than those methods currently in use.

Among numerous procedures, the one based upon Fourier analysis of the object and the image into sinusoidal components is of great theoretical interest. The striking feature of the Fourier approach is the simplicity with which sinusoidally varying luminance patterns are imaged. Diffraction and aberrations have no effect whatsoever on the sinusoidal shape of the image of such a wave, nor on its orientation. In other words, the image of a sinusoidal pattern is also sinusoidal; only the amplitude (or modulation factor) is changed by the imaging process. Thus the performance of an image-forming system may be described by a graph or table that gives the degradation of contrast as a function of spatial frequency for sinusoidal objects. Two such graphs or tables for sinusoidal objects at right angles to each other completely describe the image-forming characteristics at one point in the field.

The easiest method for determining these characteristics makes use of the multiple line object and has been used by several workers [1, 2, 3]. For this method the image of the lines is scanned by a narrow slit to measure the distribution of flux in that image. Recently an improved technique was developed at the National Bureau of Standards which scanned the aerial image directly with a slit and photo-multiplier tube [4]. The resultant variations in illuminance in the image were recorded on the moving chart recorder.

One might suppose that the proper focal plane at which the measurements should be performed would be easily determined as the one at which the response (or transfer function) was maximum. Unfortunately patterns of different frequency give different focuses for their respective maximum responses [5]. Consequently, the selection of the proper focus must be judged by additional criteria. Many investigators have reported the existence of different best focuses, depending upon the criteria. There is a focus that gives the best resolution of fine detail and, presumably, as a consequence, the sharpest reproduction of edges. There is another focus that gives best contrast of coarse detail. If only coarse detail is important, as in 530-line television pictures, this focus may be more nearly optimum than the other. In a previous work a comparison was made between the focal position for highest contrast and the position for best definition, using a coarse parellel line target [6]. Examination of photographs of the pattern indicated a maximum sharpness of edges at one focal position while the maximum amplitude of the traces gave a different position of highest contrast, with a separation of several tenths of a millimeter between the two positions for the particular lens tested. Although the results showed this discrepancy between the focal positions of highest contrast and best definition in a spatial frequency area somewhat removed from the limit of resolution, near this limit the two criteria were satisfied at the same position.

The conclusion arrived at for the location of the position of best definition were based entirely upon visual judgment. Also, all observations were carried

*Guest worker, at the National Bureau of Standards, from Ein-Shams University, Cairo, Egypt.

out using only one coarse pattern and one position, off-axis. Therefore, it was felt that further study of the problem, using more precise measurements, should be made. Thus it is the aim of this paper to present accurate procedures to locate the focuses of best definition and highest contrast and then to measure the sine-wave response and the square-wave response over a wide range of frequencies and apertures for a given lens in the hope of finding a satisfactory explanation of this phenomenon.

Although the work was done with only one lens, a Zeiss Biotar, $f/2$, with a focal length of 58 mm, the results are representative of the performance of practically all high grade photographic lenses.

2. Arrangement of Apparatus

The apparatus was as shown in figure 1. The source was a ribbon filament tungsten lamp focused by a condensing lens on the target located at the focus of the collimator. Thus illumination in this case fulfills the conditions for noncoherence [7]. The lens under test was located on a nodal slide in the collimated beam. The nodal slide and pickup unit were mounted on separate saddles riding on a length of ways alined with the optical axis established by the collimator.

Immediately behind the test lens was the pickup unit consisting of a microscope, slit, and photo tube. In operation the object was moved in the focal plane of the collimator so that the image moved across the slit.

3. Description of Targets

The two targets used in this investigation were transparencies on glass plates, with a transmittance variation forming a square wave for one and a sine wave for the second. There was no variation in transmittance in the orthogonal direction. The square-wave target consisted of 25 patterns, each of which had three dark lines, with lines and spaces of equal width. The spatial frequencies in the image of the patterns ranged from the coarsest, of 6.12 lines per millimeter, to the finest, of 97.6 lines per millimeter. The sine-wave target had 23 patterns producing spatial frequencies in the image of 3 lines per millimeter for the coarsest pattern and approximately 500 lines per millimeter for the finest pattern.

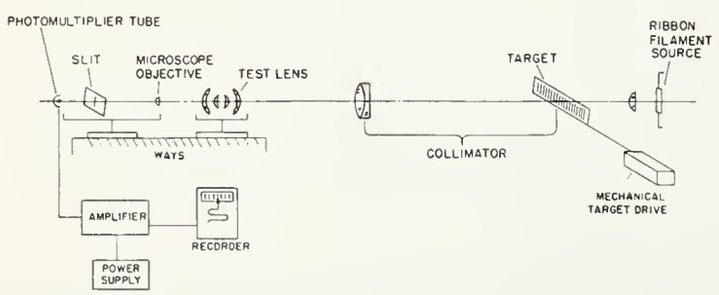


FIGURE 1. *Schematic diagram of the apparatus used.*

The target under investigation was mounted on a carriage and moved through the light beam in a plane parallel to the image plane along a line at right angles to the lines on the target. The scanning speed was constant for a given pattern but for coarse patterns it was faster than for fine patterns. An effort was made to maintain a constant scanning rate in terms of lines per minute for the entire target by changing target speed with gears.

4. Calibration of Targets

Both targets were calibrated using a Leeds and Northrup microdensitometer. The numerical apertures of the microscope and condenser lenses of the microdensitometer were approximately 0.35, whereas those of the collimator and lens being tested were very small, always less than 0.008. This difference of geometry between the system in which the targets were calibrated and that in which they were used resulted in an error. This error was discovered when the measured sine wave responses of the lens were apparently greater than one for some of the coarse patterns. This cannot, of course, be correct. The discrepancy was attributed to the fact that the contrast of photographic materials appears greater with small apertures than with large because of their scattering nature.

It is not possible to eliminate this error by performing the calibration with the aperture of the microdensitometer reduced so as to match the conditions of use. Such a procedure would so reduce the resolving power of the microdensitometer as to render the calibration useless at the higher frequencies. Consequently an indirect calibration process was adopted.

For this calibration a photographic plate was exposed to give a series of fairly broad strips, about 1 cm in width, varying in transmittance by small increments over the entire density scale of the plate. The type of plate was chosen in order to have similar scattering characteristics to that on which the target objects were made. The transmittances of the strips on this plate were measured on the microdensitometer; they were also measured on the test instrument at a numerical aperture of approximately 0.008. The widths of the strips were sufficient that the accuracy of these measurements was not reduced by the low resolving power of the system. The values obtained are exactly the same as if the lens system were free from aberrations and diffraction. A calibration curve was made by plotting transmittance values by test instrument as a function of transmittance by microdensitometer using the data obtained with the step density plate. This is shown in figure 2.

Obviously the end points, zero and unity, are the same on both instruments, but in between there is considerable difference. Correction of target modulations by means of this calibration eliminated responses greater than one and led to satisfactory agreement between measured and computed square-wave responses.

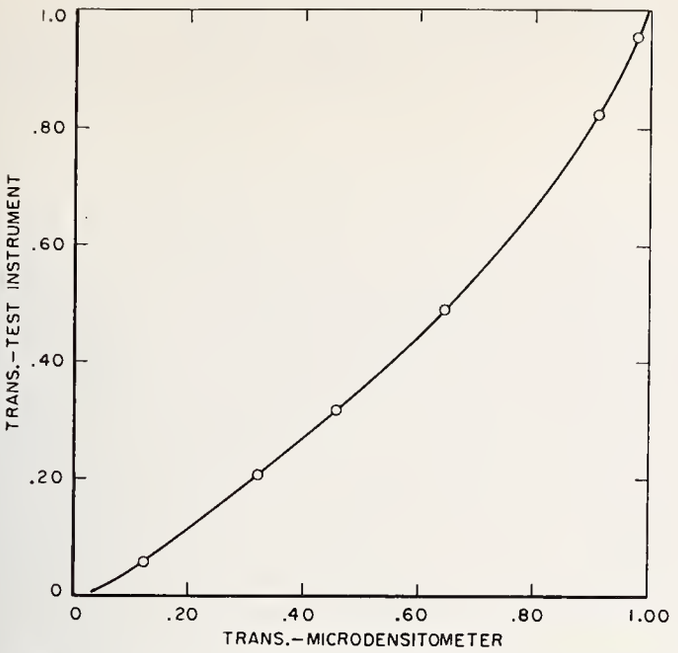


FIGURE 2. *Calibration curve for transmittance measured by the test instrument.*

5. Procedure

The initial requirement of each run was to determine the focal positions of best definition and highest contrast for the selected f-stops. Runs were made at lens stop openings of $f/2$ and $f/8$. The position of best definition was determined by using a pattern near the limit of resolution but still giving enough modulation, compared to the noise level, to be measured with moderate accuracy. Several scans were made of the same target pattern with small change in focal position between scans to determine the position yielding the largest signal or modulation. Modulation of the image was calculated using the formula

$$M_i = \frac{T_{\max} - T_{\min}}{T_{\max} + T_{\min}},$$

where M_i is the image modulation produced by the lens under test, $T_{\max}$ the illuminance of a space, and $T_{\min}$ the illuminance of the adjoining line. A plot of image modulation against focal position was drawn. The maximum modulation of this curve determined the position for best definition. The procedure was repeated using a coarse pattern on the square-wave target and a similar curve was drawn to indicate the position of highest contrast. There was little difficulty in finding these focal positions of the lens on axis but at off-axis positions the increased depth of focus of the lens gave flat-top curves, especially for the highest contrast positions, as can be seen in figure 3.

Several runs were made at other apertures, $f/2.8$, $f/4$ and $f/5$–6, where it was found that there was practically no difference between the positions of best definition and best contrast. This is an interesting phenomenon itself, which is worthy of further investigation; but, as we were interested in cases where these two focuses were distinct, no further results are reported for these apertures.

After the correct focal positions had been obtained, the response was determined for several patterns on each target and a plot was made of response as a function of frequency for each focal position. This was accomplished by moving the target through the light beam parallel to the image plane and tracing the resultant image illuminances on the recorder chart. A pattern consisted of a set of lines and spaces of constant spatial frequency. Thus, the target pattern produced a spatial variation in luminance consisting of a sinusoidal component superimposed on a constant component. The ratio of the amplitude of the sinusoidal component to the magnitude of the constant component is the target modulation. An imperfect image of this pattern was formed by the lens under test. This image was further projected onto the plane of the receiver slit with a good low-power microscope objective. Thus the distribution of illuminance in the image was measured.

When the target is imaged by the lens under test, the modulation is reduced, the degree of reduction depending on the quality of the lens. The ratio of the modulation in the image to that in the object is called the response. This factor generally diminishes as the spatial frequency of the pattern increases, the lens acting as a low-pass filter on spatial frequencies. Thus if M_t denotes the modulation of the target and M_i that of the image produced by the lens under test, then the ratio of the two is called response,

$$\Phi = \frac{M_i}{M_t}.$$

The use of the modulation automatically compensates for any factor which changes the signal output of the testing device proportionally such as the transmission factor of the lens, change in brightness of the source, change in gain of the detector amplifier, etc. It must be remembered that the complete response function is actually complex, involving both an amplitude and a phase factor, of which Φ is the amplitude factor. However, for measurements made on axis there is no phase shift involved. In most cases, for evaluation purposes, the phase shift factors can be neglected.

The square-wave response of the lens was computed from its sine-wave response, employing Fourier analysis. A general object was considered to be the summation of a set of sinusoidal waves spatially distributed in the object plane, these component waves differing from each other in amplitude, frequency, phase, and direction. The image consisted of the summation of the images of these component waves.

The square-wave test object can be analyzed into its component harmonics, each of which is attenuated by the Φ corresponding to its frequency, and the square-wave response is obtained by adding these attenuated harmonics [8]. For the response

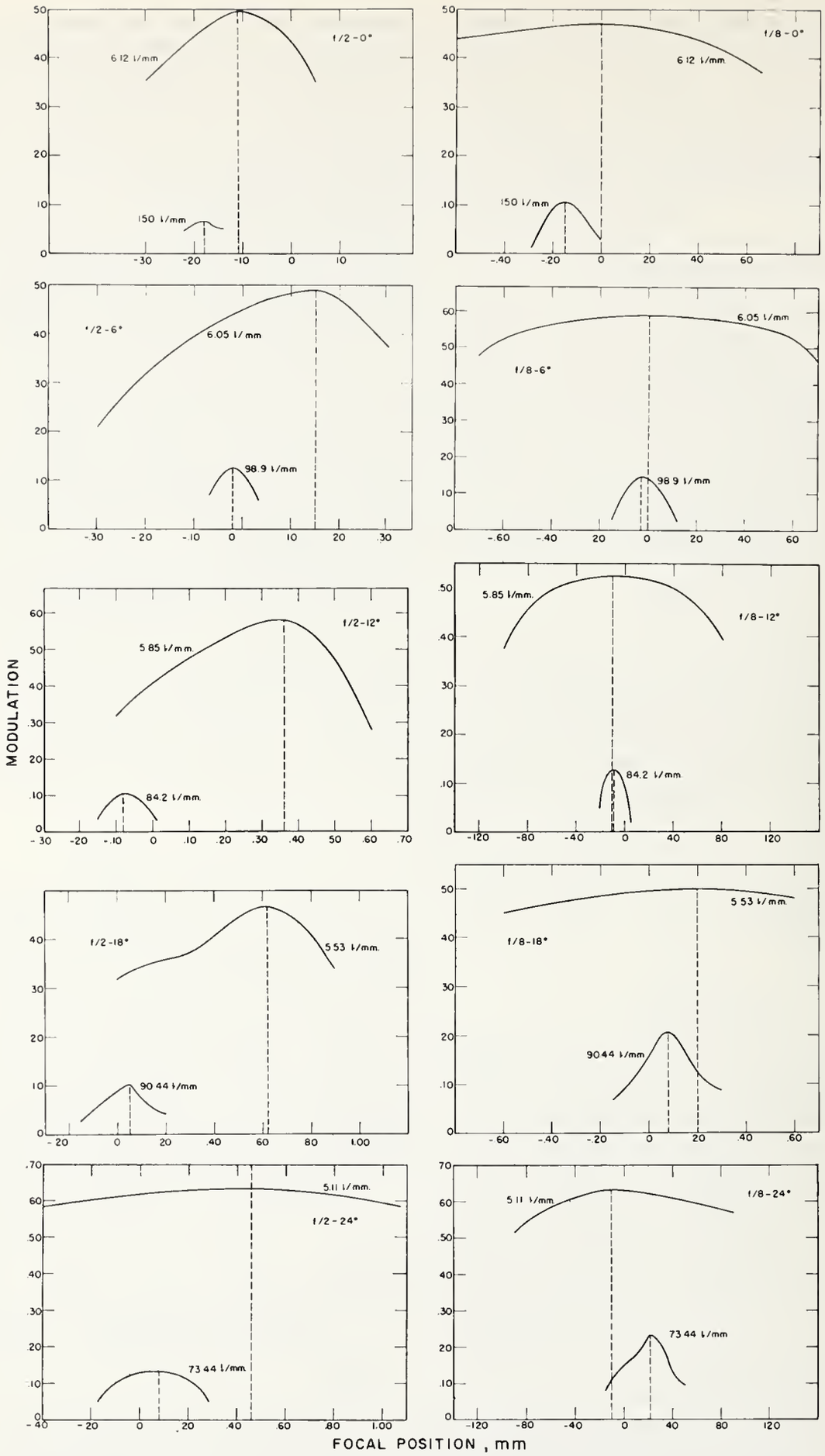


FIGURE 3. *Curves used for obtaining the focal positions for best definition and highest contrast.*

curve we are only interested in the peak-to-peak values which are obtained from the values at the centers of the lines and spaces. The square-wave response curve on axis is then given by:

$$\Psi(\omega)=\frac{4}{\pi}\left[\Phi(\omega)-\frac{1}{3}\Phi(3\omega)+\frac{1}{5}\Phi(5\omega)-\ldots\right]. \quad (1)$$

It should be noticed that there will be only a finite number of terms in the sum because there is a limiting value of ω beyond which Φ is zero for all larger values of ω. This limit exists because of the finite dimensions of the aperture of the system the larger the aperture, the higher the limiting frequency.

Figure 4 shows square-wave target curves or axis using a lens stop of $f/2$. The two solid curves are the experimental curves plotted from data taken in the focal positions of best definition and highest contrast, while the two dotted curves are obtained from values calculated from eq (1). The agreement between the experimental and the calculated curves is very good.

It is to be emphasized that comparison of measured square-wave responses with computed values was entirely unsatisfactory until averages of several runs were used for both the square-wave and sine-wave responses used in the computation. Accordingly, care has been taken to have an average of several runs for each curve shown in this investigation. To convey an idea of the magnitudes of the uncertainties of these results, tables 1 and 2 show the average responses and associated probable errors for two different focuses, best definition and highest contrast, respectively, for the square wave target. These tables represent the results of five runs each.

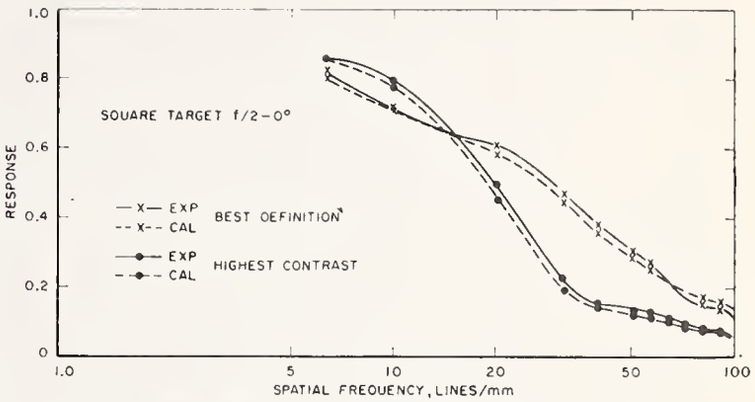


FIGURE 4. *Best definition and highest contrast curves for the square-wave target at f/2 on axis. The two solid curves are those obtained by experiment while the two dotted curves are for values calculated from the sine-wave response using Fourier analysis.*

TABLE 1. *Responses to square waves at the focus of best definition with associated probable errors; measured on axis at f/2.*

Lines/mm	6.12	9.7	19.4	30.7	38.7	49.2	55.2	62.4	69.8	78.5	88.4	97.5
Average response	0.820	0.720	0.626	0.492	0.384	0.310	0.272	0.220	0.177	0.151	0.142	0.111
Probably error	±0.025	±0.027	±0.021	±0.018	±0.018	±0.012	±0.012	±0.008	±0.008	±0.007	±0.008	±0.006

TABLE 2. *Responses to square waves at the focus of highest contrast with associated probable errors; measured on axis at f/2.*

Lines/mm	6.12	9.7	19.4	30.7	38.7	49.2	55.2	62.4	69.8	78.5	88.4	97.5
Average response	0.860	0.800	0.514	0.228	0.158	0.138	0.124	0.114	0.097	0.084	0.079	0.060
Probable error	±0.027	±0.026	±0.035	±0.015	±0.008	±0.014	±0.007	±0.006	±0.007	±0.007	±0.005	±0.002

6. Results

Figure 5 illustrates the square- and sine-wave response curves at an $f/2$ stop opening on axis and at off-axis positions of 6°, 12°, 18°, and 24°. The curves for the square-wave show an interesting phenomenon. On axis the best-definition and highest-contrast curves intersect at a spatial frequency of 14.5 lines per millimeter with a response of 0.66. At 6° off-axis best-definition and highest-contrast curves intersect at the spatial frequency of 11 lines per millimeter, where the response is 0.65. At 12° these curves intersect at a point where the frequency is 11.5 lines per millimeter at a lower response of 0.39. At 18° the frequency at which they intersect reaches a value of approximately 9.6 lines per millimeter at a response as low as 0.34. At 24° the two curves intersect at a spatial frequency of 17 lines per millimeter at a response of 0.65. The curves for the lens at the position of highest contrast show that the finest pattern resolved decreases in frequency at off-axis positions and reaches a minimum of 23 lines per millimeter at 18° but increases again at 24° reaching a value of 46 lines per millimeter, while the best-definition curves show little change.

The sine wave shows charaeteristics similar to those for the square wave. On-axis the behavior is as expected; but at 6° off-axis the highest contrast and best definition curves intersect at a frequency of 7.8 lines per millimeter, where the response is 0.71. At 12° the two curves intersect at a spatial frequency of 10.8 lines per millimeter with a decreased response of 0.31. At 18° off-axis the spread between the best definition and highest contrast decreases with coineidence at a spatial frequency of 8.2 lines per millimeter and a response of 0.36. As the lens is rotated further about its vertical axis until 24° off-axis is

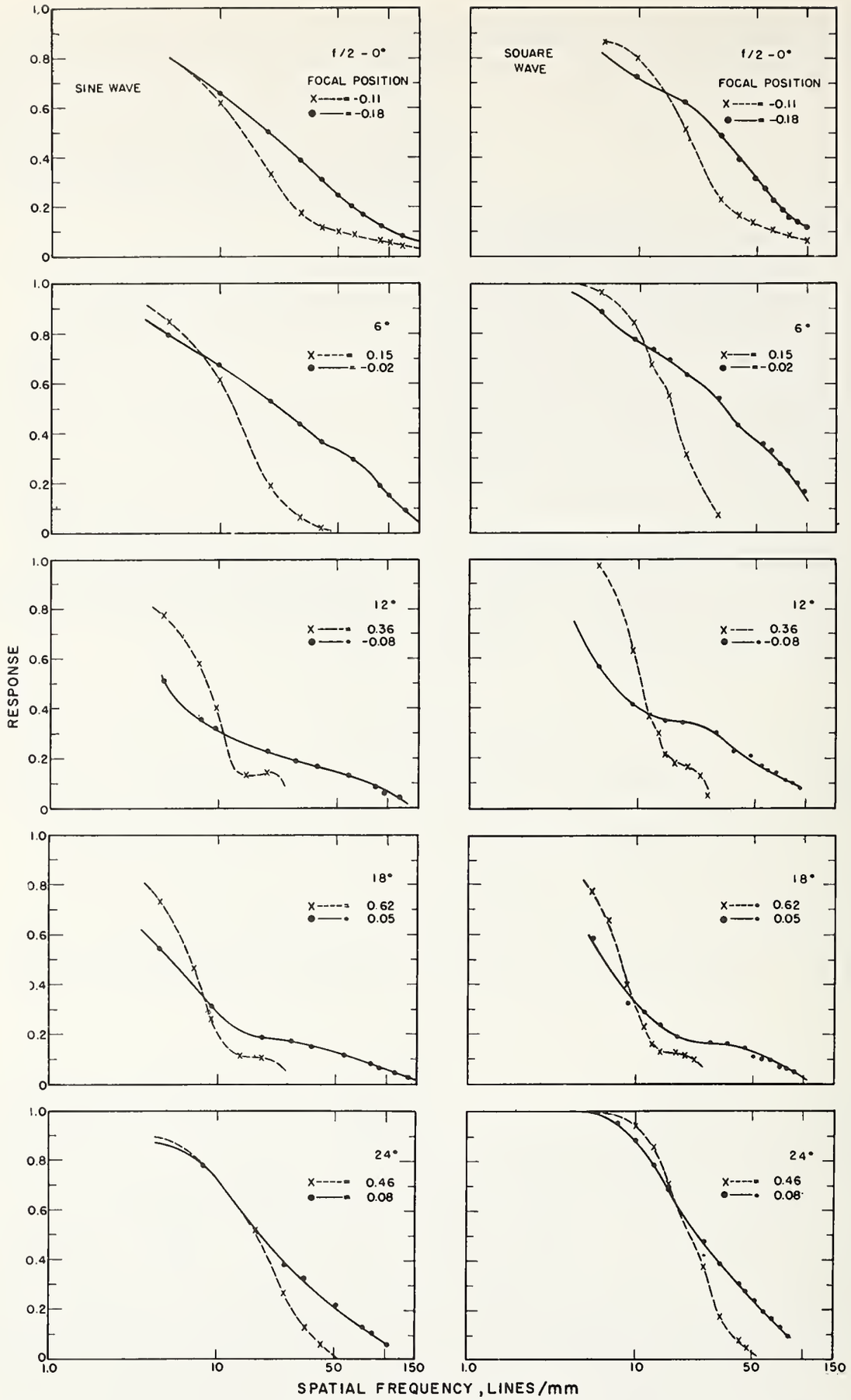


FIGURE 5. *Responses for square-wave and sine-wave targets at f/2 on axis and for four off-axis positions, 6°, 12°, 18° and 24°.*

reached the best definition and highest contrast curves intersect at a spatial frequency of 8.8 lines per millimeter with a response of 0.77. It is of interest to note that at this frequency the response has again reached a higher value, closer to that obtained on axis. The finest pattern resolved for the highest contrast curves decreases in frequency for increasing off-axis angles, reaching a minimum of 18 lines per millimeter at 18° and then increasing to a value of 42 lines per millimeter at 24°.

This suggests that for both targets there is a serious effect from aberrations reaching a maximum near 12° off-axis and then decreasing as off-axis positions approach 24°. This obviously is a consequence of the zonal nature of the corrections in photographic lenses. In this case the observations appear to be most nearly corrected on axis and at about 24° off-axis with some deterioration of image quality in between.

Figure 6 shows a set of curves for both targets using *f*/8 stop of the lens on axis and at 18° and 24° off-axis positions. It is clearly seen that the results obtained with the two targets are similar. The best definition and highest contrast curves differ considerably on axis and at 24° off-axis while the difference at 18° is smaller.

Curves at 6° and 12° with an *f*/8 stop opening are omitted because at these off-axis positions the focal positions for best definition and highest contrast were very close and no appreciable difference in response could be observed on these curves.

Further runs were made, both on- and off-axis, using the *f*/2 lens stop and four different focal positions, two of which were the positions of highest contrast and best definition. Two examples are given illustrating the effect of changing the focal positions. Figure 7 shows curves for the sine-wave target at four different focal positions on axis of the lens. The best focal position is that of best definition. Figure 8 illustrates another set of curves at 18° off-axis. Here again it is clearly seen that in the spatial frequency range of 5 to 6.8 lines per millimeter the highest contrast position (0.62) gives the highest response while in the range of 6.8 to 12 lines per millimeter the position next to highest contrast (0.40) is best. In the area 12 to 29 lines per

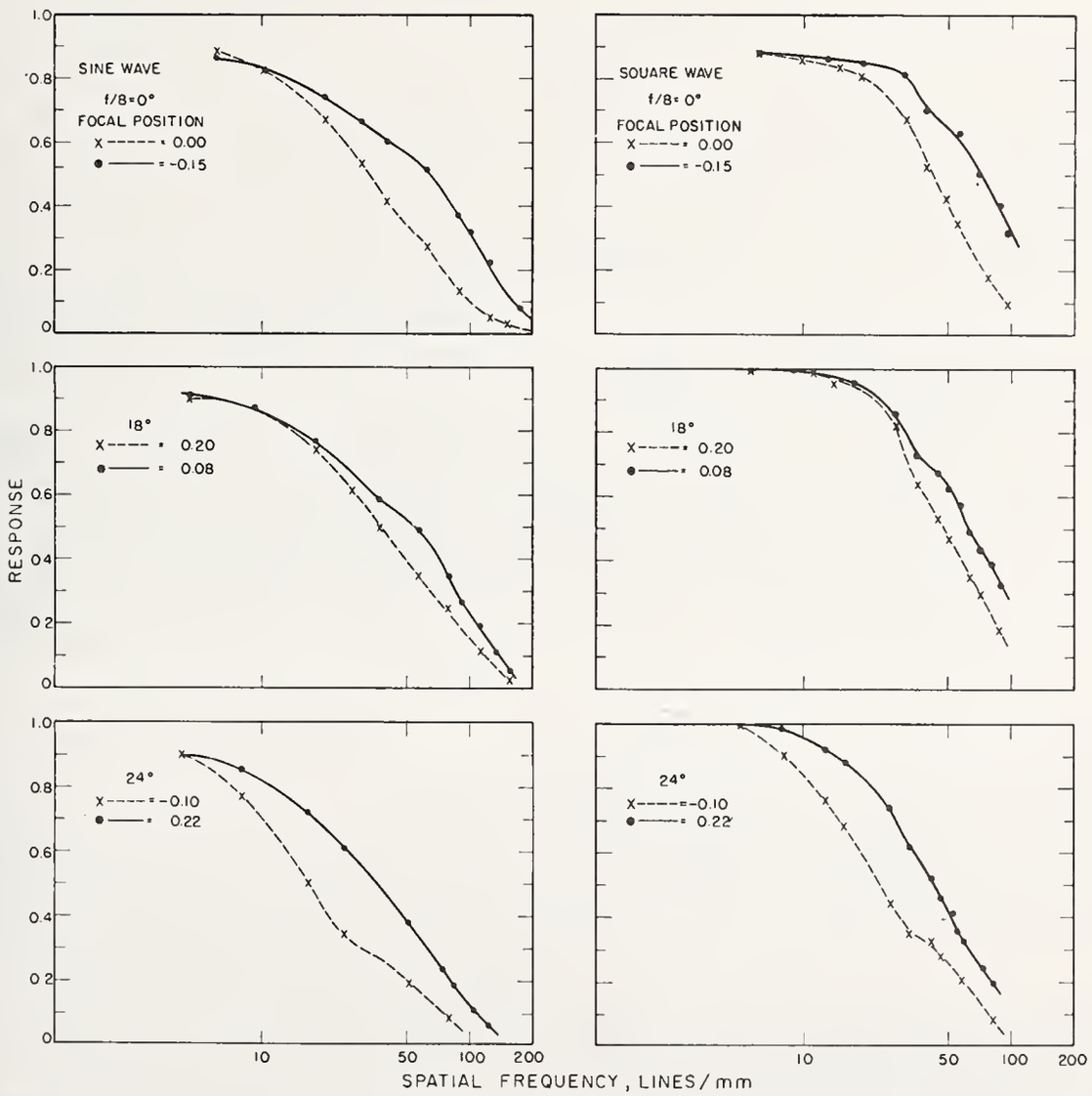


FIGURE 6. *Responses for square-wave and sine-wave targets at* f/8 *on axis and for two off-axis positions, 18° and 24°.*

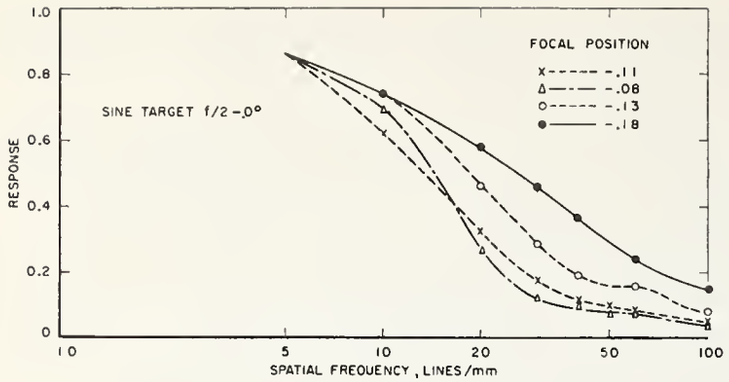


FIGURE 7. *Sine-wave response curves for four different focal positions on-axis.*

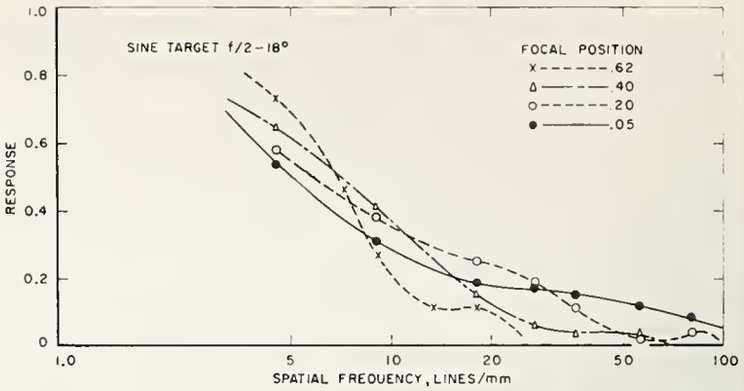


FIGURE 8. *Sine-wave response curves for four different focal positions at 18° off-axis.*

millimeter the next focal position (0.20) is best and, finally, from 29 lines per millimeter up the position of greatest definition (0.05) is best.

7. Conclusions

There are a number of conclusions which can be drawn:

1. The square-wave response can be predicted reasonably well if the sine-wave response is known. It is concluded that for a large aperture of the lens there is a certain spatial frequency at which the best definition and highest contrast curves coincide. Its value differs slightly for off-axis positions.

2. The response for both targets decreases as spatial frequency increases.

3. There is a serious effect caused by lens aberrations at off-axis positions; reaching a maximum for this lens close to 12° and decreasing again at 24°.

4. Displacement of focal positions at large apertures shows that the position of best definition may be inferior for some purposes.

5. Curves drawn for locating the focuses of best definition, and especially those for locating focuses of highest contrast become very flat topped for small apertures. Consequently these two focuses become indistinguishable. This is, of course, a manifestation of the increasing of depth of focus.

The author wishes to express her grateful thanks to the National Bureau of Standards and to A. G. McNish, chief of the Metrology Division, for providing excellent laboratory facilities and greatly appreciates the assistance of staff members. My sincere thanks are due in particular to F. W. Rosberry for his kind interest and encouragement during the course of this work and to R. E. Stephens for invaluable discussions.

8. References

[1] W. Henist, A photo-elcctric lens bench, J. Opt. Soc. Am. **37**, 472 (1947).
[2] E. Inglestam and P. J. Lindberg, A combined test procedure for camera lenses, and photoelectric examination of intensity distribution in line images, NBS Circular 526 (1951).
[3] O. Schade, A new system of measuring and specifying image definition, NBS Circular 526, 183 (1951).
[4] F. W. Rosberry, Equipment and method for photoelectric determination of image contrast suitable for using square-wave targets, J. Research NBS **64C**, 57 (1960).
[5] F. W. Rosberry, The effect of object frequency on focal position of four photographic objectives (informal communication).
[6] F. W. Rosberry, Measurement of contrast in the aerial image, Photogrammetric Eng. **27**, 155 (1961).
[7] R. V. Shack, Investigations into the correlation between photographic and photo-electric image evaluation (informal communication).
[8] R. V. Shack, Characteristics of an image-forming system, J. Research NBS **56**, 245 (1956) RP2672.

(Paper 65A6–127)

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 66C, No. 3, July–September 1962

Measurement of Longitudinal Spherical Aberration in the Extra-Axial Region of Lenses

Francis E. Washer and Walter R. Darling

(April 10, 1962)

A method of measuring longitudinal spherical and chromatic aberration in the extra-axial region of lenses is described. The method employs an especially constructed optical T-bench equipped with nodal slide and angle-measuring telescope. The determinations are based upon measurements of angular deviations in selected small regions of the collimated beam emergent from the lens under test. The underlying theory of the method is presented together with a brief description of the apparatus used and techinque of measurement. Results of measurement on three lenses are included.

1. Introduction

Longitudinal spherical aberration is one of the prime causes that prevent a given lens from yielding optimum imagery. The measurement of this aberration is therefore of some interest, as the probable image forming qualities may be inferred from an analysis of the results of such measurements. Such measurements are also of interest in a diagnostic sense in that the results of measurement serve to explain why the performance of a given lens may fall short of expectations. In an earlier paper [1] the senior author described a simple, rapid method of measuring the axial longitudinal spherical aberration in an indirect manner. In this process, the lens is used to collimate the light from a point source located in the focal plane and the angular deviations of adjacent portions of the emergent beam from parallelism are measured; the longitudinal spherical aberration is then determined by computations using the angular displacements.

The present paper extends this process to the extra-axial region. The problem is somewhat more complicated in the extra-axial region, as the lens is not rotationally symmetric about a chief ray inclined at angle β to the axis; so adequate information can be gained only by measuring in two meridians. The measurement of longitudinal chromatic aberration can readily be performed by the same process. Results of measurement are presented on three lenses.

2. Theory of the Method

The manner in which the values of longitudinal spherical aberration are derived from measurements of angular deviations from parallelism in a collimated beam of light emergent from a lens under test is described in detail in an earlier paper (see footnote 1). In this earlier paper, the discussion was confined to the situation where the luminous object was located on the optical axis at or near the focus and measurements were made in the axial collimated beam. Under these conditions, shown schematically in figure 1a, the relation governing Δf, the displacement $00'$ in the figure and ϵ, the change in angular deviation of a small portion of the collimated beam located at height h from the optical axis was shown to be

$$\Delta f = \frac{\epsilon f^2}{h} \tag{1}$$

where f is the equivalent focal length of the lens under test.

For an ideal lens, the quantity ϵ/h is invariant with varying h. For a lens affected with longitudinal spherical aberration, the quantity ϵ/h varies with h, and measurement of ϵ as a function of h permits the evaluation of longitudinal spherical abberration.

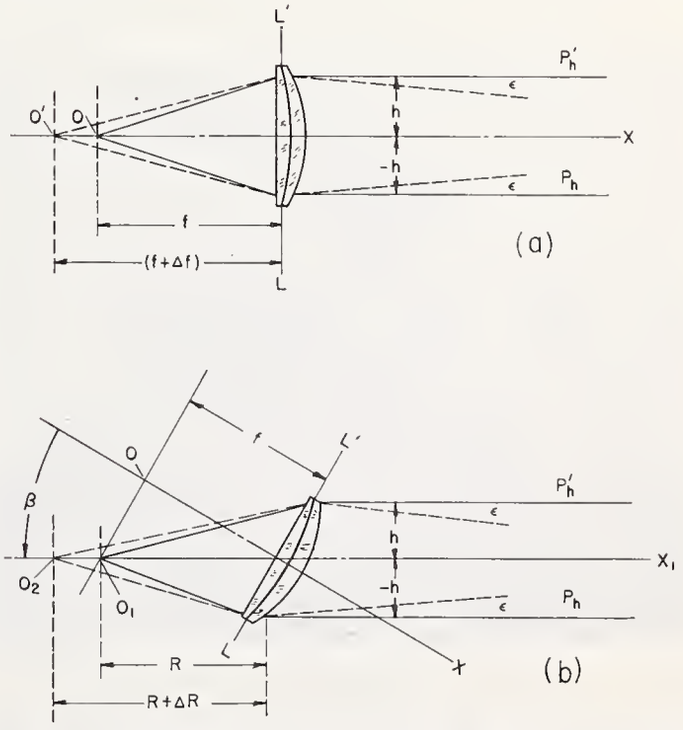


FIGURE 1. *Schematic drawing of a lens showing (a) arrangement for measurement of longitudinal spherical aberration on axis and (b) arrangement for measurement of longitudinal spherical aberration along a chief ray inclined at angle β to the optical axis.*

[1] F. E. Washer, Optical T-bench method of measuring longitudinal spherical aberration. J. Research NBS **61**, 31 (1958) RP2880.

In the present study, the chief interest is in the magnitude of the longitudinal spherical aberration as measured along a chief ray inclined at angle β to the optical axis of the lens system. It is clear from figure 1b, that the same principles are involved as on the axis. Consequently the following modification of eq (1),

$$\Delta R = \frac{\epsilon R^2}{h}, \tag{2}$$

is satisfactory in determining the displacement $0_1 0_2$ of the focus along the chief ray with zone height h. In the above equation,

$$R = f \sec \beta, \tag{3}$$

h is measured normal to the chief ray, and $\Delta R = 0_1 0_2$ is measured along the chief ray.

3. Method of Measurement

3.1. In the Tangential Fan

The method of measurement is essentially the same as that for determination of axial longitudinal spherical aberration. However, some changes in procedure are necessary to make measurements along the chief ray inclined at angle β to the axis. For measurements in the tangential fan, the apparatus is arranged as shown in figure 2. Initially the lens is placed on the nodal slide with its front facing the pentaprism through which the angular deviations are viewed. The lens is adjusted until its rear nodal point is in the vertical axis of rotation of the nodal slide and the illuminated target reticle is located at the rear focal point of the lens. In addition, the optieal axis of the lens is set parallel to the bench ways upon which the lens and target supports are located. When the target is illuminated, the light proceeding from the target falls upon the rear surface of the lens and emerges from the front as a collimated beam of light. This collimated beam is incident upon one face of the pentaprism mounted on the ways of the second bench and is directed into the viewing telescope located as shown in figure 2.

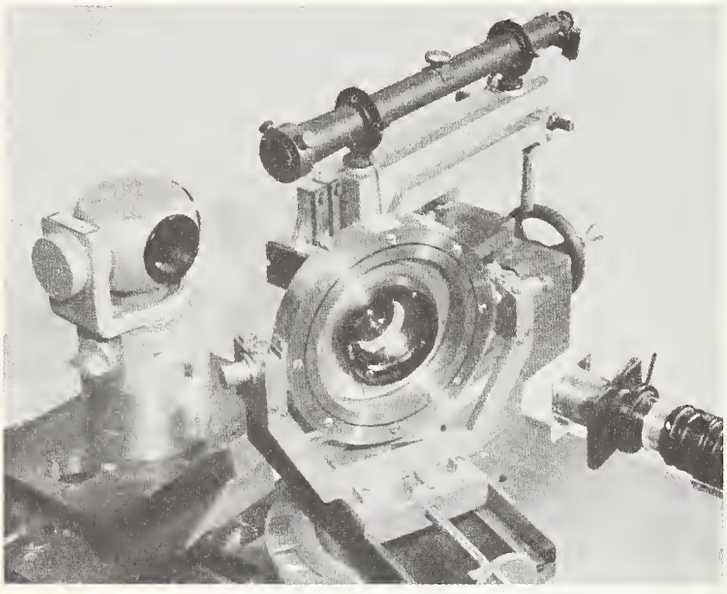


FIGURE 2. *Test equipment arranged for measurements in the tangential fan.*

The lens under test is shown mounted on the nodal slide in the eenter of the photograph; the lens has been rotated through an angle β about the vertical axis. The target reticle and illuminating system are shown at the right with the target reticle facing the rear of the lens. The parallel beam of light emergent from the front of the lens is turned through 90° by the pentaprism in the left of the seene into the viewing teleseope at the rear. The diaphragm eovering the telescope objeetive limits the beam to a small peneil of rays.

The telescope objective is equipped with a diaphragm having a small central aperture which is centered with respect to the telescope objective. When an observer looks through the telescope, he sees the image of the illuminated target (either a cross or pinhole pattern) superimposed on the crosshairs in the ocular of the viewing telescope. When the slide carrying the pentaprism is moved along the bench ways, the observer sees the target through successive small areas of the lens along its diameter. In the event of movement of the image, it can be brought into coincidence with the crosshairs by appropriate movement of the transverse micrometer. Movement of the transverse micrometer changes the direction of pointing of the viewing telescope. The lateral movement of the micrometer is directly proportional to the angular change in pointing. Hence the change in pointing is a measure of the change in direction of a small circular beam of light emergent from a given area of the lens under test from the direction of another similar area lying along the same diameter of the lens. This angular deviation ϵ is measured in scale divisions read on the transverse mierometer. The value of ϵ in scale divisions may be converted to radians on multiplying by the calibration constant of the micrometer. For comparison purposes the value of ϵ in scale divisions may be used directly, but is usually converted to radians for the final evaluation of the longitudinal spherical aberration.

It is customary to select a size of diaphragm opening small in comparison with the area of the lens under test and to select a series of steps so that measurements of the deviation of these successive small emergent beams can be made for an entire diameter of the lens. If the lens under test has been properly focused, the angular deviations noted arise from parallax (or angular spherical aberration) of the lens under test.

The seale on the bench carrying the pentaprism is used in making the successive settings at selected intervals. These settings are recorded and the difference of a given bench scale reading from the one in the central position is the zone height, h. The readings of the angular micrometer are taken for each value of h and the differenee of a given micrometer reading from that one in the central position is recorded as ϵ.

The longitudinal spherieal aberration usually varies with the wavelength of the image forming light. It is therefore necessary when making measurements to know the wavelength of the light used. While it would be desirable to use monochromatic light, useful measurements can be obtained

TABLE 1. *Filters used in controlling illumination of target reticle* *

Filter Nos. 58 and 25 transmit a moderately wide band of wavelengths; filter No. 73 is a narrow band filter. The column headed dominant wavelength gives the value of the wavelength that appears to be dominant in the light transmitted by the filter when illuminated by standard illuminant "*C*".

Filter No.	Description	Dominant wavelength mμ
58	Tricolor green	540.2
73	Yellow-green (narrow band)	574.9
25	Tricolor red	615.1

*Handbook of Chemistry and Physics (37th Ed.) pp. 2730–2734, Chemical Rubber Publ. Co., Cleveland, Ohio (1955–56).

using filters to control the quality of the light. Filters of the type used in photographic processes are suitable for the purpose. Moreover, results obtained with these filters present a picture of the manner in which longitudinal spherical aberration is operative under conditions of use. In the present study, three Wratten light filters were used; their characteristics are given in table 1. Filter No. 73 is a narrow-band filter while filters Nos. 58 and 25 transmit relatively broad regions of the spectrum. These filters are interposed between the primary light source and the target reticle, hence the spectral characteristics of the light emerging from the lens under test are determined by the filter used.

The foregoing procedure permits evaluation of longitudinal spherical aberration, Δf, on the axis of the lens. To make measurements of longitudinal spherical aberration for the tangential fan of a chief ray inclined at angle β to the axis, it is only necessary to rotate the lens through the angle β about the vertical axis of the nodal slide and to move the target away from the lens by amount f (sec $\beta-1$) which places the target in the focal plane at distance f tan β from the focal point. Measurements are then made in the same manner as was done for the axial region. A typical set of measurements is shown in table 2 for measurements made in the tangential fan for a chief ray inclined at angle $\beta=20°$ to the axis of the lens. In this table, the column headed zone height h gives the intervals along the bench scale at which observations were made. The value $h=0$ corresponds to the center of the emergent beam, plus values of h indicate positions to one side, and negative values positions to the opposite side, of center. For each value of h, observed values of the angular deviation ϵ using each of three filters are listed under the filter Nos. 58, 73, and 25; the characteristics of these filters are given in table 1. For positive values of h, negative values of ϵ indicate that the emergent beam is diverging or that the target is nearer to the lens than the focus. More generally when the sign of ϵ is the same as that of h, the emergent beam is converging; and when the sign of ϵ is the negative of the sign h, the emergent beam is diverging. The values of ϵ shown in table 2 indicate that the emergent beam is diverging and that the target is nearer to the lens than the focus for every zone. This is confirmed when the displacement ΔR_T of the target from the focal point is determined with the aid of

TABLE 2. *Variation of parallax angle ϵ and apparent target displacement ΔR_T as a function of zone height* h *for a chief ray inclined 20° with respect to the axis of the lens*

These values were obtained for lens No. 1, a photographic objective having a focal length of 152.4 mm. The values of ϵ were measured in the tangential fan for angle B=20°. Values are given for three conditions of target illumination determined by Wratten filters Nos. 58, 73, and 25.

Zone height h	ϵ for filter Nos.			ΔR_T for filter Nos.		
	58	73	25	58	73	25
mm	*sc. div.*	*sc. div.*	*s. div.*	*mm*	*mm*	*mm*
9	−34.5	−28.1	−23.5	−2.76	−2.25	−1.88
8	−29.8	−24.5	−19.4	−2.68	−2.20	−1.74
7	−23.7	−19.6	−15.0	−2.44	−2.01	−1.54
6	−17.1	−13.9	−10.3	−2.05	−1.66	−1.23
5	−11.9	−9.5	−6.4	−1.71	−1.37	−0.92
4	−7.9	−5.9	−3.5	−1.42	−1.06	−.63
3	−5.0	−3.3	−1.5	−1.20	−0.79	−.36
2	−2.6	−1.3	−0.5	−0.93	−.46	−.18
1	−1.2	−1.1	−.3	−.86	−.79	−.21
0	0	0	0	------	------	------
−1	1.4	0.5	0.2	−1.00	−.36	−.14
−2	3.3	2.0	.5	−1.18	−.71	−.18
−3	5.0	2.9	1.2	−1.20	−.70	−.29
−4	8.1	5.7	3.1	−1.46	−1.03	−.56
−5	14.4	11.0	6.4	−2.07	−1.58	−.92
−6	20.5	16.5	12.1	−2.45	−1.97	−1.45
−7	25.9	21.0	16.3	−2.66	−2.16	−1.68
−8	30.8	26.1	20.0	−2.77	−2.35	−1.80
−9	38.5	30.8	25.9	−3.08	−2.46	−2.07
Average				−1.88	−1.44	−0.99

eq (2). For these calculations, it is convenient to use the following modification of eq (2):

$$\Delta R_T=\frac{k\epsilon f^2 \sec^2\beta}{h}, \tag{4}$$

where $k=27.35\times10^{-6}$ radians/scale division is used to convert ϵ from scale division to radians, f is the focal length of the lens, and β is the inclination of the chief ray to the lens axis. Values of ΔR_T derived from the observed values of ϵ listed in table 2 are shown in the same table under the heading ΔR_T for filters Nos. 58, 73, and 25.

In table 2, the values of ΔR_T for each value of h and ϵ are negative as were indicated by opposing signs of h and ϵ. It is clear that the magnitude of ΔR_T varies appreciably with h. For comparison purposes, we may define the average of ΔR_T obtained with a given filter as the position of best focus with respect to the focal plane for use with that filter. In the present instance, it is noteworthy that the average value of ΔR_T is different for each filter and it is negative for each filter. Since the target was originally located in the plane of best axial focus, it is clear that for image-forming light incident upon the lens at an angular inclination of $\beta=20°$ the plane of best focus for use with any one of these three filters is nearer to the lens than the plane of best axial focus.

This displacement of the plane of optimum focus in the extra-axial region from the plane of best axial focus arises from curvature of field and could not necessarily be anticipated at the start of the measurements. However, it does not invalidate the results, as variation of ϵ may arise from a combination of two effects (1) out-of-focus effect and (2) angular spherical

aberration or parallax resulting from longitudinal spherical aberration inherent in the lens. These two effects are easily separable, as out-of-focus effect produces a variation of ϵ proportional to h which expressed graphically plots in a straight line of constant slope passing through the zero point. In terms of ΔR_T, it simply means the addition of a constant to all terms or displacing the entire longitudinal spherical aberration curve by a constant amount when ΔR_T is plotted as a function of h.

Examination of the values of ΔR_T for any one of the three sets of data shown in table 2 reveals an asymmetry in the manner in which ΔR_T varies with h on opposite sides of the zero point of h. Thus in the column headed 73, $\Delta R_T = -2.25$ mm at $h = +9$ mm and $\Delta R_T = -2.46$ mm at $h = -9$.mm. Differences of this nature may be ascribed to anyone of, or combination of, three causes which are (1) random error, (2) error in location of the center point, $h=0$; and (3) asymmetric aberrations in the lens. The effects of all three sources of error may be minimized by averaging the values of ΔR_T for paired values of $+h$ and $-h$. The results of such averaging of the values of ΔR_T given in table 2 are shown in table 3. These values of ΔR_T shown in table 3 are therefore the accepted values of the longitudinal spherical aberration plus a constant which measures the departure of the position of best focus for rays in the tangential fan from the plane of best axial focus for a chief ray inclined at angle $\beta = 20°$ to the axis. As a check upon the accuracy of measurements, it is customary to repeat the entire series of measurements for the chief ray inclined at angle $\beta = -20°$ to the axis. Since the lens is symmetric about the optical axis, the values found for $\beta = -20°$ are usually sufficiently nearly identical to the values found for $\beta = +20°$ that the results may be averaged to produce the final accepted values of longitudinal spherical aberration for the chief ray inclined at $\beta = 20°$ to the axis. A graphical presentation showing both the averaged values of ϵ versus h (frame A) and ΔR_T versus h (frame B) for $\beta = +20°$ and $\beta = -20°$ is given in figure 3. In this presentation, there was no averaging of values of ϵ and ΔR_T for paired values of $+h$ and $-h$, consequently the effect of small asymmetric aberrations are still present. In frame B of figure 3, the effect of this asymmetric aberration is to tilt the whole array of ΔR_T curves slightly in a clockwise manner. In the final determination, average values of ΔR_T are obtained from paired values of $+\beta$ and $-\beta$, and for paired values of $+h$ and $-h$.

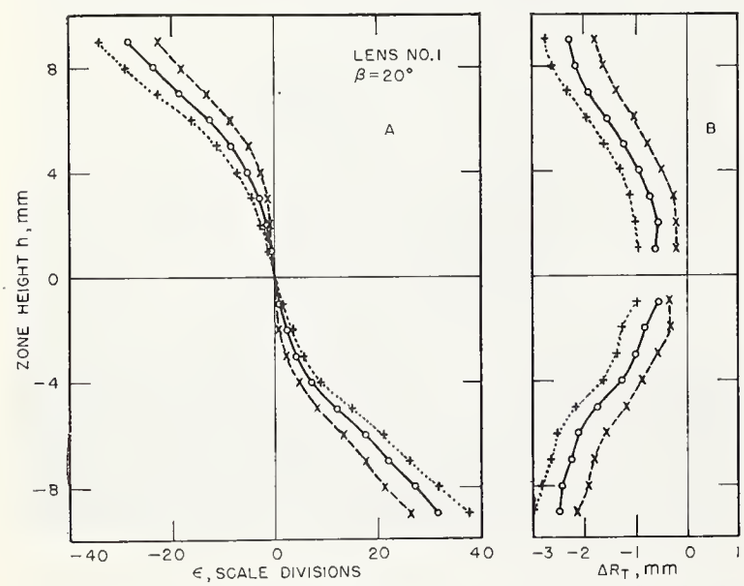


FIGURE 3. *Variation of parallax angle, ϵ, with zone height,* h, *(frame A) and derived longitudinal spherical aberration,* ΔR_T, *versus zone height* h *(frame B).*

Results are given for a chief ray inclined at $\beta = 20°$ to the optical axis. In each of the frames, the dotted-line curves show results obtained with light transmitted by filter No. 58; the solid-line curves show results for light transmitted by filter No. 73; and the dashed-line curves show results for light transmitted by filter No. 25

TABLE 3. *Values of* ΔR_T *and* $\Delta(\Delta R_T)$ *as a function of zone height* h *for a chief ray inclined at 20° to the axis of the lens*

These results are for lens No. 1. The values of ΔR_T are derived from table 2 by averaging the values obtained at the same absolute value of zone height h. The values obtained with each of three filters (Wratten Nos. 58, 73, and 25) are shown. Negative values of ΔR_T indicate apparent displacement of the target reticle along the chief ray and toward the lens with respect to the focal plane. Values of $\Delta(\Delta R_T)$ are given for the three filter pairs. All values are expressed in mm.

Zone height h	ΔR_T for filter No.			$\Delta(\Delta R_T)$ for filter pairs Nos.		
	58	73	25	73–58	25–73	25–58
1	−0.93	−0.58	−0.18	0.35	0.40	0.75
2	−1.06	− .58	− .18	.48	.40	.88
3	−1.20	− .74	− .32	.46	.42	.88
4	−1.44	−1.04	− .60	.40	.44	.84
5	−1.89	−1.48	− .92	.41	.56	.97
6	−2.25	−1.82	−1.34	.43	.48	.91
7	−2.55	−2.08	−1.61	.47	.47	.94
8	−2.72	−2.28	−1.77	.44	.51	.95
9	−2.92	−2.36	−1.98	.56	.38	.94
Average	−1.88	−1.44	−0.99	0.44	0.45	0.89
PE_s				±0.04	±0.04	±0.05

3.2. In the Sagittal Fan

To make measurements of longitudinal spherical aberration for the sagittal fan of a chief ray inclined at angle β to the axis, it is necessary to rotate the lens about a horizontal axis normal to the optical axis and passing through the rear nodal point of the lens and to move the target away from the lens by an amount $f(\sec \beta - 1)$. This places the target at distance $f \sec \beta$ from the rear nodal point of the lens as measured along the chief ray inclined at angle β to the axis and at distance $f \tan \beta$ from the axial focal point as measured in the focal plane. The arrangement of apparatus is shown in figure 4. Measurements of ϵ versus h are then made in the same manner as for the axial region. In order that the measurements are made for the same region of the picture area as those made in the tangential fan, the lens is also rotated 90° about its own optical axis from the position occupied during measurements in the tangential fan. Except for the foregoing differences, the process of measurement is essentially the same as that employed for the tangential fan. Values of the displacement ΔR_S of the target from the focal point of the lens for the given chief ray are obtained with the aid of eq (4). The final accepted values of ΔR_s are obtained from the

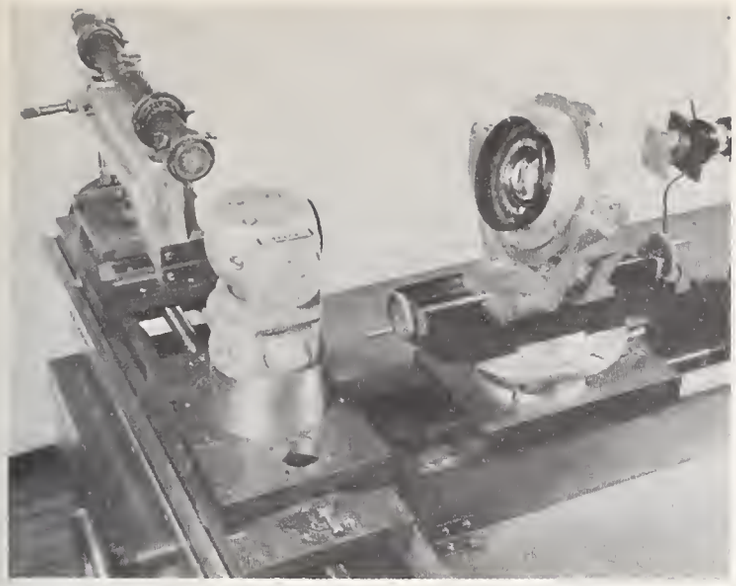


FIGURE 4. *Test equipment arranged for measurement in the sagittal fan.*

The same equipment appearing in figure 2 is shown with the lens rotated through an angle β about the horizontal axis.

average of paired values for $+\beta$ and $-\beta$ and for paired values of $+h$ and $-h$.

4. Results of Measurement

The results of measurement on three wide-angle lenses having focal lengths of approximately 150 mm are presented in this study. Two of the lenses, designated Nos. 1 and 2, are essentially distortion-free while the third lens, designated No. 3, has a moderate amount of distortion.

4.1. Longitudinal Spherical Aberration

The values of the longitudinal spherical aberration for both the tangential and sagittal fans are shown graphically in figures 5, 6, and 7. In the upper series of frames of each figure, average values of the displacement ΔR_T for the tangential fan are given as a function of zone height h at angular inclinations of the chief ray from the axis 0, 10, 20, 30, and 40 degrees. Results for the same series of filters are presented for each lens for each angular inclination. In each of the figures, the dotted-line curves show the results obtained with light transmitted by filter No. 58, the solid line curves show the results obtained for light transmitted by filter No. 73; and the dashed line curves show the results obtained for light transmitted by filter No. 25. The optical characteristics of these filters are given in table 1. In the lower series of frames in figures 5, 6, and 7 values of the displacement ΔR_s for the sagittal fan as a function of zone height h, are given for the same angular inclinations of the chief ray to the optical axis and for light transmitted by the same three filters. In the graphs, values of ΔR_s and ΔR_s are plotted for both positive and negative values of h. This was done to increase clarity for, although the lower portion of a given curve is a mirror image of the upper portion, the inclusion of both upper and lower portion permits a better visualization of the phenomenon. In contemplating these curves, it is interesting to note the variation in shape with increasing β. In all cases shown, the curves of ΔR_T versus h become increasingly concave toward the axis and the curves of ΔR_s versus h become increasingly convex with respect to the axis with increasing values of β.

Too, it is evident that the zero points of ΔR_T and ΔR_s do not always maintain the same relative position with respect to the curves of ΔR_T and ΔR_s for various values of β. This shift results from curvature of field and the magnitude of the shift is a measure of tangential field curvature in the case of the ΔR_T values of the sagittal field curvature in the case of ΔR_s values. It may also be noted that the range of values of h decreases with β in the tangential fan but remains constant for the sagittal fan. This reduction in range is caused by vignetting which has a more pronounced effect in the tangential fan.

4.2. Longitudinal Chromatic Aberration

In figures 5, 6, and 7, it is clear that the longitudinal spherical aberration curves obtained with the various filters for a given value of β are very similar in appearance but are displaced along the chief ray with respect to one another. This displacement of the curves with respect to one another may be ascribed to longitudinal chromatic aberration as the effective wavelength of the light illuminating the target reticle varies from one filter to another as may be inferred from the information contained in table 1. Hence it is reasonable to suppose that the magnitude of the displacement of one curve with respect to another may be taken as a measure of the change in focal position along the chief ray produced by the change in character of the light transmitted by one filter as compared to another. While strictly speaking the term "longitudinal chromatic aberration" applies only to differences in focal position induced by change in the nature of the image forming light from one monochromatic illuminant having a sharply defined wavelength to another monochromatic illuminant having a different sharply defined wavelength, it is nonetheless at times expedient to use the term when dealing with image forming light whose nature is determined by a filter. Hence in this study the term "longitudinal chromatic aberration" will be used as a convenient mnemonic to designate the change in focal position produced by a change in the filter controlling the natures of the illuminant. With this as a basis, the longitudinal chromatic aberration in the tangential fan for a given filter pair may be determined zone-by-zone from the relation

$$\Delta(\Delta R_T)=\Delta R_T-\Delta R'_T \tag{5}$$

where ΔR_T is the longitudinal spherical aberration in the tangential fan determined for a specified value of h when using a given light filter and $\Delta R'_T$ is that determined for the same value of h when using

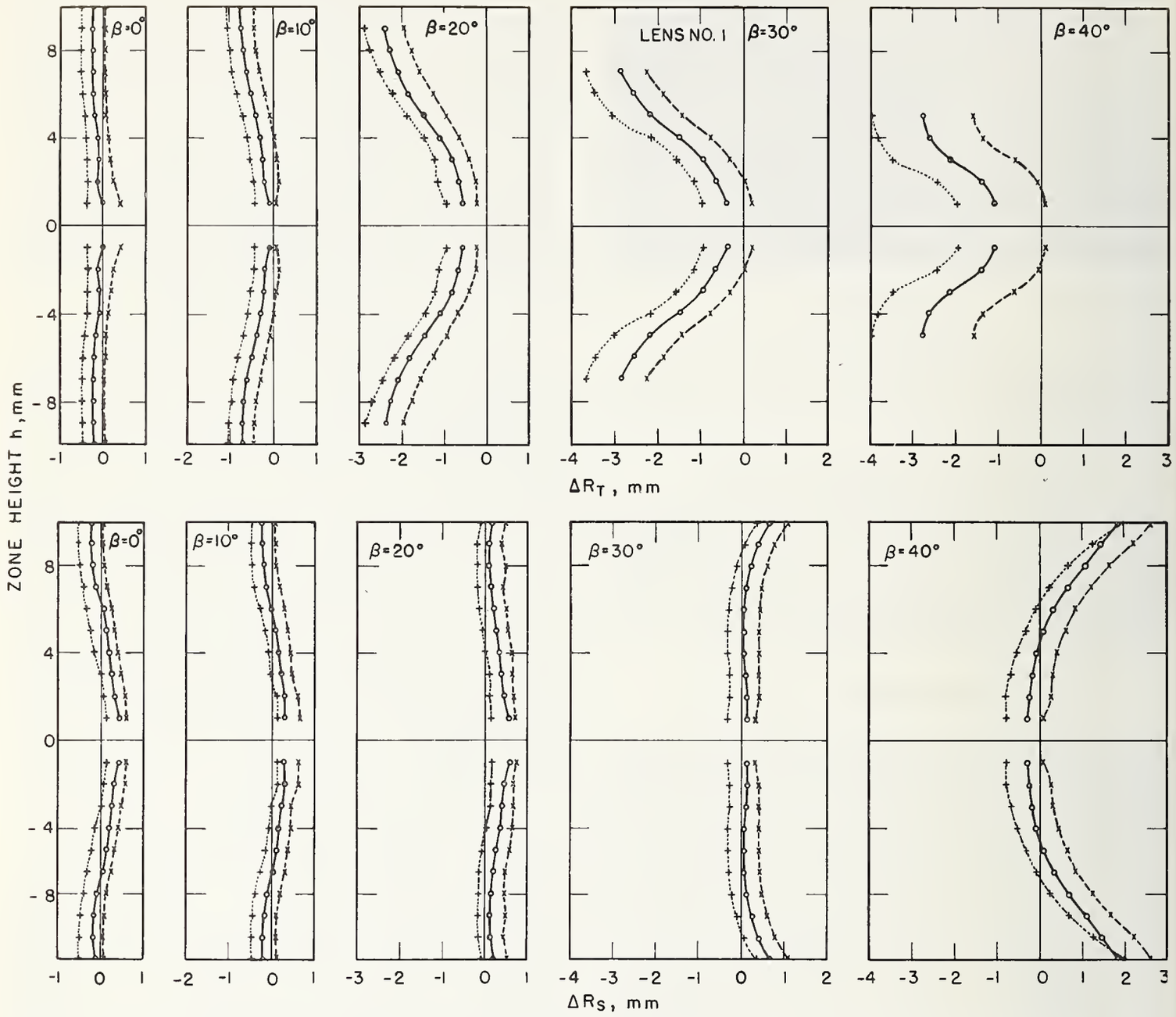


FIGURE 5. *Variation of longitudinal spherical aberration,* ΔR_T, *in the tangential fan and,* ΔR_s, *in the sagittal fan with zone height,* h, *for five values of* β.

The results are for lens No. 1, a wide-angle distortion-free lens having a focal length of 152.4 mm. The series of curves in the upper frames show the variation of ΔR_T with zone height for five values of β; the series of curves in the lower frames show the variation of ΔR_s with zone height for five values of β. In each of the frames the dotted-line curves show results obtained with light transmitted by filter No. 58; the solid-line curves show results obtained with light transmitted by filter No. 73; and the dashed-line curves show results obtained with filter No. 25.

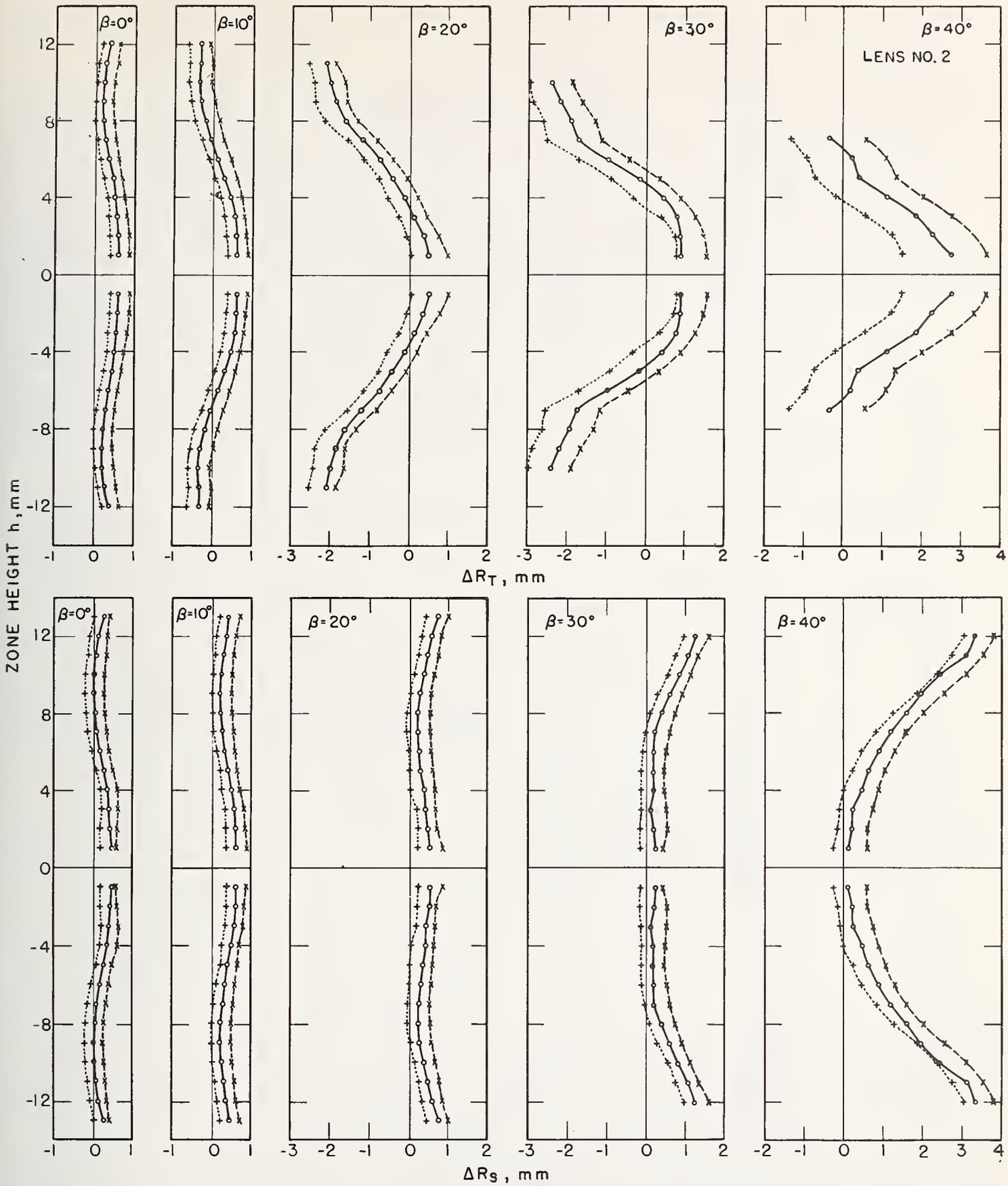


FIGURE 6. *Variation of longitudinal spherical aberration,* ΔR_T, *in the tangential fan and,* ΔR_s, *in the sagittal fan with zone height,* h, *for five values of* β.

The results are for lens No. 2, a wide-angle distortion-free lens having a focal length of 153.3 mm. The series of curves in the upper frames show the variation of ΔR_T with zone height for five values of β; the series of curves in the lower frames show the variation of ΔR_s with zone height for five values of β. In each of the frames, the dotted-line curves show results obtained with light transmitted by filter No. 58; the solid-line curves show results obtained with light transmitted by filter No. 73; and the dashed-line curves show results obtained with filter No. 25.

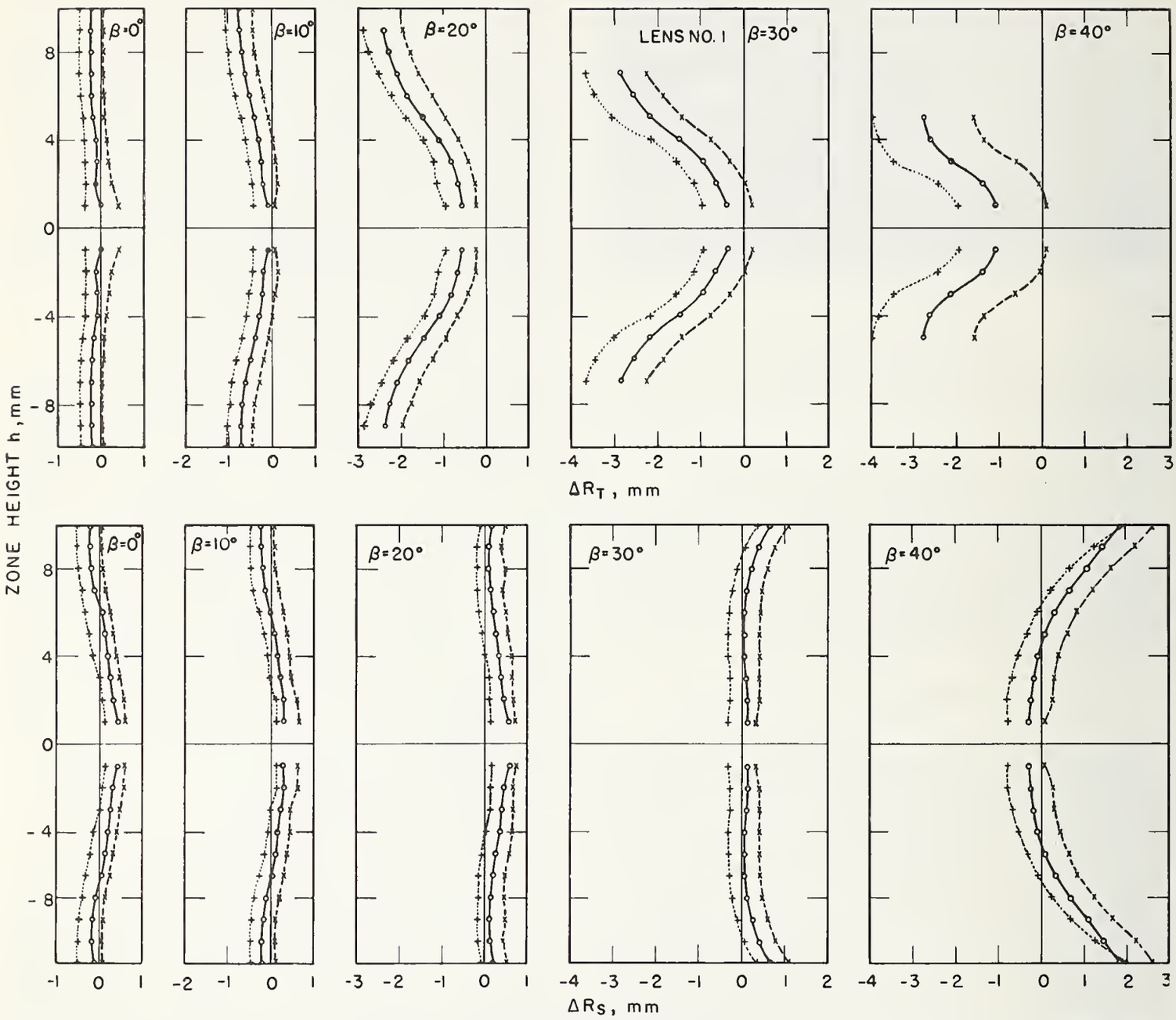


FIGURE 5. *Variation of longitudinal spherical aberration,* ΔR_T, *in the tangential fan and,* ΔR_s, *in the sagittal fan with zone height,* h, *for five values of* β.

The results are for lens No. 1, a wide-angle distortion-free lens having a focal length of 152.4 mm. The series of curves in the upper frames show the variation of ΔR_T with zone height for five values of β; the series of curves in the lower frames show the variation of ΔR_s with zone height for five values of β. In each of the frames the dotted-line curves show results obtained with light transmitted by filter No. 58; the solid-line curves show results obtained with light transmitted by filter No. 73; and the dashed-line curves show results obtained with filter No. 25.

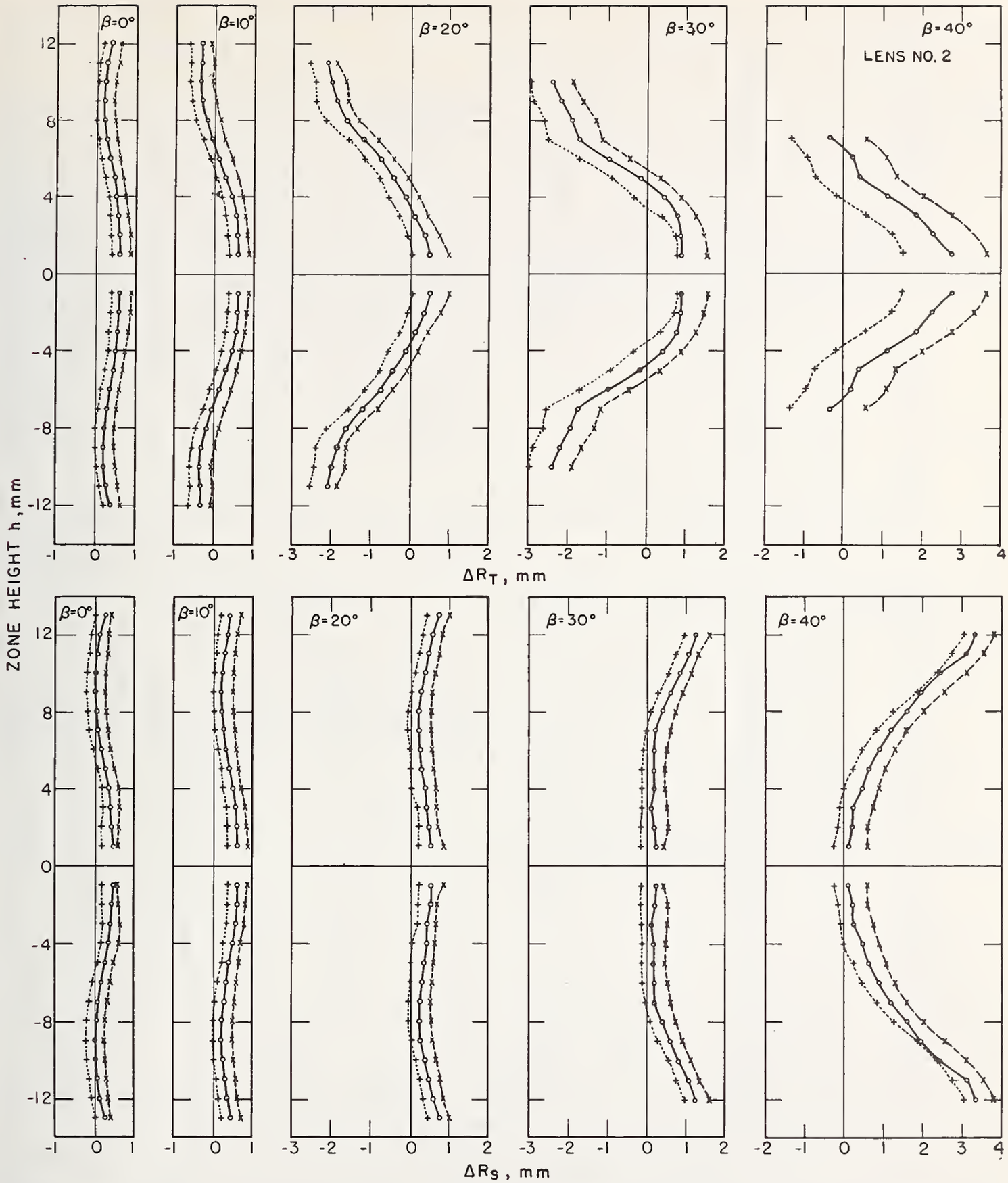


FIGURE 6. *Variation of longitudinal spherical aberration,* ΔR_T, *in the tangential fan and,* ΔR_s, *in the sagittal fan with zone height,* h, *for five values of* β.

The results are for lens No. 2, a wide-angle distortion-free lens having a focal length of 153.3 mm. The series of curves in the upper frames show the variation of ΔR_T with zone height for five values of β; the series of curves in the lower frames show the variation of ΔR_s with zone height for five values of β. In each of the frames, the dotted-line curves show results obtained with light transmitted by filter No. 58; the solid-line curves show results obtained with light transmitted by filter No. 73; and the dashed-line curves show results obtained with filter No. 25.

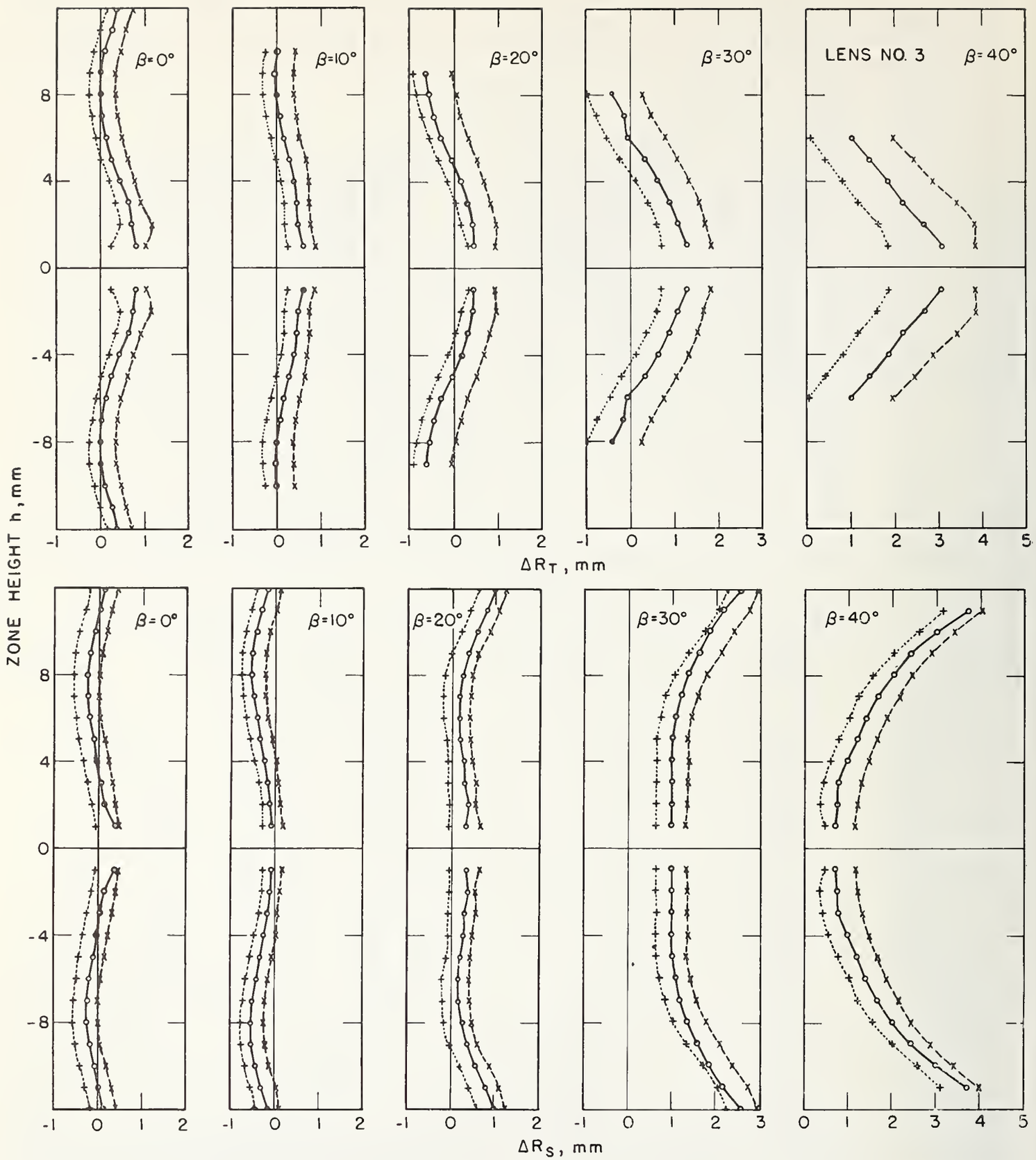


FIGURE 7. *Variation of longitudinal spherical aberration,* ΔR_T, *in the tangential fan and,* ΔR_s, *in the sagittal fan with zone height,* h, *for five values of* β.

The results are for lens No. 3, a wide-angle lens having a focal length of 153.3 mm. The series of curves in the upper frames show the variation of ΔR_T with zone height for five values of β; the series of curves in the lower frames show the variation of ΔR_s with zone height for five values of β. In each of the frames, the dotted-line curves show results obtained with light transmitted by filter No. 58; the solid-line curves show results obtained with light transmitted by filter No. 73; and the dashed-line curves show results obtained with filter No. 25.

a different light filter. In a similar manner, the relation

$$\Delta(\Delta R_s)=\Delta R_s-\Delta R_s' \tag{6}$$

may be used to determine the longitudinal chromatic aberration in the sagittal fan. It is elear from figures 5, 6, and 7 that for a given value of β, the quantities $\Delta(\Delta R_T)$ and $\Delta(\Delta R_s)$ do not vary appreeiably with zone height for these three lenses.

The invariance of $\Delta(\Delta R_T)$ with zone height is also indieated in table 3 where values of $\Delta(\Delta R_T)$ are given for three filter pairs. The average values of $\overline{\Delta(\Delta R_T)}$ are also given together with the probable error of a single observation (PE_s). In view of the various errors inherent in the measurement of ΔR_T, it is reasonable to eonelude from the low values of PE_s that the quantity $\Delta(\Delta R_T)$ may be regarded as invariant for these lenses. This leads to a simplification in the determination of $\overline{\Delta(\Delta R_T)}$ as one may determine the average value of ΔR_T for each filter used and evaluate $\overline{\Delta(\Delta R_T)}$ for any pair of filters from the relation

$$\overline{\Delta(\Delta R_T}=\overline{\Delta R_T}-\overline{\Delta R_T'}. \tag{7}$$

In a similar manner, the average value of $\overline{\Delta(\Delta R_s)}$ for the sagittal fan for a given value of β may be determined from the relation

$$\overline{\Delta(\Delta R_s)}=\overline{\Delta R_s}-\overline{\Delta R_s'}. \tag{8}$$

TABLE 4. *Values of* $\overline{\Delta R_T}$, $\overline{\Delta R_s}$, $\overline{\Delta(\Delta R_T)}$, *and* $\overline{\Delta(\Delta R_s)}$ *for the tangential and sagittal fans as a function of* β

These values are for lens No. 1 a wide-angle distortion-free lens having a foeal length of 152.4 mm. Part (a) shows the values of $\overline{\Delta(R_T)}$ and $\overline{\Delta(\Delta R_T)}$ for the tangential fan and part (b) shows the eorresponding values for the sagittal fan. Values are given for three eonditions of target illumination determined by the indieated filter numbers. All values are expressed in mm.

(a) Tangential fan

β degrees	$\overline{\Delta R_T}$ for filter No.			$\overline{\Delta(\Delta R_T)}$ for filter pairs Nos.		
	58	73	25	73–58	25–73	25–58
0	−0.41	−0.15	0.14	0.26	0.29	0.55
5	−.46	−.17	.08	.29	.25	.54
10	−.76	−.45	−.16	.31	.29	.60
15	−1.27	−.90	−.57	.37	.33	.70
20	−1.88	−1.47	−1.01	.41	.46	.87
25	−2.21	−1.64	−1.22	.57	.42	.99
30	−2.26	−1.58	−0.90	.68	.68	1.36
35	−2.44	−1.53	−.73	.91	.80	1.71
40	−3.12	−2.01	−.72	1.11	1.29	2.40
45	−5.76	−4.42	−2.84	1.34	1.58	2.92

(b) Sagittal fan

β degrees	$\overline{\Delta R_s}$ for filter No.			$\overline{\Delta(\Delta R_s)}$ for filter pairs Nos.		
	58	73	25	73–58	25–73	25–58
0	−0.24	0.08	0.30	0.32	0.22	0.54
5	.02	.28	.52	.26	.24	.50
10	−.20	.04	.32	.24	.28	.52
15	−.04	.24	.48	.28	.24	.52
20	−.02	.27	.54	.29	.27	.56
25	−.30	.02	.32	.32	.30	.62
30	−.18	.19	.52	.37	.33	.70
35	−.08	.32	.71	.40	.39	.79
40	.09	.47	1.02	.38	.55	.93
45	−.04	.54	1.02	.58	.48	1.06

Further eonsideration of the curves in figures 5, 6, and 7 indieates that both $\overline{\Delta(\Delta R_T)}$ and $\overline{\Delta(\Delta R_s)}$ increase with increasing β and that for a given value of β the value of $\overline{\Delta(\Delta R_T)}$ is greater than that of $\overline{\Delta(\Delta R_s)}$. The variation of $\overline{\Delta(\Delta R_T)}$ and $\overline{\Delta(\Delta R_s)}$ with β is shown in tables 4, 5, and 6 for three filter pairs for lenses Nos. 1, 2, and 3. In each table, the values of $\overline{\Delta R_T}$ and $\overline{\Delta R_s}$ obtained for light trans-

TABLE 5. *Values of* $\overline{\Delta R_T}$, $\overline{\Delta R_s}$, $\overline{\Delta(\Delta R_T)}$, *and* $\overline{\Delta(\Delta R_s)}$ *for the tangential and sagittal fans as a function of* β.

These values are for lens No. 2, a wide-angle distortion-free lens having a foeal length of 153.3 mm. Part (a) shows the values of $\overline{\Delta R_T}$ and $\overline{\Delta(\Delta R_T)}$ for the tangential fan and part (b) shows the eorresponding values for the sagittal fan. Values are given for three eonditions of target illumination determined by the indieated filter Nos. All values are expressed in mm.

(a) Tangential fan

β degrees	$\overline{\Delta R_T}$ for filter No.			$\overline{\Delta(\Delta R_T)}$ for filter pairs Nos.		
	58	73	25	73–58	25–73	25–58
0	0.19	0.39	0.65	0.20	0.26	0.46
10	−.15	.10	.39	.25	.29	.54
20	−1.25	−.82	−.48	.43	.34	.77
30	−1.20	−.65	−.10	.55	.55	1.10
40	.00	1.16	2.11	1.16	.95	2.11

(b) Sagittal fan

β degrees	$\overline{\Delta R_s}$ for filter No.			$\overline{\Delta(\Delta R_s)}$ for filter pairs Nos.		
	58	73	25	73–58	25–73	25–58
0	−0.01	0.21	0.44	0.22	0.23	0.45
10	.18	.41	.66	.23	.25	.48
20	.15	.43	.70	.28	.27	.55
30	.14	.44	.77	.30	.33	.63
40	.81	1.17	1.62	.36	.45	.81

TABLE 6. *Values of* $\overline{\Delta R_T}$, $\overline{\Delta R_s}$, $\overline{\Delta(\Delta R_T)}$, *and* $\overline{\Delta(\Delta R_s)}$ *for the tangential and sagittal fans as a function of* β.

The values are for lens No. 3, a wide-angle lens having a foeal length of 153.3 mm. Part (a) shows the values of $\overline{\Delta R_T}$ and $\overline{\Delta(\Delta R_T)}$ for the tangential fan and part (b) shows the eorresponding values for the sagittal fan. Values are given for three eonditions of target illumination determined by the indieated filter Nos. All values are expressed in mm.

(a) Tangential fan

β degrees	$\overline{\Delta R_T}$ for filter No.			$\overline{\Delta(\Delta R_T)}$ for filter pairs Nos.		
	58	73	25	73–58	25–73	25–58
0	0.05	0.31	0.67	0.26	0.36	0.62
10	−.07	.22	.56	.29	.34	.63
20	−.33	−.08	.49	.25	.57	.82
30	−.10	.44	1.12	.54	.68	1.22
40	.99	2.04	3.06	1.05	1.02	2.07

(b) Sagittal fan

β degrees	$\overline{\Delta R_s}$ for filter No.			$\overline{\Delta(\Delta R_s)}$ for filter pairs Nos.		
	58	73	25	73–58	25–73	25–58
0	−0.35	−0.03	0.23	0.32	0.26	0.58
10	.45	.67	.96	.22	.29	.51
20	.02	.41	.65	.39	.24	.63
30	1.10	1.40	1.81	.30	.41	.71
40	1.30	1.70	2.14	.40	.44	.84

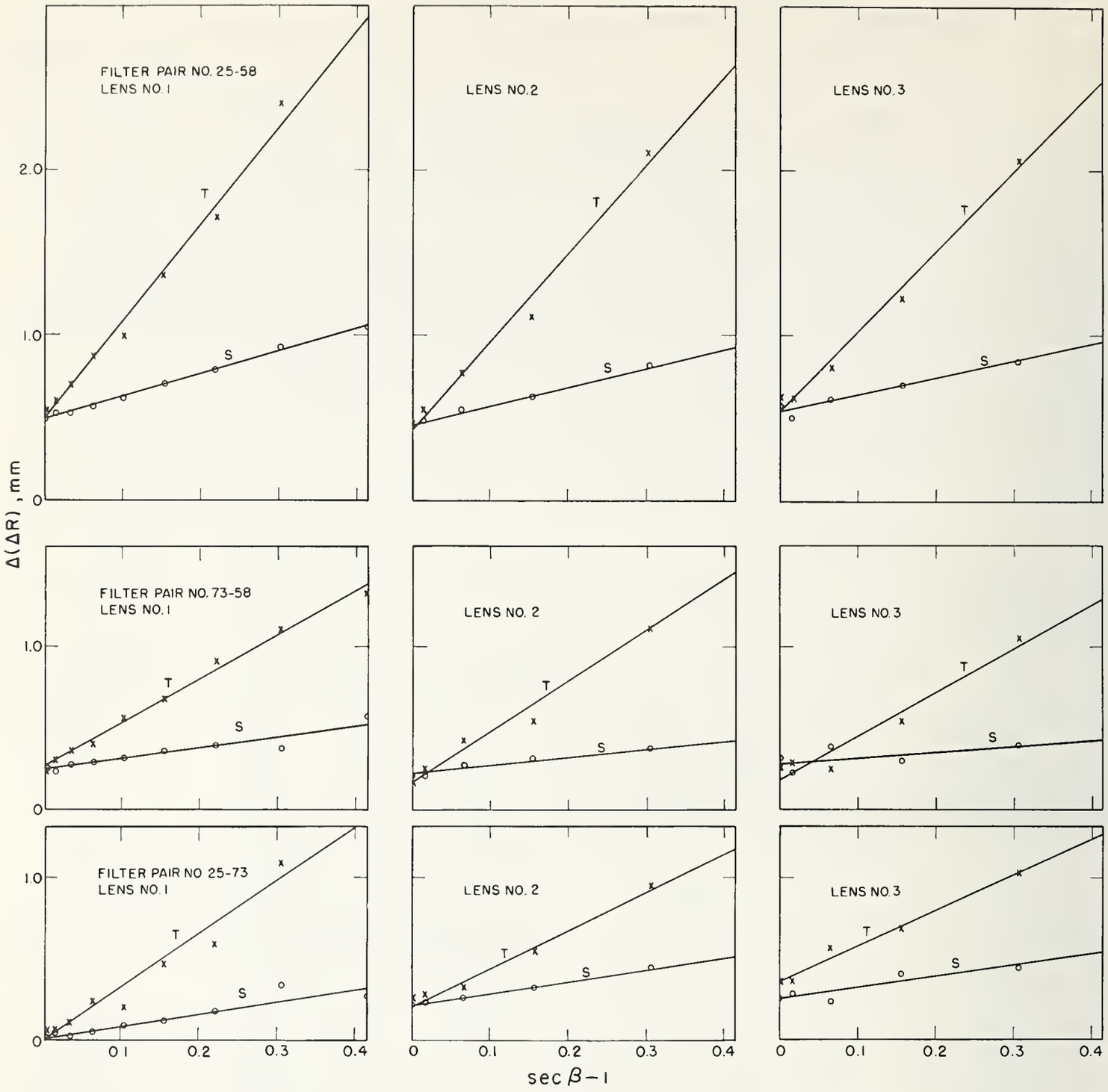


FIGURE 8. $\overline{\Delta(\Delta R)}$ *versus* (sec β–1) *curves for three lenses for three filter pairs.*

Observed values of $\overline{\Delta(\Delta R_T)}$ are plotted as X's in each frame; the straight line, marked *T*, is the regression line of $\overline{\Delta(\Delta R_T)}$ on (sec β–1) computed by least-squares. Observed values of $\overline{\Delta(\Delta R_s)}$ are shown as circles; the straight line, marked *S*, is the regression line of $\overline{\Delta(\Delta R_s)}$ on (sec β–1) computed by least squares. The plotted values are taken from table 5.

mitted by each filter for each value of β are also given. The values of $\overline{\Delta(\Delta R_T)}$ and $\overline{\Delta(\Delta R_s)}$ for each filter pair are derived from these values of $\overline{\Delta R_T}$ and $\overline{\Delta R_s}$ with the aid of eqs (7) and (8).

When the values of $\overline{\Delta(\Delta R_T)}$ and $\overline{\Delta(\Delta R_s)}$ are plotted as a function of (sec β—1) as shown in figure 8, it found that the variation of these quantities with (sec β—1) is close to linear. Accordingly one may write an empirical formula connecting $\overline{\Delta(\Delta R_T)}$ with (sec β—1) as follows:

$$\overline{\Delta(\Delta R_T)}=a+b(\sec\beta-1) \qquad (9)$$

which is the equation of a straight line with a as the y-intercept and b as the slope of the line. A similar empirical formula of the same nature may be written connecting $\overline{\Delta(\Delta R_s)}$ and (sec β—1). The constants a and b may be determined graphically by drawing the best fitting straight line or analytically by

TABLE 7. *Values of the constants* a *and* b *for fitting an equation of the form* $\overline{\Delta(\Delta R)}=a+b$ (sec $\beta-1$) *to the observed values of* $\overline{\Delta(R_T)}$ *and* $\overline{\Delta(\Delta R_s)}$ *as a function of* (sec $\beta-1$).

Values of a and b are given for three filter pairs for both $\overline{\Delta(\Delta R_T)}$ and $\overline{\Delta(\Delta R_s)}$ for each of three lenses. All values are expressed in mm.

Lens No.	Filter pairs Nos.	Values of a and b for use in equation of form $\overline{\Delta(\Delta R_T)}=a+b$ (sec $\beta-1$) for			
		$\overline{\Delta(\Delta R_T)}$		$\overline{\Delta(\Delta R_s)}$	
		a	b	a	b
1	73–58	0.27	2.67	0.26	0.64
	25–73	.22	3.21	.23	.73
	25–58	.49	5.88	.49	1.37
2	73–58	0.19	3.02	0.23	0.47
	25–73	.23	2.27	.23	.71
	25–58	.43	5.28	.46	1.17
3	73–58	0.19	2.63	0.29	0.33
	25–73	.36	2.18	.26	.67
	25–58	.55	4.82	.55	1.01

applying the method of least squares. It is evident from the curves shown in figure 8 that the constant a is approximately equal to the value of $\overline{\Delta(\Delta R_T)}$ or $\overline{\Delta(\Delta R_s)}$ at $\beta=0°$. It is also evident that the increase in $\overline{\Delta(\Delta R_T)}$ with (sec $\beta-1$) is substantially greater than that of $\overline{\Delta(\Delta R_s)}$; accordingly the slope b is greater for the tangential fan than for the sagittal fan.

Determinations of the constants a and b that are usable in equations of the form given in eq (9) for showing the relation of values of $\overline{\Delta(\Delta R_T)}$ or $\overline{\Delta(\Delta R_s)}$ to (sec $\beta-1$) were made for each of the filter pairs for each of the three lenses using the values given in tables 4, 5, and 6. The values of a and b for each combination are given in table 7. The straight line shown in each of the frames in figure 8 is drawn to satisfy the equation of the line fixed by the least squares determination of values of a and b for the particular filter pair. It is clear from the manner in which the points in any one of the graphs are distributed with respect to the straight line in that graph that the assumption of a linear relationship between values of $\overline{\Delta(\Delta R_T)}$ and (sec $\beta-1$) and between values of $\overline{\Delta(\Delta R_s)}$ and (sec $\beta-1$) is a valid one.

For a given filter pair, it is clear that the value of the slope b found for the $\overline{\Delta(\Delta R_T)}$ data is not the same as that for the $\overline{\Delta(\Delta R_s)}$ data. The value of the slope b for the $\overline{\Delta(\Delta R_T)}$ data is in all instances greater by a ratio ranging from 3.2 to 7.9 times greater than that for the $\overline{\Delta(\Delta R_s)}$ data. In view of the relative magnitude of this ratio, it is believed that it may be regarded as established that the variation of $\overline{\Delta(\Delta R_T)}$ with (see $\beta-1$) is significantly greater than that of $\overline{\Delta(\Delta R_s)}$ for the three lenses.

5. Discussion

In this study, a visual method of determining longitudinal spherical and chromatic aberration from measurements of angular deviation in the collimated beam emergent from a lens has been extended to the extra-axial region. Results are reported for measurements made on three lenses at five angular inclinations to the optical axis. It is of especial interest to note the increase in the values of longitudinal spherical aberration with increasing β and to note the large magnitudes attained. At $\beta=40°$, a range of 3 mm in the measured values of the longitudinal spherical aberration is not unusual; of course if these values were referred to the normal to the focal plane they would be reduced by a cosine factor but would still remain quite large.

The authors express their appreciation to other members of the staff of the National Bureau of Standards for assistance during this work and in particular to Edgar C. Watts who prepared the illustrations.

(Paper 66C3–95)

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 66C, No. 4, October–December 1962

Biprism Method of Determining the Equivalent Focal Length of Flat Field Lenses

Walter R. Darling

(June 25, 1962)

A device is described that permits the rapid determination of the equivalent focal length of a lens. A transmitting biprism mounted between a collimated light source and a lens, divides the light incident upon the front of the lens into two parallel beams making a fixed angle with one another. On passing through the lens, two images are formed in the focal plane. The magnitude of the lateral separation of the images is determined by the angular separation of the two incident beams and the focal length of the lens. The focal length of the imaging lens may be determined from the measured separation of the images at the focal plane of the lens and the known angle of deviation of the two incident beams produced by the biprism.

1. Introduction

The National Bureau of Standards is frequently called upon to determine the equivalent focal length and related constants of lenses with a high degree of accuracy. For many types of lenses these measurements are customarily performed visually on the precision optical bench to within ± 0.10 mm on lenses whose equivalent focal lengths range up to 200 mm, ± 0.15 mm for lenses whose equivalent focal lengths range between 200 and 800 mm. ± 0.25 mm for lenses whose equivalent focal lengths range from 800 to 1,200 mm, and ± 0.50 mm for lenses whose equivalent focal lengths range from 1,200 to 1,800 mm.[1]

Frequently it is desirable to determine only the equivalent focal length of lenses for use in evaluation of certain metrical qualities, such as distortion and resolution at finite and infinite distances. For these determinations the values of the focal lengths need not be as accurately known as those obtained by the precision optical bench method, yet it is desirable to know these values to a greater degree of accuracy than the value marked on the front of the lens, which is referred to as the nominal focal length. The tolerance for the nominal focal length is given by the ASA Standards [2] as ± 4 percent.

There is small likelihood of gross error in optical bench methods, but to preclude the possibility of such error it is desirable to have a method that permits a quick check of the value of the equivalent focal length of a given lens obtained with the optical bench. It is moreover desirable that the check measurement be based upon a different principle of measurement and employing different data to minimize the danger of systematic error. In this paper a simple check method for measuring the equivalent focal length of a lens is presented. A biprism placed in front of the lens under test splits the incoming collimated beam into two parallel beams with a fixed angular separation. With this prism, values of focal length accurate to within ± 1 percent may be obtained. While the range of uncertainty is somewhat higher than that attained by the optical bench method it nonetheless permits one to evaluate the equivalent focal length of a given lens with a degree of precision well within the tolerances set forth in the ASA Standards for nominal focal length.

[1] The differences in accuracy depend more upon the relative apertures than upon focal lengths, as a consequence of the differences in depth of focus.
[2] ASA Standards—Focal Length Marking of Lenses. PH.3.13-1958.

2. Description of the Instrument

An optical wedge of borosilicate optical glass approximately 2¾ in. in diameter was ground and polished with an angle of approximately 2 degrees between the faces. The finished wedge was then cut on a diamond saw in the principal section across its diameter. One half of the wedge was inverted with respect to the remaining half so that the deviations were in opposite directions. These two halves were then cemented to a flat base of ground and polished optical glass of approximately 3⅛ in. in diameter and ½ in. in thickness. For small wedge angles α, the deviation angle θ for normal incidence is:

$$\theta=(n-1)\alpha, \tag{1}$$

n being the index of refraction of the material used for the biprism. For this borosilicate glass $(n-1)=0.519$. For a collimated beam of light incident normally upon one side of the biprism, the angle separating the two emergent beams is 2θ. The magnitude of the angular deviation, 2θ, was measured in the refractometry laboratory with white light and was found to be 2.059 degrees. The probable error of this measurement is approximately 6 sec. The rest of the instrument consists of a cell to hold the biprism and a rod attached to the cell for adjustment of the entire unit. The component parts of the instrument are illustrated in figure 1.

The wedge is not achromatic but the error resulting from the lack of achromatism is negligible. When white light passes through a prism (fig. 2), the components of different wavelengths are deviated by

FIGURE 1. *The component parts of the instrument are illustrated.*

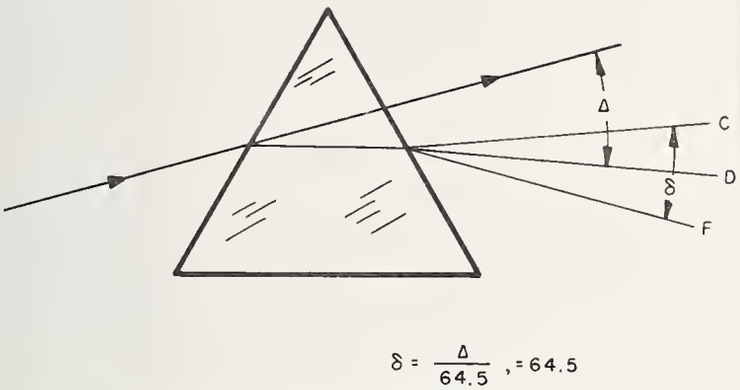


FIGURE 2. *When a white light beam passes through the wedge, the difference in the deviation of the red beam and the blue beam forms an angle* δ.

different amounts. The magnitude of the dispersion is customarily measured by the difference between the deviations for red light, hydrogen 656, and blue light, hydrogen 486. Let this difference in deviation be called δ. It may easily be deduced from eq (1) that

$$\delta=\frac{\theta}{\nu}$$

where ν is the Abbe number of the glass. Our biprism is made of borosilicate glass, $\nu=64.5$. Consequently

$$\delta=0.015\theta.$$

The wavelength of maximum visibility, for which visual settings are made when white light is used, falls almost midway between the wavelengths F and C. Since light at F and at C is definitely colored and also considerably less bright than that of maximum visibility, it is very unlikely that a setting would be off as far as ½ δ. Further, the deviation of the biprism consists of deviation to the left of θ plus deviation to the right of θ. Consequently the dispersion is also half in one image and half in the other.

The error in setting, on one of the images, resulting from chromatic aberration should not exceed $\frac{1}{4}\frac{\theta}{\nu}$, where 2θ is the total deviation of both prisms, and the error in the measured separation of the two images from this source then should not exceed $\frac{1.4x}{(4)(64.5)}$, where x is the separation of the images, or approximately one quarter of 1 percent of the measured interval. This represents an upper limit of error that should never be exceeded, not the probable error, which would be one fourth as large.

3. Method of Measurement

The lens under test is placed in the chuck of the nodal slide assembly of the optical bench. The nodal slide assembly is moved along the ways of the optical bench until the image of a collimated source formed by the lens under test is viewed in the microscope of the optical bench. The reticle used in the collimator consists of a vertical and horizontal line at right angles to each other forming a cross. The nodal slide assembly is now adjusted by the lead screw of the optical bench until the image formed by the lens of the collimated beam proceeding from the illuminated reticle is in sharp focus. The biprism is now inserted close to the face of the lens under test between the lens and the collimated beam from the illuminated reticle. This is illustrated in figure 3. The image of the reticle formed by the lens under test is split by the biprism into two images which are viewed in the ocular of the viewing microscope. Figure 4 is a photograph of the two split images as seen through the microscope. Care must be exercised in the orientation of the biprism to achieve maximum displacement of the split images. This orientation is produced by rotating the biprism within its cell about its optical axis until the two horizontal lines of the viewed images, if continued, would appear as one continuous line. The horizontal crosshair in the viewing microscope is used as a guide in making this adjustment. The biprism has no spherical refracting power so the position of best focus is not changed by the introduction of the biprism. The viewing microscope is equipped with a lateral adjustment and scale that allows measurements to be made to within ± 1 μ. The vertical arm of the crosshair in the microscope is brought into coincidence with the vertical line in the image of one of the two images formed by the biprism and lens by the lateral adjustment of the viewing microscope. A reading R_1 is taken of this position on the lateral scale of the microscope. The microscope is now traversed laterally until the vertical line in the crosshair of the microscope is in coincidence with the vertical line in the second image of the cross and a reading R_2 is taken on the lateral scale of the viewing microscope. Readings R_1 and R_2 are successively taken until an average of at least five readings of each position has been obtained. The focal length can then be determined from the following relation:

$$D=R_{1a}-R_{2a}=2f\tan\theta$$

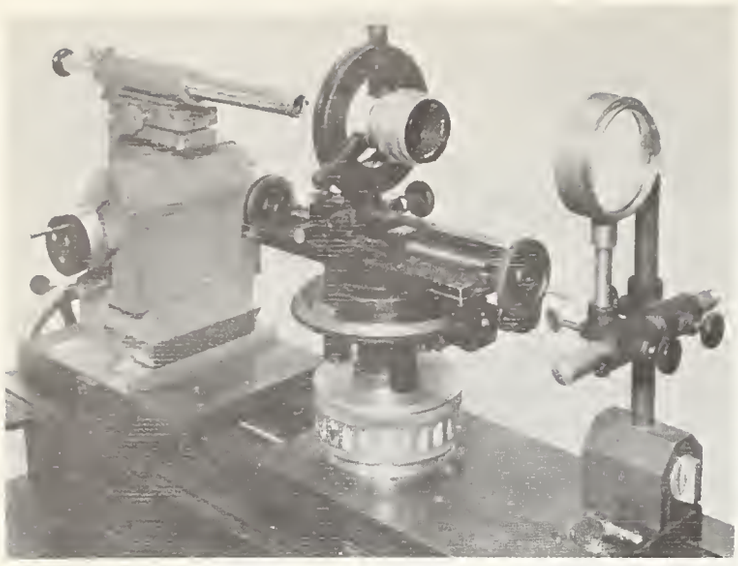


FIGURE 3. *The briprism is inserted close to the face of the lens under test, between the lens and the collimated beam from the illuminoted reticle.*

FIGURE 4. *A photograph of the two split images as seen through the microscope.*

or

$$f=0.5\ D\ \cot\ \theta$$

where $D=R_{1a}-R_{2a}$ or the difference of the averages of the lateral scale measurements of R_1 and R_2. The quantity 0.5 cot θ is a constant of the instrument. Hence the value of the equivalent focal length may be obtained directly from the formula,

$$f=27.81\ D. \qquad (2)$$

4. Precision Optical Bench

The operation of the precision optical bench has been discussed in a previous paper.[3] It may be mentioned here that the lens under test by the optical bench method of measurement is subjected to a number of critical time-consuming adjustments before any measurements are made. Depending upon the physical size of the lens and the focal length as much as two hours time is needed for adjustment of the lens on the precision optical bench.

5. Results of Measurement

The following table contains the results of measurements made on a number of lenses of varying focal lengths.

TABLE 1

Lens	Nominal	Values of equivalent focal length		Percent difference
		Measured		
		by optical bench method	by using biprism	
	f	f_0	f_b	$100\ \frac{f_b-f_0}{f_0}$
	mm	*mm*	*mm*	
1	58	58.39	58.38	−0.02
2	58	58.36	58.37	+.02
3	58	58.50	58.79	+.50
4	58	58.26	58.22	−.07
5	58	58.14	58.09	−.08
6	58	58.55	58.42	−.22
7	58	58.22	58.22	+.00
8	150	149.94	150.29	+.23
9	180	180.50	179.38	−.62
10	306	305.83	306.52	+.22
11	483	486.58	486.67	+.02
12	610	610.67	610.59	−.01
13	763	763.09	763.44	+.04
14	1065	1065.55	1064.99	−.05

Average deviation ±0.15%

Improper orientation of the prism may lead to error, but even this is unlikely to lead to serious error as long as the prism axis does not deviate from normality with the ways of the measuring microscope by an amount exceeding ±2.5 degrees. A misorientation of this magnitude is not likely to occur as the splitting of the horizontal arm of the cross would immediately make manifest this type of error.

6. Limitations in Measurements

The maximum travel of the viewing microscope available for the measurement of R^1 and R^2 was 50 mm. With the biprism whose angle of deviation being 2.059 degrees, the maximum equivalent focal length that can be determined is 1390 mm. Longer focal lengths can be determined by decreasing the angle of deviation in the biprism or by increasing the lateral travel of the viewing microscope. Care must

[3] F. O. Washer, W. R. Darling, Factors affecting the accuracy of distortion measurements made on the nodal slide optical bench, J. Opt. Soc. Am. **49**, 517 (1959).

be exercised in using a prism of this nature, one must use a reasonably monochromatic light source, and the initial calibration of the prism must be made with near monochromatic light of approximately the same wavelength as that which will be used in the focal length determination.

Upon examination of the results of measurements obtained in the experimental data the focal length determinations by the biprism method were found to be biased. The values reported in table 1 using the biprism method were adjusted in the amount of −0.07 percent. A 0.07 percent change in the focal length amounts to approximately 5 sec change in deviation of the biprism, this is well within the tolerance of the measurement of the biprism. With the deviation from the average not greater than ±0.15 percent one can be confident that a focal length determination by means of the biprism is very unlikely to be error by as much as 1 percent.

The focal length determinations in table 1 were made with a Wratten No. 73 filter with an effective wavelength of approximately 575 mμ.

7. Sources of Error

Since $f=0.5D \cot \theta$, error can be produced by errors in D and by errors in θ. Assuming θ to be without error, from the relation in eq (1) $0.5D \cot \theta$ is a constant equal to 27.81. Therefore the error in the equivalent focal length as a result of error in the lateral scale readings is given by the expression: $\Delta f=27.81\ \Delta D$, eq (2).

From the above it is clear that a 2μ error in D leads to an error of ±0.055 mm in f which can usually be considered negligible. Errors in f arising from errors in the measured value of θ can be neglected as it is unlikely to exceed ±0.2 percent for a probable error of 6 sec in an angle of 2.059 degrees.

8. Discussion

It is evident from the information contained in the foregoing sections that the biprism method is a satisfactory means for quickly determining the equivalent focal length of flat field lenses with an error that does not exceed ±1 percent. It is also pointed out that a nodal slide assembly is not a necessity in making the focal length determination. Any method that allows the lens to be held firmly and in a reasonable alinement with the collimated incident beam is sufficient.

Acknowledgment is made to F. E. Washer and R. E. Stephens for assistance and suggestions in the preparation of this paper.

(Paper 66C4–106)

Calibration of Photogrammetric Lenses and Cameras at the National Bureau of Standards

Francis E. Washer.
National Bureau of Standards
Washington, D. C.

Abstract: *A summary of calibrations performed at the National Bureau of Standards on lenses and cameras that are used in precise photogrammetric work is given. Brief description of the photographic and visual calibrations most frequently required are given. This paper includes a list of publications by members of the NBS staff that pertain to problems of lens and camera calibration.*

1.0 Introduction

The success of a given mapping program depends in large measure upon the quality of the lenses used in the various operations intervening between the initial photography from the air and the production of a usable map. This was early recognized and various tests were developed to measure the optical constants and imaging properties of lenses. Emphasis was placed initially upon the characteristics of the lens in the aerial camera and many types of tests and calibrations were developed which proved useful. Subsequently, the performance of lenses used in other phases of the operation was investigated in greater detail with the aim of achieving the minimum loss in information content of the original photography.

A great deal of the early work in the measurement of lens performance was done at the National Bureau of Standards at the request of various mapping agencies. In the process of making these tests, information was gathered which when correlated with the quantitative and qualitative properties of the aerial photograph formed the basis for numerous lens and camera specifications.

At present, lenses that are to be used in the government mapping projects are required to satisfy minimum standards of performance as determined by the National Bureau of Standards or an accredited testing agency.

The present article presents a brief description of a number of the more commonly required calibrations together with definitions of the quantities measured.

2.0 Calibration of Aerial Camera Lenses

The nature of the calibration of a given lens is determined by the requirements of the user. For some lenses, an accurate determination of the equivalent focal-length suffices. For lenses used in precision mapping cameras, accurate information on all the metrical properties of the lens that have a bearing upon the accuracy of the first map is required. These requirements are contained in specifications that list the specific requirements and set limits on the range in magnitude of the various quantities to be measured that will be acceptable for the given use.

2.1 calibrations using photographic methods

It is customary to perform the calibrations on most lenses that are to be used in aerial photography by photographic methods. For this work the Bureau employs the precision lens testing camera (1) and the camera calibrator (31).

2.1.1 *Calibrations on lenses not mounted in cameras*

The lens-testing camera is used primarily for determination of such quantities as equivalent focal-length, *EFL;* back focal-distance, *BF;* distortion, *D;* and resolving-power, *RP.* This instrument consists of 10 collimators, spaced at 5° intervals which provide infinitely distant objects within the confines of the laboratory. Each collimator is equipped with a resolution chart in addition

to the cross-lines in the reticle that are used in the precise determination of the angle separating the collimators. With this instrument, the entire region of usable imagery of the lens under test is sampled at intervals photographically to determine the plane of best average definition. When this plane has been located by examination of the imagery which is recorded on the test negative, the distance r_β separating the image at known angle β from the axial image is measured. The value of EFL is then obtained using the relation

$$EFL = r_\beta \cot \beta \tag{1}$$

The distance separating the rear vertex of the lens from the position occupied by the emulsion surface of the photographic plate when in the plane of best average definition is measured with the aid of a viewing microscope and a standard scale mounted on the bench ways; this is the back focal-distance, BF.

The measured value of BF is of prime interest to the camera maker in that it permits the camera maker to position the camera lens with respect to the fixed focal-plane with a high degree of confidence that optimum performance will be achieved provided that the focal-plane is placed at a distance from the rear vertex of the lens corresponding to the value of the back focal distance, BF, for best average definition.

The measured value of the equivalent focal length, EFL, is valid as a true scale-factor for use in interpreting transverse distances in the central region of the negative. For the outer regions of the negative, the relation

$$f = EFL = r_\beta \cot \beta \tag{2}$$

yields varying values of EFL. It is customary to attribute this value to distortion, D_β, defined by the equation

$$D_\beta = r_\beta - f \tan \beta \tag{3}$$

where EFL is written f for brevity. Values of D_β are determined from the measurements of r_β made on the original test negative. It frequently happens that D_β so defined may reach a high value at large values of β and produce intolerable error for the given scale-factor because of wide variation in the measured magnitude of equal areas located in different regions of the negative. This may be minimized by selecting a value of f which when used in Equation 3 equalizes the magnitude of the maximum positive and the maximum negative distortion. This particular value of f is designated f_c or calibrated focal-length, CFL.

Values of the resolving power are read from the images registered on the negative in the plane of best average definition. At this Bureau, high-contrast line charts are used and values of resolving power for tangential and radial are reported at each angular interval. Values are given in lines-per-millimeter and indicate that the given pattern and all coarser patterns are clearly resolved. The values of resolving power vary with contrast of target and emulsion used to register the image. Knowledge of the measured value of the resolving power is of interest to the lens user in that it gives an indication of the quality of definition that can be obtained with the lens. High values indicate good definition. The foregoing measurements which are made on the precision-lens testing camera are made on lenses not mounted in cameras. When the values obtained indicate that the lens is suitable for use in a precision-mapping camera, the lens is mounted in its camera and returned for calibration on the camera calibrator.

2.1.2 *Calibrations on lenses mounted in precision-type cameras*

The camera calibrator is equipped with 25 collimators consisting of four arrays of collimators spaced at 7.5° intervals. The four arrays make successive angles of 90° to form a cross-shaped arrangement. The central collimator is common to all. This device permits the valuation of the distortion at 7.5° intervals for four separate radii of the picture area. In addition, the device is used for locating the principal-point of collimation and the relative positions of the four collimation-markers mounted for registry in the focal plane. A schematic drawing of a typical test negative is shown in Figure 1. The location of the collimation index markers, designated A, B, C, D are registered on the photographic plate at the same time as the images of the collimator targets along the radii, designated I, II, III, and IV. While a single negative contains all of the information necessary for the calibration of a precision-camera, two are made with the camera rotated 180° in azimuth between exposures and the measured results averaged to insure accuracy.

With this device, the equivalent and calibrated focal-lengths of the lens are determined for the actual focal-plane of the camera. In addition, values of the radial distortion are determined for four radii spaced at azimuths 0°, 90°, 180°, and 270°.

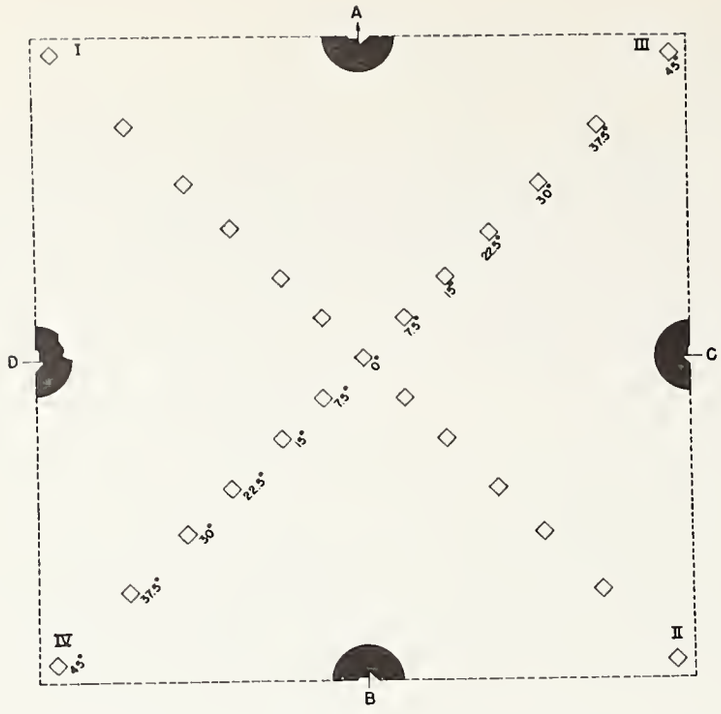


FIG. 1. Schematic drawing of test negative obtained with camera calibrator.

Variations in the value of D_β for a given value of β with azimuth indicate the presence of radial asymmetric-distortion.

To illustrate the effect of radial asymmetric-distortion measured values of the distortion D_β for β ranging from 0° to 45° for four azimuths are shown for two lenses in Table 1. In this table, it is clear that for lens No. 1, the values of D_β for a given value of β show relatively small variation with ϕ, the azimuthal angle: moreover the average departure $\overline{\Delta D_\beta}$ from the average $\overline{D_\beta}$ is not appreciably greater than the probable error of measurement. On the other hand, the values of D_β for a given value of β for lens No. 2 show marked variation with ϕ. In addition, the average departure $\overline{\Delta D_\beta}$ from $\overline{D_\beta}$ is in two instances greater than $\overline{D_\beta}$; and $\overline{\Delta D_\beta}$ is also much greater than the probable error of measurement.

If the target images for radii 0° and 180° do not fall in a straight line, tangential-distortion is indicated. Also the failure of images for radii 90° and 270° to fall in the same straight line indicates tangential-distortion. The use of the two negatives taken at azimuth 0° and 180° of camera rotation about its axis increases the accuracy of determination of tangential-distortion.

The function of the collimation markers is to locate the position of principal-point of autocollimation on each negative produced by the camera in use. Consequently extreme care must be taken in the design and installation of such markers to ensure their precise positioning and to permit precise measurement of their separations and locations relative to each other. Their configuration simplifies the orientation of successive negatives relative to each other; indicates direction of flight; and the measured separation of their images on the film permits a comparison with the separations listed in the calibration report for use in checks on differential film shrinkage. Close tolerances are usually set on the location of these markers both for the angle between the lines joining opposite pairs of markers and the position of the intersection point (center of collimation) of these lines and the principal-point of autocollimation. Because of these close tolerances, it is usually necessary to adjust the location of the markers following preliminary calibration in such a manner that the prescribed tolerances are satisfied. Following such adjustment, the calibration is repeated to determine whether or not the requirements have been met.

The platen is checked for flatness in the course of calibration. Tolerances for flatness are usually set as low as feasible because departures from flatness introduce appreciable amounts of distortion. In addition, measurements of the departures from paral-

Table 1. Variation of the Distortion D_β Referred to the Calibrated Focal-length with Angular Separation β from the Axis and Azimuthal Angle ϕ About the Axis

These values are shown for wide-angle lens No. 1, which has low asymmetric-distortion (part a), and lens No. 2 which has appreciable asymmetric-distortion (part b). Values of the average $\overline{D_\beta}$ and the average departure from the average $\overline{\Delta D_\beta}$ are also given. All values are expressed in microns.

(a) *Lens No. 1*
D_β *at* β =

β / *Azimuth*	0°	7.5°	15°	22.5°	30°	37.5°	45°
degrees							
0	0	2	4	−2	− 9	3	−11
90	0	2	1	−9	− 9	3	− 8
180	0	2	−1	−5	− 6	6	− 6
270	0	2	2	−8	− 7	6	− 1
$\overline{D_\beta}$	0	2	2	−6	− 8	5	− 6
$\overline{\Delta D_\beta}$	0	0	±1	±2	± 1	± 2	± 3

(b) *Lens No. 2*
D_β *at* β =

0	0	−1	−1	−2	−15	−12	−22
90	0	−2	−5	0	1	16	28
180	0	−1	−5	−4	− 6	9	27
270	0	0	2	2	− 3	0	−10
$\overline{D_\beta}$	0	−1	−2	−1	− 6	3	6
$\overline{\Delta D_\beta}$	0	0	±3	±2	± 5	± 9	±23

lelism of surfaces of the filter mounted on the front of the lens are made. This is done because a filter whose surfaces are appreciably out of parallel behaves like a thin prism and introduces both radial, asymmetric, and tangential-distortion into the final negative.

It must be mentioned that the foregoing calibrations are intended for precision-type mapping cameras, which do not have the collimation index markers and focal-plane located in detachable magazines. In addition, the calibration negatives are made without activating the shutter mechanism. Consequently, it is necessary that cameras submitted for this type of calibration should be in such condition that the shutter can be maintained in the open position during test.

3.0 Calibration of Lenses Used at Finite Distances

In general, the lenses used at finite distances include such a wide range of sizes and focal-lengths that it is impracticable to use photographic methods because of space limitations. It is therefore customary to use a visual optical bench for such lenses.

3.1 Calibrations using the visual optical bench

Measurements commonly made on the visual optical bench (21) include focal-lengths; back and front focal-distances; lens thickness; nodal-point separation; and distortion.

The optical bench, presently in use at the National Bureau of Standards, consists of a set of bench ways upon which are mounted slides carrying a viewing microscope and nodal-slide assembly. A collimated beam of light is provided by a parabolic mirror mounted at one end of the bench ways. Length measurements are made with respect to a standard scale mounted beside the bench ways or in the case of distortion with respect to a transverse scale that indicates sideways displacement of the viewing microscope. The nodal slide assembly upon which the lens under test is mounted is provided with a variety of adjustments for positioning the rear nodal-point of the lens in the vertical axis of rotation. In addition, provision is made for measured angular displacements of the optical-axis of the lens with

respect to the axis of the collimated beam from the mirror.

To measure the equivalent focal-length of a lens, the lens is so positioned that its rear nodal-point is located in the vertical axis of rotation of the nodal-slide and the entire assembly moved along the bench ways until the illuminated target located in the focal-plane of the lens is imaged by the lens in the object-plane of the viewing microscope. The distance separating the image formed by the lens and its rear nodal-point is the equivalent focal-length, *EFL*, of the lens. The distance separating the image and the vertex of the back surface of the lens is the back focal-distance, *BF*.

A simple focal-length determination usually suffices where the user is concerned only with the scale-factor in the image for very distant objects and is not concerned with distortion.

For lenses used at finite distances, the front equivalent focal-length, front focal-distance (*FF*) nodal-point separation (*NPS*), and lens-thickness (*T*) are also measured. These front and back equivalent focal-lengths can usually be regarded as equal. These quantities are related as follows:

$$2\,EFL + NPS = BF + FF + T \tag{4}$$

It is necessary to have reliable values of the foregoing quantities for proper positioning of object-plane, lens, and image-plane in a camera used at finite distances at specified values of the reduction ratio $(1/M)$ where M is the ratio of object-to-image size, shown in Figure 2. In such instances, the object-distance d_o and image-distance d_i are given by the following formula:

$$d_o = f(M + 1) \tag{5}$$

$$d_i = f\left(1 + \frac{1}{M}\right) \tag{6}$$

and the distance D separating object and image planes is

$$D = d_o + d_i + NPS = \frac{f}{M}(M + 1)^2 + NPS \tag{7}$$

If the lens is intended for use in highly precise copying or projection work, the values of the distortion in the image-plane at one or more specified values of the reduction ratio are also measured.

Values of the resolving power can also be determined with the visual optical bench. These values are useful in indicating the probable performance of the lens although they are likely to be appreciably higher than those determined by photographic means.

4.0 Miscellaneous Calibrations

In the foregoing sections, a brief account of calibrations that are frequently performed at the National Bureau of Standards is given. In addition, many calibrations are performed of a less routine nature. These calibrations include items such as determination of geometric *f*-number, *T*-number, lens transmit-

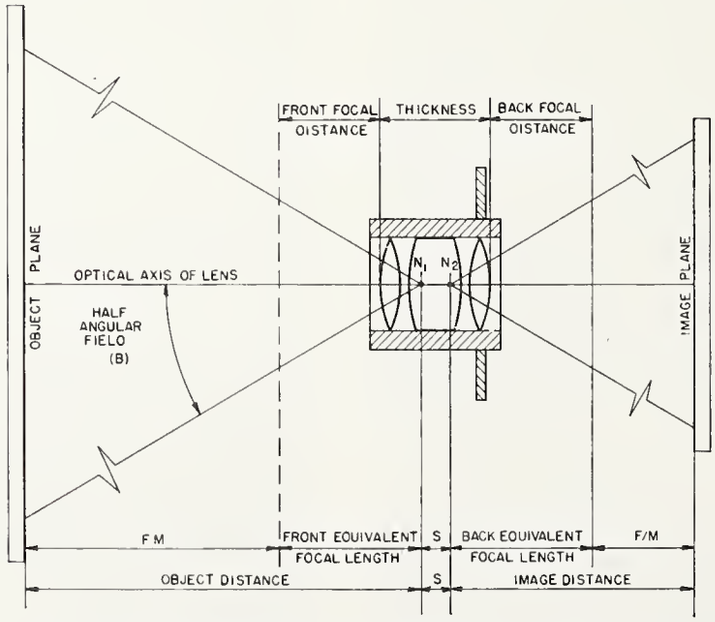


FIG. 2. Schematic drawing showing optical constants of photographic objective and the relation between object and image distances.

tance, relative illumination in the focal-plane, longitudinal spherical and chromatic-aberration, sine-wave response, and many tests of a specialized nature.

5.0 Fees for Calibration

The calibrations described in the foregoing sections are performed on a fee basis with the fee determined by the actual cost of labor and materials used in the work. Calibrations that are frequently performed are usually on the basis of a fixed fee which are listed in the Test Fee Schedule published in the *Federal Register*. The schedule presently in use was published Apirl 8, 1961 in Volume 26, Number 67 of the *Federal Register*. The fee for calibrations of a specialized nature are based upon actual labor costs and computed at the time of the test.

These calibrations are performed as promptly as possible following receipt of the material in condition for test; a specific request authorizing the test; and instructions describing the nature of the test required. It is usually possible to complete a given calibration within three weeks following recipt of the material. Further information on specific details may be obtained on request from the Refractometry Section, Metrology Division (2.2), National Bureau of Standards, Washington 25, D. C.

6.0 List of Publications by Members of Staff of the NBS that Pertain to the Problem of Lens and Camera Calibration

6.1 lens performance

1. Gardner, I. C. and Case, F. A., "Precision Camera for Testing Lenses," *J. Research NBS* 18, 449 (1937) RP984.
2. Washer, F. E., "Resolving Power and Distortion of Typical Airplane-camera Lenses," *J. Research NBS* 22, 729 (1939) RP1216.
3. Gardner, I. C., "A Test of Lens Resolution for the Photographer," *NBS Circular* C428 (1941). (Superseded by C533).
4. Washer, F. C., "Characteristics of Wide-angle Airplane-camera Lenses," *J. Research NBS* 29, 233 (1942) RP1498.
5. Washer, F. E., "Region of Usable Imagery in Airplane-camera Lenses," *J. Research NBS* 34, 175 (1945) RP1636.
6. Gardner, I. C., "Compensation of the Aperture Ratio Markings of a Photographic Lens for Absorption, Reflection, and Vignetting Losses," *J. Research NBS*, 643 (1947) RP1803.
7. Gardner, I. C. and Washer, F. E., "Lenses of Extremely Wide Angle for Airplane Mapping," *J. Research NBS* 40, 93 (1948) RP1858. *J. Opt. Soc. Am.* 38, 421 (1948).
8. Washer, F. E., "Sources of Error in and Calibration of the *f*-number of Photographic Lenses," *J. Soc. Motion Picture Engr.*, 51, 242 (1948); *J. Research NBS* 41, 301 (1948 RP1927.
9. Washer, F. E., and Rosberry, F. W., "New Resolving Power Test Chart," *J. Opt. Soc. Am.* 41, 597 (1951).
10. Hotchkiss, R. N., Washer, F. E., and Rosberry, F. W., "Spurious Resolution of Photographic Lenses," *J. Opt. Soc. Am.* 41, 600 1951).
11. Washer, F. E., and Gardner, I. C., "Method for Determining the Resolving Power of Photographic Lenses," *NBS Circular* 533 (1953).
12. Washer, F. E., "Testing of Photographic Lenses at NBS," Photogrammetric Engineering V, 37–53 (1954).
13. "Optical Image Evaluation Symposium Proceedings," NBS Circular No. 526 (1954), Washer, F. E., "Resolving Power of Airplane-camera Lenses," No. 15, 208–218 (1954), Shack, R. V., "A Proposed Approach to Image Evaluation," No. 20, 275–286 (1954), Gardner, I. C., "The Experimental Evaluation of Lens Performance," *International Archives of Photogrammetry* XI, 224 (1954).
14. Magill, A. A., "Variation in Distortion with Magnification," *NBS* 54, 135–142 (1955) RP2574.
15. Rosberry, F. W., "Effect of Object Frequency on Focal Position of Four Photographic Objectives," *J. Research NBS* 57, 17 (1956) RP2688.
16. Gardner, I. C. and Bennett, A. H. "A Modified Hartmann Test Based on Interference," *J. Opt. Soc. Am.* and *Rev. Sci. Instr.* 11, No. 4, 441 (1925). *Zeitschrift fur Instrumentenkunde* 47, 197 (1927) Translation.
17. Washer, F. E. and Scott, L. W. "Influence of the Atmosphere upon the Precision of Telescope Pointing," *J. Research NBS* 39, 297 (1947) RP1829.
18. Washer, F. E., "An Instrument for Measuring Longitudinal Spherical Aberration of Lenses," *J. Research NBS* 43, 137–144 (1949) RP2015
19. Washer, F. E., "Optical T-Bench Method of Measuring Longitudinal Spherical Aberration," *J. Research NBS* 61, No. 1, July (1958) RP2880.
20. Washer, F. E., Tayman, W. P., and Darling, W. R., "Evaluation of Lens Distortion by Visual and Photographic Methods," *J. Research NBS* 61, No. 6 (1958) RP2920.
21. Washer, F. E. and Darling, W. R., "Factors Affecting the Accuracy of Distortion Measurements Made on the Nodal Slide Optical Bench," *J. Opt. Soc. Am.* 49, No. 6, 517–534 June 1959.
22. Washer, F. E. and Darling, W. R., "Evaluation of Lens Distortion by the Inverse Nodal Slide," *J. Research NBS* 63C, No. 2 (1959).
23. Washer, F. E., and Darling, W. R., "Evaluation of Lens Distortion by the Modified Goniometric Method," *J. Research NBS* 63C No. 2.
24. Rosberry, F. W., "Equipment and Method for Photoelectric Determination of Image Contrast Suitable for Using Square Wave Targets," *J. Research NBS* 64C-1, (1960).
25. Washer, F. E. and Tayman, W. P., "Location of the Plane of Best Average Definition for Airplane Camera Lenses," Photogrammetric Engineering XXVI, No. 3, June (1960).
26. Washer, F. E. and Tayman, W. P., "Variation

of Resolving Power and Type of Test Pattern," *J. Research NBS* 64C-3, (1960).
27. Stephens, R. E., "Magnifications of a Telescope," *J. Opt. Soc. Am.* 51, No. 7, 803–804, July (1961).
28. Washer, F. E. and Tayman, W. P. "Location of the Plane of Best Average Definition with Low Contrast Resolution Patterns," *J. Research NBS* 65C No. 3 (1961).

6.2 CAMERA CALIBRATIONS

29. Washer, F. E., "Locating the Principal Point of Precision Airplane Mapping Cameras," *J. Research NBS* 27, 405 (1941) RP1428.
30. Washer, F. E. and Case, F. A., "New Precision Camera Calibrator," *Tech. News Bull.* 33, No. 1 (1949).
31. Washer, F. E. and Case, F. A., "Calibration of Precision Airplane Mapping Cameras," PHOTOGRAMMETRIC ENGINEERING XXVI, 619 (1950); *J. Research NBS* 45, 1–16 (1950) RP2108.
32. Washer, F. E., "Effect of Camera Tipping on the Location of the Principal Point," *J. Research NBS* 57, 31 (1956) RP2691.
33. Washer, F. E., Sources of Error in Various Methods of Airplane Camera Calibration," PHOTOGRAMMETRIC ENGINEERING XXII, 727 (1956).
34. Washer, F. E., "A Simplified Method of Locating the Point of Symmetry," PHOTOGRAMMETRIC ENGINEERING XXIII, 75 (1957).
35. Washer, F. E., "The Effect of Prism on the Location of the Principal Point," PHOTOGRAMMETRIC ENGINEERING XXIII, 520 (1957).
36. Washer, F. E., "Prism Effect, Camera Tipping, and Tangential Distortion," PHOTOGRAMMETRIC ENGINEERING XXIII, 721 (1957).
37. Washer, F. E., "Calibration of Airplane Cameras," PHOTOGRAMMETRIC ENGINEERING XXIII, 890 (1957).

Reprinted from
PHOTOGRAMMETRIC ENGINEERING
January
1963

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 67C, No. 4, October–December 1963

Determination of Optical Path Difference for a Photographic Objective

Francis E. Washer and Walter R. Darling

(June 6, 1963)

A method of measuring longitudinal spherical aberration and optical path differences by a direct visual means is described. The method employs a nodal slide optical bench and a movable-slit system. The underlying theory of the method is presented together with a brief description of the apparatus used. The results of measurement on a typical lens and a procedure for checking the consistency of measurements are included.

1. Introduction

One of the criteria used in evaluating the potential image forming qualities of a photographic objective is the magnitude of the longitudinal spherical aberration which measures the variation in axial focus with radius of annular zone of the lens. It is sometimes desirable however to use a related quantity more meaningful in terms of physical optics, namely the optical path difference existing between the paraxial region of the lens and the various annular zones located between the axial and marginal regions of the lens. The optical path difference expressed in units of wavelength yields information on the manner in which light converging from a given annular zone of the lens to a specified focal point may be expected to contribute to image formation. In specifying the performance of a high quality telescope objective, a common requirement is that the optical path difference over the entire wave front emergent from a lens shall not depart by more than 0.25 wavelength from a true sphere.

A method has therefore been developed for the measurement of the departure of the emergent wave front from a true spherical surface in units of wavelength. The method is essentially a refinement of the method reported by Chalmers [1] and similar to the two slit method used by Väisälä.[2] The present method is well adapted for use with lenses having a wide range of focal lengths.

The accuracy of the final results compares favorably with that of results obtained by more elaborate interferometric methods. Moreover the results are in such a form as to simplify the computation of changes in optical path difference with shift of the focal plane.

2. Theory of the Method

2.1. Determination of Optical Path Difference, P

When a lens is affected by longitudinal spherical aberration, the position of best axial focus changes from zone to zone of the lens. The magnitude of this focal shift can be readily measured by a variety of methods, one of which is reported in an earlier paper.[3] The effect of longitudinal spherical aberration may also be described as a distortion of the wave front from a true spherical form by action of the lens. This distortion of the wave front produces variation in the optical path separating different portions of the wave front emerging from the lens and the focal point. It is possible to measure this variation using a technique similar to that reported by Chalmers (see footnote 1). In this method, two parallel slits, separated by distance h_0, are placed in front of a lens, the image of an infinitely distant bright line object formed by the lens then appears as a series of parallel interference fringes in the focal plane of the lens. The central fringe is the brightest of the group, the others growing fainter with increasing order of interference.

The separation of these fringes is given by the relation

$$S=\frac{\lambda f}{h_0} \tag{1}$$

where S is the measured separation of the central fringe and an adjacent fringe, h_0 is the slit separation, f the equivalent focal length of the lens, and λ the wavelength of the incident light.

If the slits are so placed that one is on the axis of the lens, the second slit will be at distance h_0 from the axis. A difference in focus between the paraxial zone and the annular zone of radius h_0, will shift the image of the central fringe laterally by amount δ_1 and the change in optical path, ΔP_1, is given by the relation

$$\Delta P_1=\frac{\delta_1 h_0}{f}. \tag{2}$$

[1] S. D. Chalmers: Proceedings of the Optical Convention **2**, 156 (1912).
[2] Y. Väisälä, Ann. Univ. Fenn. Abo. **1** 3 (1922).
[3] F. E. Washer and W. R. Darling, Optical T-bench method of measuring longitudinal spherical aberration, J. Research NBS **61**, 31 (1958) RP2880.

On moving the slits outward along a radius of the lens in such a manner that the first slit occupies the position formerly occupied by the second slit, a new value δ_2 will be found and the change in optical path difference ΔP_2 is given by the relation $\Delta P_2 = \frac{\delta_2 h_0}{f}$. This process is continued until the last setting is made with the second slit near the margin of the lens, and the corresponding value of ΔP is $\Delta P_m = \frac{\delta_m h_0}{f}$. The total optical path difference P between the center and a given zone designated by zone number $m (m = 1, 2, 3, \ldots)$ is given by the expression

$$P = \Sigma \Delta P = \frac{h_0}{f} \Sigma \delta. \tag{3}$$

This may be expressed in units of wavelength by dividing by λ giving the optical path difference in wavelengths, n_λ, which is

$$n_\lambda = \frac{P}{\lambda} = \frac{h_0}{\lambda f} \Sigma \delta. \tag{4}$$

With the aid of eq (4), the variation in n_λ, which may be interpreted as phase shift, can be determined for any point of the wave front with respect to any other point of the wave front located along the same diameter and for the particular image plane in which the measurements are made.

2.2. Effect of Focal Shift

The relation, shown in eq (4), is valid for any single position of the focal plane. For a focal plane separated by distance Δf from the initial position, a different set of values of n_λ will be found. The measured values of δ change by amount $\Delta\delta$ which changes are related to Δf as shown in the following equation

$$\frac{\Delta \delta}{h} = \frac{\Delta f}{f} \tag{5}$$

where h is the zone height. The change Δn_λ in n_λ is given by the relation

$$\Delta n_\lambda = \frac{h_0}{\lambda f} \cdot \frac{\Delta f}{f} \Sigma h. \tag{6}$$

For the case of one slit located initially on the lens axis, $h = h_0 (m - \frac{1}{2})$ and Δn_λ is given by the relation

$$\Delta n_\lambda = \frac{h_0^2}{\lambda f^2} \Delta f \Sigma \left(m - \frac{1}{2} \right) \tag{7}$$

For a specified value of the zone number m, eq (7) may be written

$$\Delta n_\lambda = \frac{h_0^2}{\lambda f^2} \cdot \Delta f \cdot \frac{m^2}{2}. \tag{8}$$

For the case of the lens axis located initially midway of the two slits, the zone height h is given by the relation

$$h = m h_0. \tag{9}$$

The value of Δn_λ is given by

$$\Delta n_\lambda = \frac{h_0^2}{\lambda f^2} \cdot \Delta f \Sigma m \tag{10}$$

which for a specified value of m becomes

$$\Delta n_\lambda = \frac{h_0^2 \Delta f}{\lambda f^2} \cdot \frac{m(m+1)}{2}. \tag{11}$$

2.3. Determination of the Longitudinal Spherical Aberration

The same initial observations δ from which the distortion of the wave front in units of wavelength are determined may also be used to derive the values of longitudinal spherical aberration $\Delta f'$, expressed as a focal shift. The values of $\Delta f'$, δ, and h the zone height are related as follows:

$$\Delta f' = \frac{\delta f}{h}. \tag{12}$$

For the case of one slit on the lens axis at the start of observations, eq (8) becomes

$$\Delta f' = \frac{\delta f}{(m - \frac{1}{2}) h_0} \tag{13}$$

and for the case of the zero point of observations occurring midway of two slits, eq (8) becomes

$$\Delta f' = \frac{\delta f}{m h_0}. \tag{14}$$

3. Method of Measurement

The optical bench, shown in figure 1, is used in making the measurements required for the determination of the distortion of the wave front, n_λ, and the longitudinal spherical aberration $\Delta f'$. Initially the lens under test is mounted in its holder on the nodal slide assembly and its optical axis alined with the optical axis of the collimator. The reticle, located in the focal plane of the collimator, consists of a luminous cross composed of intersecting vertical and horizontal lines. The collimated beam is incident upon the lens under test and forms an image of the luminous reticle in the focal plane of the lens. The viewing microscope is adjusted using the transverse and longitudinal motions provided until the microscope is focused on the image formed by the lens and the vertical line of the image coincides with the intersection of the crosshairs in the ocular of the viewing microscope.

A diaphragm containing two parallel slits is placed in front of the lens under test. This diaphragm is so mounted that the double slit may be moved transversely across the front of the lens by measured amounts. The slits are alined parallel to the vertical line of the target reticle and so positioned that the center of the slit system will lie in the plane determined by the horizontal front diameter of the lens and the center of the target reticle for all positions to be occupied by the slit system throughout its range of transverse movement. At the start of a run, the slit system is brought to a position as near the edge of the lens as feasible. Light from the collimator on passing through the double slit and lens forms a set of interference fringes in the focal plane of the lens. A photograph of the fringe system is shown in figure 2. The crosshair of the viewing microscope is brought into coincidence with the image of the central fringe using the transverse micrometer movement of the microscope to make the setting. This setting of the transverse micrometer is recorded. The slits are now moved transversely by an amount equal to the slit separation, h_0, and the new location of the central fringe is recorded. This process is repeated until the entire diameter of the lens is spanned. The microscope setting made when one of the slits lies on the lens axis is taken as the zero point of the displacement, δ, and the corresponding setting of the double slit is taken as the zero point of the zone number, m. From the series of observations made at the measured intervals across the lens diameter, the values of δ as a function of m are obtained for the particular focal setting. In addition, it is customary to determine the average fringe separation $\bar{S}_0$ in the vicinity of the lens axis as it can be used in eq (1) for the determination of f when λ is known or for the evaluation of $\Sigma\delta$ in units of wavelength in eq (4).

It is recognized that the use of a collimator with the illuminated reticle located in its focal plane instead of an object at infinite distance may introduce a systematic error in the determination of the longitudinal spherical aberration and the optical path difference in units of wavelength. However, such systematic errors may be minimized by using a collimator having negligible longitudinal spherical aberration. The measurements, reported in the following section, were made using an off-axis paraboloidal mirror to collimate the light from the target reticle. For the area of the emergent beam used in the measurements the error arising from this source does not exceed ± 0.01 wavelength in the determination of n_λ and hence may be regarded as negligible.

4. Results of Measurement

A series of measurements was made on a photographic objective of a type commonly used in aerial photography. The lens has a measured focal length of 210.4 mm and a maximum nominal f-number of f/6.8. The slit separation was 1.453 mm and the

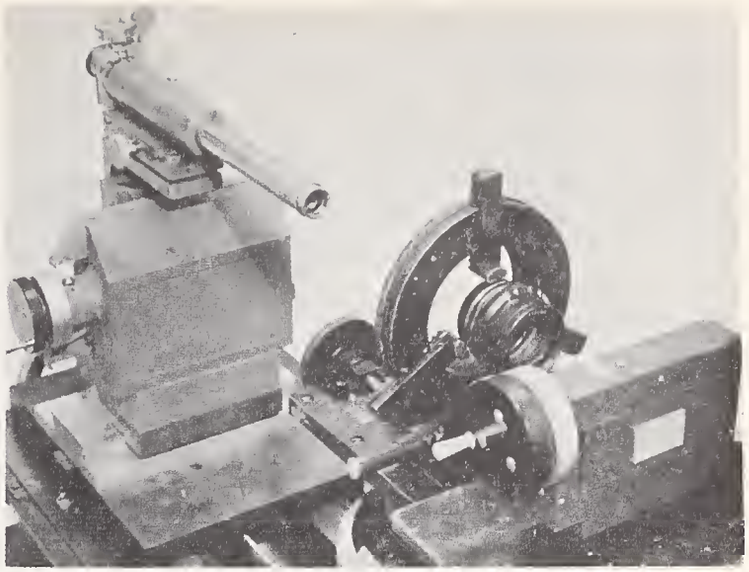


FIGURE 1. *Photograph showing optical bench with lens in position for test.*

The lens is mounted in the holder of the nodal slide. The double slit system is mounted in front of the lens on a transverse micrometer slide. Parallel light incident on the slit system passes through the lens and forms a fringe system in the focal plane of the lens under test. The fringe system is viewed with the microscope having a transverse micrometer movement.

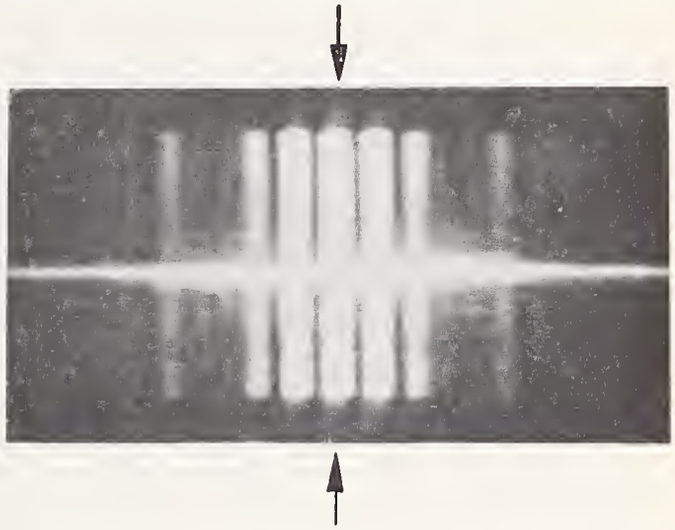


FIGURE 2. *Photograph of fringe system formed in the focal plane of the lens under test.*

In the photograph, the central fringe is indicated by the vertical arrows. The zero, first, and second order fringes are clearly visible; higher orders are present and may be seen through the microscope. The bright horizontal line is the image of the horizontal crossline in the collimator target.

effective wavelength of the light incident [4] on the lens was approximately 589.3 mμ. Observations of δ were made in the position of best focus and for two positions on either side of best focus. Values of δ for these three runs are listed in tables 1 to 3. Ten observations were made on either side of the optical axis and are designated by zone number m ranging from -10 to $+10$. Actual values of the zone height h are shown under the column headed $(m-\frac{1}{2})h_0$.

The observed values of the displacement δ of the central fringe with transverse displacement of the slit systems are shown graphically in the left hand frame of figure 3 which shows the variation of δ as a function of the zone height h for three positions of the image plane.

[4] The wavelength of the incident light is controlled by a narrow band interference filter whose peak transmittance is at $\lambda = 589.3$.

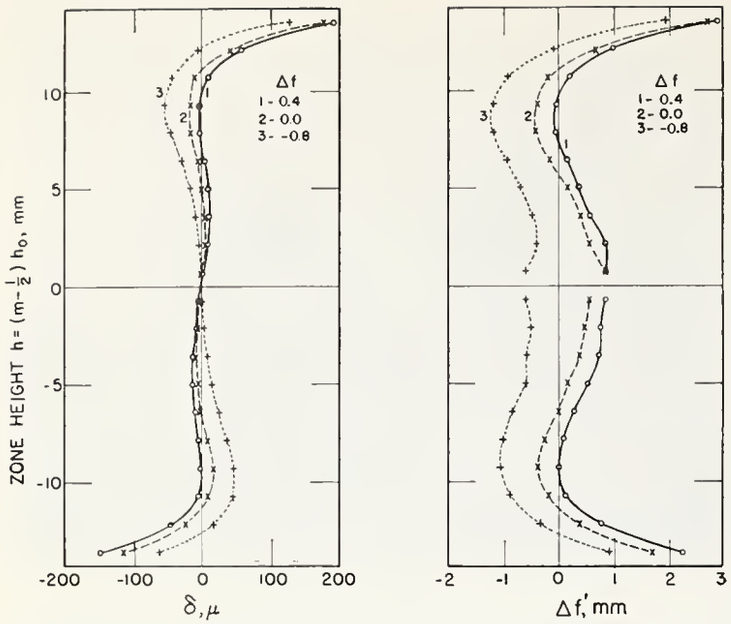


FIGURE 3. *Variation of transverse displacement, δ, of central fringe and computed value of longitudinal spherical aberration with zone height* h.

The left-hand frame shows the variation of δ with h for three positions of the focal plane. Curve 1 shows the results for a focal plane 0.4 farther from the lens than the position of best visual focus; curve 2 shows the results for the plane of best visual focus; and curve 3 the results for a plane 0.8 mm nearer to the lens than the plane of best visual focus. The right hand frame shows the values of the longitudinal spherical aberration $\Delta f'$ derived from the measured values of δ for the conditions described for the left hand frame.

TABLE 1. *Displacement δ in the image plane of central fringe as a function of zone height,* $h=(m-\frac{1}{2})h_0$

These values were obtained for a focal plane 0.4 mm farther from the lens than the plane of best focus of a lens having a focal length of 210.4 mm. The double slits are separated by h_0=1.453 mm. Wavelength of the incident light was 589.3 mm. Values of the longitudinal spherical aberration $\Delta f'$ and the distortion of the wave front, n_λ, in units of wavelength are also given.

m	$(m-\frac{1}{2})h_0$	δ	$\Delta f'$	$\Sigma\delta$	n_λ*
	mm	μ	*mm*	μ	
10	13.8	193	2.94	296	3.47
9	12.4	59	1.00	103	1.21
8	10.9	11	0.21	44	0.52
7	9.4	− 1	−.02	33	.39
6	8.0	− 2	−.05	34	.40
5	6.5	5	.16	36	.42
4	5.1	9	.37	31	.36
3	3.6	10	.58	22	.26
2	2.2	9	.87	12	.14
1	0.7	3	.87	3	.04
0	0	0			
−1	−0.7	− 3	0.87	− 3	0.04
−2	−2.2	− 8	.77	−11	.13
−3	−3.6	−13	.75	−24	.28
−4	−5.1	−13	.54	−37	.43
−5	−6.5	− 9	.29	−46	.54
−6	−8.0	− 4	−.10	−50	.59
−7	−9.4	− 1	.02	−51	.60
−8	−10.9	− 6	.12	−57	.67
−9	−12.4	−46	.78	−103	1.21
−10	−13.8	−148	2.26	−251	2.94

*$n_\lambda=\frac{(h_0)}{\lambda f}\sum\delta$.

4.1. Evaluation of Longitudinal Spherical Aberration $\Delta f'$

It is usually convenient to evaluate first the longitudinal spherical aberration $\Delta f'$ as it can be done readily using the observed values of δ and eq (13). The values of $\Delta f'$ are listed in tables 1 to 3 and are shown graphically in the right-hand frame of figure 3. It is clear that the form of the curves is essentially the same for the three focal positions but the curves are displaced with respect to each other by amounts closely approximating the nominal separations of planes in which the measurements were made.

TABLE 2. *Displacement δ in the image plane of central fringe as a function of zone height,* $h=(m-\frac{1}{2})h_0$

These values were obtained for the plane of best focus of a lens having a focal length of 210.4 mm. The double slits are separated by h_0=1.453 mm. Wavelength of the incident light was 589.3 mm. Values of the longitudinal spherical aberration $\Delta f'$ and the distortion of the wave front, n_λ, in units of wavelength are also given.

m	$(m-\frac{1}{2})h_0$	δ	$\Delta f'$	$\Sigma\delta$	n_λ*
	mm	μ	*mm*	μ	
10	13.8	178	2.71	192	2.25
9	12.4	41	0.70	14	0.16
8	10.9	−10	−.19	−27	−.32
7	9.4	−16	−.36	−17	−.20
6	8.0	−16	−.42	− 1	−.01
5	6.5	− 5	−.16	15	.18
4	5.1	4	.16	20	.23
3	3.6	7	.40	16	.19
2	2.2	6	.58	9	.10
1	0.7	3	.87	3	.04
0	0				
−1	−0.7	− 2	0.58	− 2	0.02
−2	−2.2	− 5	.48	− 7	.08
−3	−3.6	− 7	.40	−14	.16
−4	−5.1	− 4	.16	−18	.21
−5	−6.5	0	.00	−18	.21
−6	−8.0	9	−.24	− 9	.10
−7	−9.4	17	−.38	+ 8	− 09
−8	−10.9	9	−.17	17	−.20
−9	−12.4	−23	.39	− 6	.07
−10	−13.8	−114	1.74	−120	1.41

TABLE 3. *Displacement δ in the image plane of central fringe as a function of zone height,* $h=(m-\frac{1}{2})h_0$

These values were obtained for a focal plane 0.8 mm nearer to the lens than the plane of best focus of a lens having a focal length of 210.4 mm. The double slits are separated by h_0=1.453 mm. Wavelength of the incident light was 589.3 mm. Values of the longitudinal spherical aberration $\Delta f'$ and the distortion of the wave front, n_λ, in units of wavelength are also given.

m	$(m-\frac{1}{2})h_0$	δ	$\Delta f'$	$\Sigma\delta$	n_λ
	mm	μ	*mm*	μ	
10	13.8	130	1.98	−77	−0.90
9	12.4	− 4	− .07	−207	−2.42
8	10.9	−47	− .91	−203	−2.38
7	9.4	−53	−1.19	−156	−1.83
6	8.0	−44	−1.16	−103	−1.21
5	6.5	−29	−0.93	− 59	−0.69
4	5.1	−16	− .66	− 30	− .35
3	3.6	− 8	− .46	− 14	− .16
2	2.2	− 4	− .39	− 6	− .07
1	0.7	− 2	− .58	− 2	− .02
0	0				
−1	−0.7	2	−0.58	2	−0.02
−2	−2.2	5	− .48	7	− .08
−3	−3.6	10	− .58	17	− .20
−4	−5.1	14	− .58	31	− .36
−4	−6.5	26	− .84	57	− .67
−6	−8.0	38	−1.00	95	−1.11
−7	−9.4	47	−1.05	142	−1.66
−8	−10.9	46	− .89	188	−2.20
−9	−12.4	18	− .31	206	−2.41
−10	−13.8	−62	.94	144	−1.69

4.2. Check on Consistency of Observed Data

In work of this nature it is desirable to have and use check methods to ensure that errors are held to a minimum and to verify the selection of the form of equation used in the determination. This can be done by analysis of the variation $\overline{\delta}$, the average of δ,

for paired values of $+m$ and $-m$, for two focal positions. Consideration of eq (12) indicates that the difference $\Delta\bar{\delta}$ obtained from the differences in $\bar{\delta}$ for a given value of m at two different focal positions is a linear function of m. In table 4, values of $\Delta\bar{\delta}$ obtained by taking the differences between the corresponding values of $\bar{\delta}$ for focal positions $+0.4$ and -0.8 are listed under the caption, observed values of $\Delta\bar{\delta}$. When these values of $\Delta\bar{\delta}$ are plotted against zone number m as shown in curve 1 of figure 4, it is clear that a linear relation exists between $\Delta\bar{\delta}$ and $m-\frac{1}{2}$. This indicates that it was proper to use eq (13) in the evaluation of $\Delta f'$. The slope a of the best fitting line through the points on the graph is given by the expression

$$a=\frac{\Sigma\Delta\bar{\delta}}{\Sigma(m-\frac{1}{2})} \tag{15}$$

which in this instance gives a value

$$a=383/50=7.66\ \mu.$$

The equation of the best fitting line is

$$\Delta\bar{\delta}_c=a(m-\tfrac{1}{2}). \tag{16}$$

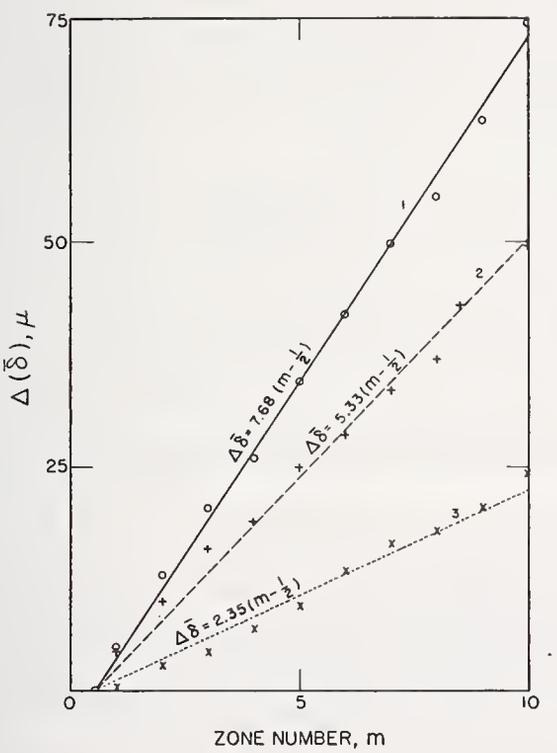


FIGURE 4. *Variation of the change in average transverse fringe displacement* $\Delta\bar{\delta}$ *between two specified focal planes with zone number,* m.

Curve 1 shows the variation in $\Delta\bar{\delta}$ for focal planes designated $\Delta f=0.4$ and $\Delta f=-0.8$; curve 2 shows the variation in $\Delta\delta$ for the focal planes designated $\Delta f=0.0$ and 0.8; and curve 3 shows the results for focal planes designated $\Delta f=0.4$ and $\Delta f=0.0$ mm. The equations of the best fitting straight line are shown adjacent to each curve.

Values of $\Delta\delta_c$ obtained with this equation for $a=7.66\ \mu$ are listed in table 4 together with the departures from the observed values of $\Delta\bar{\delta}$. It is clear that the agreement is excellent. Values of $\Delta\bar{\delta}$ for the other two combinations of focal position ($+0.4$ and 0.0, curve 3; 0.0 and -0.8, curve 2) are also shown in the graph and indicate that all three sets of data are reasonably consistent.

TABLE 4. *Comparison of values of* $\Delta\bar{\delta}$ *obtained for focal positions* $+0.4$ *and* -0.8 *as a function of* m *with values computed from relation* $\Delta\bar{\delta}=7.66\ (m-\frac{1}{2})$

m	$\Delta\bar{\delta}$ in microns		
	Observed	Comp	O−C
1	5	4	1
2	13	11	2
3	20	19	1
4	26	27	−1
5	34	34	0
6	42	42	0
7	50	50	0
8	55	57	−2
9	64	65	−1
10	74	73	1

From the values of the slope a, for the three combinations of $\Delta\delta$ versus m, it is possible to determine the average displacement of the focal plane for the three focal positions using the relation

$$\Delta(\overline{\Delta f})=\frac{af}{h_0}. \tag{17}$$

Values of $\Delta(\overline{\Delta f})$ derived from the analysis of $\overline{\Delta\delta}$ for the three combinations are given in table 5. The focal positions were known initially within ± 0.1. mm so it is clear that good agreement exists between the nominal values of $\Delta(\overline{\Delta f})$ and those derived from analysis of $\Delta\bar{\delta}$ as a function of m.

TABLE 5. *Comparison of nominal values of the relative separation* $\Delta(\Delta f)$ *of the three focal planes with that derived from the slopes of* $\Delta\delta$ *versus* $m-\frac{1}{2}$ *and with that derived from the slope of* $\Delta\bar{n}_\lambda$ *versus* m^2

Focal positions	Nominal	Value of $\Delta(\Delta f)$	
		From slope of $\Delta\delta$ vs $m-\frac{1}{2}$	From slope of Δn_λ vs m^2
	mm	*mm*	*mm*
0.0 and 0.4	0.4	0.34	0.32
0.0 and −0.8	.8	.77	.80
0.4 and −0.8	1.2	1.11	1.12

4.3. Evaluation of Optical Path Difference, n_λ

The values of the optical path difference, n_λ, are obtained with the aid of eq (4). Here the value of

518-743 O - 73 - 40

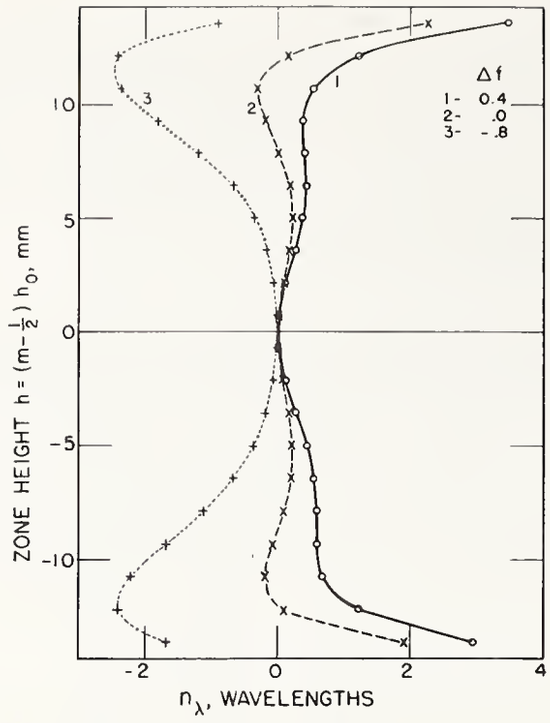


FIGURE 5. *Variation of the optical path difference expressed in number of wavelengths,* n_λ, *as a function of zone height* h.

Values of n_λ are shown for three focal positions. Curve 1 shows the results for $\Delta f=0.4$, curve 2 for $\Delta f=0.0$, and curve 3 for $\Delta f=-0.8$. These values of n_λ are derived from the values of δ shown in figure 2.

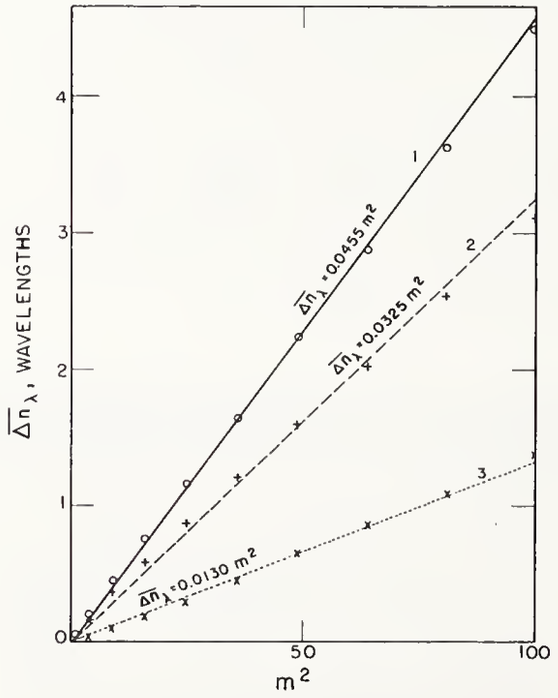


FIGURE 6. *Variation of the average change* $\Delta\bar{n}_\lambda$ *between two specified focal planes with* m^2, *where* m *is the zone number.*

Curve 1 shows the variation in $\Delta\bar{n}_\lambda$ for focal planes designated $\Delta f=0.4$ and $\Delta f=-0.8$ mm; curve 2 shows the variation for the focal planes designated $\Delta f=0.0$ and -0.8 mm; and curve 3 shows the variation for focal planes designated $\Delta f=0.0$ and 0.4 mm. The equations of the best fitting straight lines are shown adjacent to each curve.

the term $h_o/\lambda f$ is equal to 0.01172 wavelengths/micron. Values of $\Sigma\delta$ and n_λ for the three focal positions are shown in tables 1 to 3. The values of n_λ are also shown graphically in figure 5.

4.4. Check on Consistency of Values of n_λ

While it is possible to evaluate n_λ from data obtained for a single focal position, confidence in the results is increased by comparison with results for a different focal position. Hence it is desirable to check the inner consistency of results obtained for several focal positions to ensure that errors have been held to a minimum and to verify the form of equation to be used in determing the change in n_λ with shift of focus. This can be done by analysis of $\overline{n_\lambda}$, the average value of $\overline{n_\lambda}$ for paired values of $+m$ and $-m$ for two focal positions. Consideration of eq (8) indicates that the differences $\Delta\overline{n}_\lambda$ obtained from the differences in n_λ for a given value of m at two different focal positions is a linear function of m^2. In table 6, values of $\Delta\overline{n}_\lambda$ obtained by differencing the corresponding values of $\overline{n}_\lambda$ for focal positions $+0.4$ and -0.8 are listed under the caption, observed values of $\overline{\Delta n_\lambda}$. When these values of $\overline{\Delta n_\lambda}$ are plotted against m^2 as shown in curve 1 of figure 6, it is clear that a linear relation exists between $\overline{\Delta n_\lambda}$ and m^2. This also indicates that eq (8) was the proper choice for use in the interpretation of the relation between $\overline{\Delta n_\lambda}$ and m. The slope b of the best fitting line through the points on the graph is given by the expression

$$b=\frac{\Sigma\overline{\Delta n}\lambda}{\Sigma m^2} \tag{18}$$

which in this instance gives a value of

$$b=17.53/385=0.04553$$

The equation of the best fitting line is

$$\overline{\Delta n_\lambda}=bm^2. \tag{19}$$

Values of $\overline{\Delta n_\lambda}$ obtained with this equation for $b=0.04553$ are listed in table 6 under the caption, computed values of $\overline{\Delta n_\lambda}$ together with the departures from the observed values under the caption, $0-C$. It is clear that the agreement is excellent. Values of $\overline{\Delta n_\lambda}$ for the other two combinations of focal position (0.4 and 0.0, curve 3; 0.0 and -0.8, curve 2) are also shown in the graph and good consistency is indicated for the three sets of data although the values of $0-C$ are larger at $m=9$ and $m=10$.

From the values of the slope b for the three combinations it is possible to determine the average displacement of the focal plane for the three focal positions using the relation

$$\Delta(\overline{\Delta f})=\frac{2\lambda f^2}{h_o^2}\cdot b. \tag{20}$$

Values of $\Delta(\overline{\Delta f})$ derived from analysis of $\overline{\Delta n_\lambda}$ for the three combinations are also listed in table 5. It is clear that good agreement also exists between the nominal values of $\Delta(\overline{\Delta f})$ and those obtained in the foregoing manner.

TABLE 6. *Comparison of values of* $\overline{\Delta n_\lambda}$ *obtained for focal position* $+0.4$ *and* -0.8 *as a function of* m *with those computed from the relation* $\overline{\Delta n_\lambda}=0.04553m^2$

m	Values of $\overline{\Delta n_\lambda}$		$0-C$
	Observed	Computed	
1	0.06	0.05	0.01
2	.21	.18	.03
3	.45	.41	.04
4	.76	.73	.03
5	1.16	1.14	.02
6	1.65	1.64	.01
7	2.24	2.23	.01
8	2.88	2.91	−.03
9	3.63	3.69	−.06
10	4.49	4.55	−.06

4.5. Effect of Focal Shift Δf on the Values of n_λ

One of the manifest advantages of this method of measuring optical path difference n_λ, is that it permits the evaluation of n_λ as a function of zone height h for any given focal position separated by Δf from the focal plane in which measurements were made. Hence it is not necessary to make the measurements of δ at the position of best axial focus. It is only necessary that the focal plane in which measurements are made is that of a reasonably good focus as the form of the curve can be computed for a selected focal plane if one has the measured values of n_λ as a function of m for a nearby plane. The suitability of such a procedure is illustrated by comparison of the 3 sets of values of n_o, the adjusted values of $\overline{n_\lambda}$, obtained from $\overline{n_o}$, the observed values of $\overline{n_\lambda}$ for the three focal positions for which results are given in tables 1, 2, and 3.

Values of n_0, the observed values of n_λ, as a function of zone number are given for the three focal positions in table 7. It is clear that the values differ by appreciable amounts. It is worthwhile to adjust these values by a focal shift of such amount that n_c, the adjusted values of n_λ, shall be zero at $m=7$. This may be done using the relation

$$n_0-\Delta n_\lambda=n_c$$

where

$$\Delta n_\lambda=\left(\frac{n_0'}{49}\right)m^2 \tag{21}$$

and n_0' is the observed value of n_λ for zone number $m=7$. Values of n_c obtained in this manner are also listed in table 7 together with $\overline{n}_c$, the average for the three focal planes. It is clear that there is good agreement among the three values of n_c although the determinations are based upon measurements made at markedly different focal positions. Consideration of the departures from the average $\overline{n}_c$ shows that the average error in n_λ does not exceed ± 0.02 for the range from $m=1$ to $m=8$ and does not exceed ± 0.05 for values of $m=9$ and $m=10$.

TABLE 7. *Values of* n_c *as a function of zone number* m *for three focal planes*

Values of n_c obtained from n_0, the observed values of n_λ, for three focal planes. Values of Δn_λ are so chosen that $n_c=0$ at $m=7$. The value of $\overline{n}_c$, the average of n_c for the three focal planes, for each value of m is also given.

m	$\Delta f=0.4$ mm			$\Delta f=0.0$ mm			$\Delta f=-0.8$ mm			$\overline{n}_0$
	n_0	Δn_λ	n_c	n_0	Δn_λ	n_c	n_0	Δn_λ	n_c	
1	0.04	0.01	0.03	0.03	0.00	0.03	−0.02	−0.04	0.02	0.02
2	.13	.04	.09	.09	−.01	.10	−.08	−.14	.06	.08
3	.27	.09	.18	.18	−.03	.21	−.18	−.32	.14	.18
4	.40	.16	.24	.22	−.05	.27	−.36	−.57	.21	.24
5	.48	.25	.23	.19	−.08	.27	−.68	−.89	.21	.24
6	.49	.36	.13	.05	−.11	.16	−1.16	−1.29	.13	.14
7	.49	.49	.00	−.15	−.15	.00	−1.75	−1.75	.00	.00
8	.59	.64	−.05	−.26	−.20	−.06	−2.29	−2.29	.00	−.04
9	1.21	.81	.40	.12	−.25	.37	−2.42	−2.89	.47	.41
10	3.20	1.00	2.20	1.83	−.31	2.14	−1.29	−3.57	2.28	2.21

The average value $\overline{n}_c$ versus zone height h is shown in figure 7. It is clear that for this particular focal position the optical path difference does not exceed ¼ wavelength for the range from $m=1$ to $m=8$. From this it may be inferred that the optical

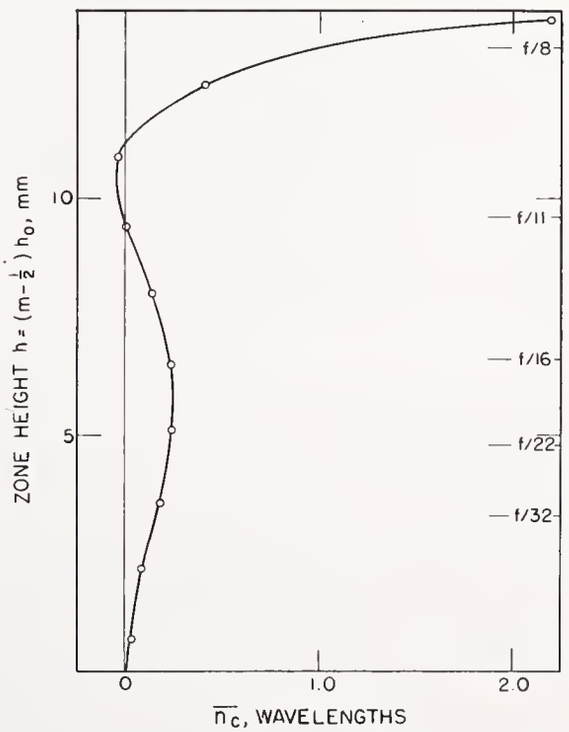


FIGURE 7. *Variation of the average adjusted value,* n_c *with zone height* $h=(m-\frac{1}{2})h_0$ *where* m *is the zone number and* h_0 *the slit separation.*

The value of $\overline{n}_c$ is the average value of $\overline{n}_\lambda$ for the three focal planes designated $\Delta f=0.4$, 0.0, and -0.8 following adjustment such that $n_c=0$ at $m=7$ in all three cases. The scale on the right-hand ordinate of the graph shows the value of the f/number corresponding to the indicated value of h, the zone height.

path difference will not exceed ¼ wavelength for one-half the effective area of the lens at maximum aperture. Moreover, it can be inferred that this lens should yield excellent imagery at f/11, good imagery at f/10, and that the quality of imagery can be expected to lessen appreciably at f/8 and smaller f/values.

5. Discussion

In this study, a visual method of determining longitudinal spherical aberration and departure of the wave front arriving at the image point from a true sphere is described. Results are reported for measurements made on a single lens for three positions of the focal plane. Tests for consistency of observations are described which permit an estimate of the reliability of the measurements.

The authors express their appreciation to other members of the staff for assistance during this work and particularly to Edgar C. Watts who designed the slit system and prepared the illustrations; and to George Cassiday and Peter Costello who performed many of the measurements.

(Paper 67C4–142)

Experimental Verification of Superachromatism

ROBERT E. STEPHENS
National Bureau of Standards, Washington, D. C. 20234
(Received 20 September 1965)

To verify superachromatism experimentally a five-element prism, to deviate a beam of light approximately 30 degrees without dispersion or distortion, has been designed and constructed from three types of Schott glass. When tested on a spectrometer, viewing a narrow slit with a telescope magnification of approximately 30 times, no dispersion could be seen. Pointings on 7 isolated spectrum lines separated by a monochromator show a maximum dispersion of 3 sec, 1 part in 36 000.

INDEX HEADINGS: Dispersion; Refractive index; Aberration; Lens design; Refraction.

PHOTOGRAPHIC lenses of focal lengths up to about 6 in. usually have resolving powers greater than that of the film that is used to record the image. Consequently, less information is recorded on the film than was present in the image projected upon it by the lens. The obvious remedy for this situation is to increase the magnification until the resolving power of the film matches that of the lens. Then, by the use of a correspondingly larger film, the information in the image may be recorded. The angular resolving power of the lens is not changed by this procedure if the aperture is kept constant, and the aberrations are suitably corrected. The magnification is expeditiously produced by an increase in focal length. It would be expected then that the resolving power and the recording of information could be increased indefinitely by increasing both the aperture and the focal length, keeping the resolving power of the film equal to or greater than that of the lens. That this belief is widely held is attested to by the extensive use of telephoto lenses for aerial reconnaissance with apertures of 6 to 10 in. and focal lengths from 30 to 50 in., also by the construction of an experimental lens of 30-in. aperture and 20-ft focal length.

Photographs made with such large lenses always lack the sharpness and crispness of those made with lenses of modest aperture and focal lengths up to about 6 in. Calculations have shown that residual secondary longitudinal chromatic aberration drastically limits the information that can be obtained by photography. Only if the focal length is as long as given by Conrady's formula can the secondary chromatic aberration be made innocuous and the theoretical resolving power be realized.[1] This results in a relative aperture that is much too small for most photographic purposes.

In photographic lenses of usable relative aperture, the residual chromatic aberration limits the usable aperture independently of the focal length. The limit appears to be reached for telephoto and anastigmatic lenses at an aperture of 2 to 3 in. The Petzval type is less limited, thus may be extended with satisfactory results to apertures of 4 to 5 in.

The deductions mentioned above were made on the assumption that photographs are to be made with light containing the whole visible spectrum. This is, of course, the case if color photographs are to be made or if panchromatic film is used. Apertures are also limited, but to larger values, if the spectral range is restricted by filters or the use of orthochromatic film.

Until recently very little could be done about the residual chromatic aberration. Three-color achromats have been made for telescope and microscope objectives. These have a larger usable spectral range but the residual tertiary chromatic aberration is almost as large as the secondary chromatic aberration of a two-color achromat. I am not aware of any photographic lens that has three-color achromatism although two lenses containing fluorite have been reported.[2,3]

Several years ago Herzberger[4] predicted that a lens achromatized for four wavelengths would have no residual chromatic aberration in the range from 0.365 to 1 μ. He also predicted that 4-wavelength correction could be accomplished by the use of only 3 properly chosen glasses. Somewhat later Stephens[5] devised a procedure for selecting suitable glasses and showed that there exists at least 1 set of 3 glasses that is marginally suitable. A lens with this type of chromatic correction has been named a *superachromat.*

It was deemed desirable to test *superachromatism* experimentally. To do so with a lens would require the design and construction of an anastigmatic lens of at least 5-in. aperture or a Petzval type of at least 10-in. aperture to definitely demonstrate its superiority over two-color achromats. The corrections of the monochromatic aberrations would have to be superb to obtain the benefit of the excellent chromatic correction. The design and construction of such a lens is impracticable for us at this time.

It occurred to me that *superachromatism* could be tested more cheaply, more easily, and better by the design and construction of a *superachromatic* prism. A

[1] A. E. Conrady, *Applied Optics and Optical Design* (Dover Publications, Inc., New York, 1957), p. 201.

[2] J. G. Baker, J. Opt. Soc. Am. 36, 722A (1946).
[3] F. H. Perrin and H. O. Hoadley, J. Opt. Soc. Am. 38, 1040 (1948).
[4] Max Herzberger, Opt. Acta 6, 197 (1959).
[5] R. E. Stephens, J. Opt. Soc. Am. 50, 1016 (1960).

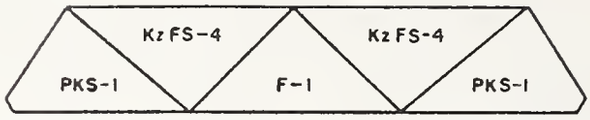


FIG. 1. Cross section of superachromatic prism showing types of glass (Schott) used. The width of the incident and emergent beams is about 15 mm, not enough to resolve a three-second line separation.

prism used to deviate collimated light near the position of minimum deviation does not produce monochromatic aberrations or distortion. Consequently, any residual chromatic aberration would be easily detectable.

Such a prism has been designed and constructed. The Schott glass types and principal section are shown in Fig. 1. The initial design was computed using the catalog values of indexes of refraction. It was anticipated that this preliminary design would not be successful; consequently, all three surfaces of the components were polished so that an acute angle would be available for making index measurements.

When the components according to the preliminary design were assembled and the deviation examined, it was found that there was indeed a moderately large residual of chromatic aberration. It was not measured. Instead the indexes of the individual components were measured for 3 wavelengths. The design was then recomputed for the actual indexes. The components were sent back to the optical shop for slight modifications to conform to the new design.

The five components of the prism were not cemented together. A small drop of cedar oil was used between adjacent prisms. The capillary effect tends to pull the prisms together and to produce a film of oil of uniform thickness. This reduces the chance of some disperison resulting from wedges of cement.

The finished prism deviates a beam of light approximately 30 degrees, which is equivalent to the deviation by the marginal zone of a lens with a relative aperture of $f/1.0$. When tested on a spectrometer with a telescope magnifying approximately 30 times, no trace of dispersion could be seen. A narrow slit, illuminated by an incandescent lamp, a mercury discharge lamp, or a cadmium discharge lamp, was viewed.

To detect any dispersion it was necessary to separate single spectrum lines from the light of the discharge lamps by means of a monochromator. Pointings were made on 7 spectrum lines from Cd 4800 Å to Cd 6438 Å (including some mercury lines). The spectrometer telescope was clamped, pointings and readings being made by means of the eyepiece micrometer. Some spectrum lines outside this range were tried, but those nearer the ends of the visible spectrum were so weak after coming through the monochromator that the cross hairs were not illuminated enough to be seen; consequently a setting could not be made. The maximum separation of the 7 lines that were usable was 3 sec, approximately 1 part in 36 000. This contrasts with approximately 1 part in 2000 over the same spectral range for two-color achromats.

Although I have been unable to carry my measurements to the desired limits in the spectrum, nothing was observed from 4800 to 6438 Å to suggest distrust of the computations. Consequently, it is considered that the existence of *superachromatism* has been verified both experimentally and by computation.

The difficulties encountered in the construction of a *superachromatic* system are formidable. As success depends upon a delicate balance of large positive and negative dispersions, the indexes of refraction used in the computations must be as accurate as possible. In general, the index of each piece of glass to be used in such a system should be measured with an accuracy of $\pm 1 \times 10^{-5}$ and the measured values used in computing the design. It is quite unlikely that the prototype indexes given in the glass catalog will suffice. After the design is completed, the elements must be manufactured as near to the specifications as practicable. In the case of the prism the angles of the components were accurate to within ± 1 min. That ± 1 min was sufficient in this case may be fortuitous since the refraction of a 1-min prism is about 30 sec and its dispersion 30/V sec. For the flint glasses this could amount to almost a second between Fraunhofer lines *F* and *C*.

Reprinted from:
JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 62, NUMBER 2 FEBRUARY 1972

Conditions for Microdensitometer Linearity*

RICHARD E. SWING
National Bureau of Standards, Washington, D. C. 20234
(Received 4 December 1970)

A simple microdensitometer system is analyzed using the principles and analytical techniques of the theory of partial coherence. A specification of the physical conditions under which the instrument is linear is obtained, for incoherent illumination. The illuminating mutual intensity is then generalized, by the Van Cittert–Zernike theorem, and conditions on the partial coherence of the preslit illumination necessary for effective incoherence are determined. The new conditions determine a mode of linear operation for the microdensitometer for which an optical transfer function may be unambiguously defined.

INDEX HEADINGS: Microdensitometer; Coherence.

The one essential instrument for the evaluation of optical and photographic images is the microdensitometer. This instrument can provide the analyst with precise measurements of the density microstructure of an image, from which he may then deduce information about the original luminance distribution or assess the quality of the system. Both functions require an instrument whose response is independent of the sample being evaluated. This is also the fundamental requisite for the quantitative relation of interlaboratory measurements of the same sample. The requirement is an attribute of a linear system.

Optical systems with partially coherent illumination have been rigorously shown to be nonlinear[1,2] for edges and sinusoids. Prior to these studies, a large body of diverse evidence had accumulated which showed that the microdensitometer and similar optical imaging devices were generally nonlinear for other types of objects. Calculations were carried out for images of specific objects such as opaque disks and lines, with great attention to the specification of partial coherence.[3–8] Collaterally, experiments were carried out[9,10] that confirmed one of the important results of the calculations: the significant parameter clearly affecting the linearity (related to incoherence of the illumination) was the ratio of the numerical aperture of the influx objective to that of the efflux objective. Whenever this ratio exceeded unity, results approached the theoretical predictions for incoherent illumination. This effect had previously been documented[11] in 1950 in terms of the resolution of two pinholes of equal brightness. Recently, the relation of this ratio to the resolution limit of the efflux optics and the limiting sample spatial frequency, and its attribution to a condition of "strong incoherence" was cited,[12] although the essential theoretical development was omitted.

The literature to this point has generally considered systems that were incoherent in the sample plane or in the pupil of the illuminating condenser. This presupposed certain conditions on the illumination which might not be completely realizable in actual systems. The question of whether optical-system linearity is a natural concomitant of incoherent illumination has only been partially answered, and distinctions between linearity and incoherence are difficult to make when it is necessary to account for the state of the illumination at the source slit. It is therefore essential to differentiate between the two and to determine the conditions under which the microdensitometer system performs linearly and the conditions under which the illumination may be considered incoherent, or effectively incoherent.

This paper is concerned only with the theoretical aspects of these two problems. Image formation with the microdensitometer will be formulated with the aid of the theory of partial coherence. The microdensitometer is commonly used in a scanning mode, the output of which is a measure of specular density (or transmittance) as a function of time. It is not the purpose of this paper to consider the scanning aspects of the problem, inasmuch as they have no real effect on the basic problems of linearity and incoherence. The essential problem is the relation of measured (imaged) flux transmittance to the corresponding transmittance value at the sample, as a function of the optical-system parameters that affect its linearity.

ANALYSIS

Optical System

A typical microdensitometer optical system is shown[13] in Fig. 1. The lenses are conventionally high-quality microscope objectives. The coordinate systems are defined in the figure and the object and image distances are shown. The source slit is imaged onto the sample (nearly always at a reduction), and the sample is imaged onto the sensor aperture. Thus, with the system in focus, object and image distances satisfy the gaussian thin-lens equation, with the magnifications given by

$$m_1 = -z_2/z_1; \tag{1}$$

$$m_2 = -\xi_2/\xi_1. \tag{2}$$

The $\bar{\mu}$, $\bar{w}$, and $\bar{x}$ planes are therefore conjugate. The pupil functions of the influx and efflux optics are $F_1(\bar{\alpha})$ and $F_2(\bar{\beta})$, respectively. The Fourier transforms of these functions are the complex impulse responses of the respective lenses. The tilde ($\sim$) over a function denotes the Fourier transform of that function in its conjugate variable.

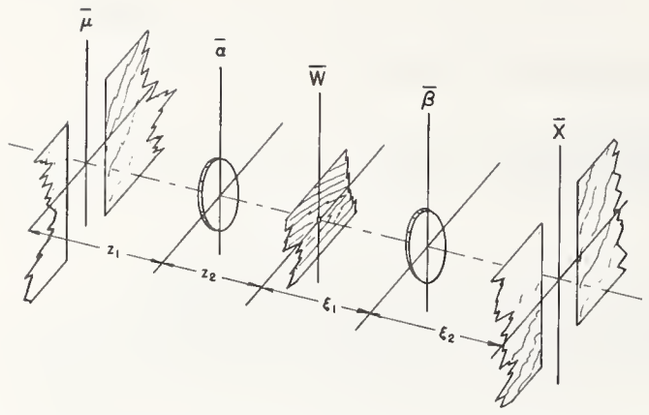


FIG. 1. Typical microdensitometer optical system, showing location of the coordinate systems and the object and image distances for the optical elements. Component nomenclature and the corresponding coordinate systems are: $\bar{\mu}$—source aperture; $\bar{\alpha}$—influx optics; $\bar{w}$—sample plane; $\bar{\beta}$—efflux optics; $\bar{x}$—image plane and sensor aperture.

Assumptions and Definitions

The microdensitometer optics, constrained by the slit configuration, operate effectively on axis. Further, because the slit widths used in microdensitometry are very small compared with the propagation distances along the optical axis, standard paraxial approximations are appropriate. For these reasons also, system stationarity will be assumed; i.e., the system impulse response will be considered constant over the field of interest. To reduce the number of variables and integrations, the analysis will be limited to one dimension.

The methods of the theory of partial coherence will be employed. To expedite the development, the quasi-monochromatic approximation[14] will be taken. This provides a useful solution for some mean wavelength, $\bar{\lambda}$. The dependence of the system components on $\bar{\lambda}$ will be neglected.[15] The theory of partial coherence requires a description of the incident illuminating field. This will be given by a quantity called the mutual intensity. This function characterizes the state of coherence in the illumination, by an implicit point-by-point comparison or correlation across the field. Optical systems respond linearly to mutual intensity, and an analysis of an imaging system can be reduced to characterizing the propagation of mutual intensity from source to image. The notation to be used generally follows that employed by Beran and Parrent.[14]

The radiance characteristics of the source slit are described by $T(\mu)$. In terms of complex quantities, this is given by

$$T(\mu)=t(\mu)t^*(\mu), \tag{3}$$

where the asterisk signifies the complex conjugate.

The sample, inserted in the w plane, has radiance characteristics denoted by $\psi(w)$. In complex notation, by analogy with Eq. (3),

$$\psi(w)=\varphi(w)\varphi^*(w). \tag{4}$$

This is a general functional description of a spatial distribution of sample transmittance. Its actual value, as measured with the microdensitometer, will depend on the size of the sample area under examination as well as the cone angle of the collecting optics (degree of specularity). $\psi(w)$ is usually a photographic transparency, and therefore physically band limited. Denote this upper spatial frequency value by η_0. The amplitude spectrum of the transparency is given by[16]

$$\tilde{\varphi}(\eta)=\int \varphi(w)\exp[-2\pi i\eta w]dw. \tag{5}$$

Now

$$\tilde{\psi}(\eta)=\int \varphi(w)\varphi^*(w)\exp[-2\pi i\eta w]dw$$

$$=\int \tilde{\varphi}(\eta+\theta)\tilde{\varphi}^*(\theta)d\theta, \tag{6}$$

and this is clearly the autocorrelation of the complex spectrum. Because this will have a width twice that of the correlating functions, $\tilde{\varphi}(\eta)$ and $\tilde{\varphi}^*(\eta)$ will be assumed to be band limited to $\eta_0/2$. To avoid problems associated with specular measurement of photographic emulsions, sample materials will be considered grainless for present purposes.

Image Mutual Intensity

Consider an illuminating field incident on the source slit and denote the general mutual intensity by $\Gamma_0(\mu_1,\mu_2)$. This propagates through the system to produce the image mutual intensity, $\Gamma(x_1,x_2)$, in the $\bar{x}$ plane. The equation that describes the mutual intensity in that image can be written in terms of complex impulse responses,[17] with several quadratic phase factors,[18]

$$\Gamma(x_1,x_2)=\left\{\exp\left[\frac{ik(x_1^2-x_2^2)}{2\xi_2}\right]\right\}\int\Gamma_0(\mu_1,\mu_2)$$

$$\cdot t(\mu_1)t^*(\mu_2)\exp\left[\frac{ik(\mu_1^2-\mu_2^2)}{2z_1}\right]$$

$$\cdot\varphi(w_1)\varphi^*(w_2)\exp\left[ik\left(\frac{w_1^2-w_2^2}{2}\right)\left(\frac{1}{z_2}-\frac{1}{\xi_1}\right)\right]$$

$$\cdot\tilde{F}_1\left(\frac{w_1}{\lambda z_2}+\frac{\mu_1}{\lambda z_1}\right)\tilde{F}_1^*\left(\frac{w_2}{\lambda z_2}+\frac{\mu_2}{\lambda z_1}\right)$$

$$\cdot\tilde{F}_2\left(\frac{x_1}{\lambda\xi_2}+\frac{w_1}{\lambda\xi_1}\right)\tilde{F}_2^*\left(\frac{x_2}{\lambda\xi_2}+\frac{w_2}{\lambda\xi_1}\right)d\mu_1 d\mu_2 dw_1 dw_2. \tag{7}$$

For normal microdensitometer operation, $T(\mu)$, representing the slit, is a rectangular function. This

could be incorporated with the mutual intensity to describe the system illumination in the manner utilized by Mielenz.[19] However, it will be useful to keep these characterizations explicit, so that the interrelation of slit, sample, and illumination may be more advantageously studied.

Before undertaking the reduction of Eq. (7), consider the quadratic expressions in μ and w. The physical constraints afforded by normal microdensitometer operation serve to limit their effect. In the instrument, for nearly all modes of operation, the source slit width is very small compared with the object distance, z_1. The quadratic in μ can therefore be neglected.[20] Further, since this slit is imaged at a reduction onto the $\bar{w}$ plane, the extent of the illuminated sample in w will similarly be small compared to the propagation distances involved. This quadratic factor will also be neglected.

Linearity

It will now be useful to establish a criterion for linearity, and then to develop the analysis towards its application. One of the best-known attributes of linear systems is that the image spectrum equals the object spectrum times the transfer function. This implies an analytical separation of the system transfer function and the object spectrum. Thus, when the image spectrum is calculated, the system will be considered linear when transfer function and object spectrum are independent. The equation expressing this is

$$\tilde{I}(\sigma)=\tilde{I}_0(\sigma)\tau(\sigma), \tag{8}$$

where $\tilde{I}_0(\sigma)$ is the object spectrum, $\tilde{I}(\sigma)$ the image spectrum, $\tau(\sigma)$ the system transfer function.

The analysis can now continue, with a different emphasis. Because it is necessary to calculate the image spectrum, the irradiance limit

$$\begin{aligned} I(x)&=\lim\Gamma(x_1,x_2),\\ x_1&\to x,\\ x_2&\to x, \end{aligned} \tag{9}$$

will first be applied to Eq. (7). Then the image spectrum will be obtained through application of the Fourier transform

$$\tilde{I}(\sigma)=\int I(x)\exp[-2\pi i\sigma x]dx. \tag{10}$$

The Image Spectrum

When the quadratic terms are neglected, and the operations of Eqs. (9) and (10) are applied to Eq. (7), an expression for the general image spectrum is obtained. After a lengthy calculation, the expression takes the form

$$\begin{aligned} \tilde{I}(\sigma)=\int & F_1[-\zeta S-\lambda\xi_2\sigma S-\lambda z_2\eta]F_1^*[-\zeta S-\lambda z_2\eta']\\ &\cdot F_2[\lambda\xi_2\sigma+\zeta]F_2^*[\zeta]\tilde{\varphi}(\eta)\tilde{\varphi}^*(\eta')\\ &\cdot\left\{\int\Gamma_0(\mu_1,\mu_2)l(\mu_1)l^*(\mu_2)\right.\\ &\cdot\exp\left[-2\pi i\mu_1 m_1\left(\sigma m_2-\frac{\zeta}{\lambda\xi_1}-\eta\right)\right]\\ &\left.\cdot\exp\left[-2\pi i\mu_2 m_1\left(\frac{\zeta}{\lambda\xi_1}+\eta'\right)\right]d\mu_1 d\mu_2\right\}d\eta d\eta' d\zeta, \end{aligned} \tag{11}$$

where

$$S=z_2/\xi_1, \tag{12}$$

and η and η' are the spatial frequency variables in the sample plane. Because the incident mutual intensity remains general, it will not be possible to continue without considering specific functional forms of $\Gamma_0(\mu_1,\mu_2)$.

The Incoherent Limit

Preliminary to a broader generalization, it will be useful to consider the case of incoherent illumination. This is obtained when the mutual intensity is represented by[21]

$$\Gamma_0(\mu_1,\mu_2)=I_0(\mu_1)\delta(\mu_1-\mu_2). \tag{13}$$

This illumination over the source slit will be assumed constant, with an effective value of unity for this analysis. When Eq. (13) is substituted in Eq. (11), and the relation of Eq. (3) is invoked during the ensuing calculations, the image spectrum becomes

$$\begin{aligned} \tilde{I}(\sigma)=\int & F_1[-\zeta S-\lambda\xi_2\sigma S-\lambda z_2\eta]F_1^*[-\zeta S-\lambda z_2\eta']\\ &\cdot F_2[\lambda\xi_2\sigma+\zeta]F_2^*[\zeta]\tilde{\varphi}(\eta)\tilde{\varphi}^*(\eta')\\ &\cdot\tilde{T}[\sigma m_1 m_2-m_1(\eta-\eta')]d\eta d\eta' d\zeta. \end{aligned} \tag{14}$$

Constant Sample Transmittance

The sample has been illuminated with the reduced image of the source slit. When this image width is smaller than any detail variation in $\varphi(w)$, the transmittance is effectively constant. For this,

$$\begin{aligned} \tilde{\varphi}(\eta)&\cong c_1\delta(\eta),\\ \tilde{\varphi}^*(\eta')&\cong c_1^*\delta(\eta'), \end{aligned} \tag{15}$$

c_1 is the complex sample transmittance and

$$c=c_1c_1^*. \tag{16}$$

This condition is exemplified by a sample of neutral density, or a density tablet step. When Eqs. (15) are substituted in Eq. (14) and the integrations over η and

η' carried out, the image spectrum becomes

$$\tilde{I}(\sigma)=c\cdot\tilde{T}(\sigma m_1 m_2)$$
$$\cdot\int F_1[-\zeta S-\lambda\xi_2\sigma S]F_1^*[-\zeta S]F_2[\lambda\xi_2\sigma+\zeta]F_2^*[\zeta]d\zeta$$
$$=c\cdot\tilde{T}(\sigma m_1 m_2)\cdot\tau(\sigma). \quad (17)$$

Here the image spectrum consists of an object spectrum, $c\tilde{T}(\sigma m_1 m_2)$, multiplied by the transfer function, for the integral is the autocorrelation of the effective aperture distribution.[22] The integral is the convolution of the four complex pupil functions whose respective arguments are identical except for the scale factor, $-S$. We shall assume that the optics are diffraction limited. The pupil functions are therefore even functions, and the negative sign can be neglected. These functions are rectangular; the width of the influx-optics pupil function will have a value of $2a/S$, and that of the efflux optics will be $2b$. To reduce the elements of the convolution to those of the efflux pupil function and its complex conjugate,[23] it is necessary that

$$a/S\geq b, \quad (18)$$

which may be determined from a consideration of the integrand of Eq. (17). Upon the application of Eq. (12), the inequality reduces to

$$a/z_2\geq b/\xi_1. \quad (19)$$

Because

$$a/z_2=(\mathrm{NA})_{\mathrm{inf}} \quad (20)$$

and

$$b/\xi_1=(\mathrm{NA})_{\mathrm{eff}}, \quad (21)$$

$$\frac{(\mathrm{NA})_{\mathrm{inf}}}{(\mathrm{NA})_{\mathrm{eff}}}\geq 1. \quad (22)$$

When this condition is met,

$$\tilde{I}(\sigma)=c\cdot\tilde{T}(\sigma m_1,m_2)\int F_2[\lambda\xi_2\sigma+\zeta]F_2^*[\zeta]d\zeta, \quad (23)$$

and the system is independent of the influx optics, a well-known result.[24] No sensor aperture in the image plane is technically necessary because the source slit image is reproduced almost perfectly. This ignores the effects of flare light. The only modification lies in the radiance of the image, and this is proportional to the transmittance of the sample: precisely the quantity the equipment is intended to measure! It is not difficult to show that when the reduced slit image is smaller than any detail variation within the sample, the system will manifest the property that the flux transmittance will be linearly proportional to the sample transmittance, regardless of the degree of partial coherence existing at the source slit. Therefore, if this size relationship between sample detail and reduced slit image could be maintained for all sample distributions, no further consideration of this problem would be necessary.

However, there are limitations to this simple relationship. There is a lower size limit to the reduced slit image that is determined by physical optics. When the source slit is small enough to be unresolved (taking the magnification, m_1, into account), the image will be the line impulse response of the influx optics. Its profile will be a one-dimensional representation of the Airy pattern, and hence illumination of the sample will be highly coherent. Further reduction of source slit size or changes of illumination partial coherence will not affect the nature or size of this image. Provided that this image illuminates an area of constant transmittance (much smaller than any transmittance variation in the sample) no problem of system linearity arises.

On the other hand, many samples contain detail variations that are smaller than this image. Because the illumination of such samples is highly coherent, the efflux optics will image the distributions nonlinearly. It is therefore essential to consider illumination in which the degree of partial coherence can be related to the (assumed) incoherently illuminated source slit. This condition will be achieved when the slit image is purely a geometric reduction of the slit, determined by the magnification, m_1; i.e., the slit image must be resolved by the influx optics. An alternative statement of this condition is that the slit reduction must occur for dimensions in which the magnifications have geometrical meaning.

For these circumstances, then, the reduced slit image does not illuminate an area of constant transmittance, and a more general analytical relation is necessary to describe the imagery.

Relatively Large Slit Image

The other incoherent limiting case occurs when the slit image is very large compared to the detail variation in the sample. This does not necessarily call for a large source slit, and in fact it can be exactly the size that it might have been for the case of constant sample transmittance. It is the size of the reduced slit image relative to the variations in transmittance of the sample within the illuminated region that is significant. The flux transmittance that will now be measured will be an average over this region, and transmittances within the region must be measured with the aid of a smaller sampling or sensor aperture located in the image plane.

For this condition, the slit spectrum may be approximated by

$$\tilde{T}[\sigma m_1 m_2-m_1(\eta-\eta')]\simeq\delta[\sigma m_1 m_2-m_1(\eta-\eta')]. \quad (24)$$

The relation is now substituted in Eq. (14), and integration over η' produces

$$\tilde{I}(\sigma)=\int F_2[\lambda\xi_2\sigma+\zeta]F_2^*[\zeta]\cdot\left\{\int F_1[-\zeta S-\lambda\xi_2\sigma S-\lambda z_2\eta]\cdot F_1^*[-\zeta S-\lambda\xi_2\sigma S-\lambda z_2\eta]\cdot\tilde{\varphi}(\eta)\tilde{\varphi}^*(\sigma m_2-\eta)d\eta\right\}d\zeta. \quad (25)$$

The arguments of the complex pupil functions are identical. Because these functions are rectangular by hypothesis, the ensuing quadrature does not affect shape or width. It is necessary, however, that the resulting function be wide enough to encompass the convolution of the complex sample spectrum with its shifted conjugate. By definition, the sample spectrum is band limited to $\eta_0/2$. Therefore, the necessary condition that the inner integral reduce to the convolution of the complex object spectra is

$$\eta_0/2+\sigma m_2\leq a/\lambda z_2-\xi_2\sigma S/z_2-\zeta S/\lambda z_2, \quad (26)$$

which may be determined from an examination of the pertinent integrand of Eq. (25). For the maximum extent of the efflux pupil function, ζ can be replaced by its maximum value, b, as before. Further, with a diffraction-limited lens, the resolution limit is a function of numerical aperture. For the efflux optics, the resolution limit will therefore be

$$\sigma_0=2(\mathrm{NA})_{\mathrm{eff}}/\lambda. \quad (27)$$

Upon application of Eqs. (2), (12), (20), and (21), the necessary condition reduces to

$$\eta_0/2+b/\lambda\xi_1\leq a/\lambda z_2 \quad (28)$$

and

$$(\mathrm{NA})_{\mathrm{inf}}/(\mathrm{NA})_{\mathrm{eff}}\geq 1+\eta_0/\sigma_0. \quad (29)$$

When this condition is met, Eq. (4) is invoked, and the convolution theorem applied to Eq. (25). The image spectrum becomes

$$\tilde{I}(\sigma)=\tilde{\psi}(-\sigma m_2)\int F_2[\lambda\xi_2\sigma+\zeta]F_2^*[\zeta]d\zeta. \quad (30)$$

This is again linear, but now the sample is imaged onto the x plane, and a sensor aperture is clearly necessary for detail-transmittance measurement. The system has operated as a projector.

The essential difference between the two extreme cases is the location or definition of what can be called the sampling aperture. In the first case, the reduced image of the slit was imaged with sufficient fidelity by the system to act as the photometric or sampling aperture. Nothing was needed (in principle) in the image plane to restrict unwanted light. In the case just considered, it would be necessary to insert an aperture in the image plane small enough to block out the unwanted portions of the image, so that a measurement over a region of constant transmittance could be made.

General Linearity Conditions

The two extreme cases have been examined in detail because they provide insight into the solution of the general case that lies between the two. The essential problem can be formulated easily from Eqs. (14), (17), and (30). The double integral that defines the interactive components is

$$\tilde{I}_1(\zeta,\sigma)=\int\int F_1[-\zeta S-\lambda\xi_2\sigma S-\lambda z_2\eta]F_1^*[-\zeta S-\lambda z_2\eta']\cdot\tilde{\varphi}(\eta)\tilde{\varphi}^*(\eta')\tilde{T}[\sigma m_1m_2-m_1(\eta-\eta')]d\eta d\eta'; \quad (31)$$

this is contained within an integral that is a function of the efflux optics alone,

$$\tilde{I}(\sigma)=\int\tilde{I}(\zeta,\sigma)F_2[\lambda\xi_2\sigma+\zeta]F_2^*[\zeta]d\zeta. \quad (32)$$

Now, when Eq. (31) reduces to

$$\tilde{I}_0(\sigma)=\int\int\tilde{\varphi}(\eta)\tilde{\varphi}^*(\eta')\tilde{T}[\sigma m_1m_2-m_1(\eta-\eta')]d\eta d\eta' \quad (33)$$

(i.e., the sample spectrum no longer depends on the influx optics), the system can be regarded as linear according to the criterion of Eq. (8).

At this point, a clarification of the designation, object, is in order. Because the μ, w, and x planes are conjugate, the microdensitometer images the source slit and sample distribution together in the x plane. This dual aspect of the system object is shown in the double convolution of Eq. (33). When the reduced slit image is much smaller than the sample details, we have shown that the slit characteristics predominate in the image. When the slit image is larger than the detail transmittance variations of the sample, the sample characteristics predominate. In all cases, the light flux in the image plane is related to the transmittance of the sample, but the microdensitometric object is not a clearly defined entity located in a single plane. Although the desired goal of microdensitometer measurements is the quantitative evaluation of the sample transmittance, the analysis must contend with this ambiguous specification of the object. There need be no confusion in practice; the sampling or sensor aperture in the image plane eliminates the ambiguities of transmittance measurement.

The conditions under which Eq. (31) generally reduces to Eq. (33) are perhaps best displayed graphically. The influx pupil function has a physical width $2a$ and will be considered diffraction limited. The integrand of Eq. (31) is sketched in Fig. 2, with the pupil functions scaled by the appropriate reduced coordinates. The

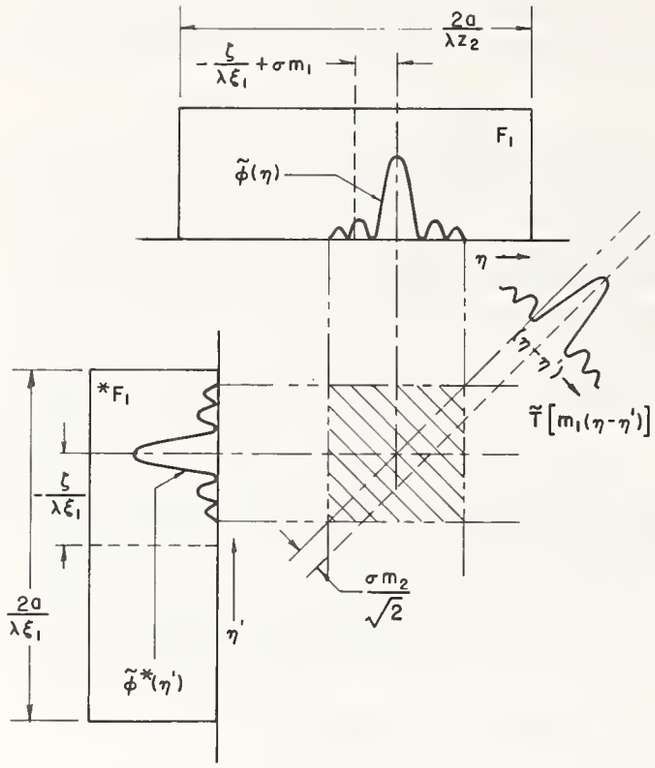


FIG. 2. Sketch of the integrand of Eq. (31).

cross-hatched area indicates the region in which the complete object spectrum is contained, with the amplitude spectra, $\tilde{\varphi}(\eta)$ and $\tilde{\varphi}^*(\eta)$, band limited to $\eta_0/2$.

The slit spectrum, a function of coordinate differences, is oriented diagonally with respect to the two coordinate systems and will sweep across the cross-hatched area during the integrations. The particular value of the convolved functions depends on the displacement, $(\sigma m_2)/\sqrt{2}$, at a given frequency, and drops to zero when the displacement exceeds the cutoff frequency of the efflux optics. From Fig. 2, it is clear that the conditions for which the sample spectra are independent of the complex influx pupil functions are

$$a/\lambda z_2 \geq \eta_0/2+\zeta/\lambda\xi_1-\sigma m_2 \tag{34}$$

and

$$a/\lambda z_2 \geq \eta_0'/2+\zeta/\lambda\xi_1. \tag{35}$$

The variables ζ and σ are in the image space and are technically unrestricted in range. However, ζ defines the extent of the pupil functions and σ is thereby also restricted. The efflux pupil function has a width of $2b$. The optics are assumed to be diffraction limited and the limiting or cutoff frequency of the efflux optics is defined by Eq. (27).

Because σ is an image frequency, σm_2 is the image frequency scaled to the sample space in which η and η' are contained. Because the sample is band limited to $\eta_0/2$, the maximum frequency in image space will be related to this through

$$(\sigma m_2)_{\max}=\eta_0/2. \tag{36}$$

The condition of Eq. (34) therefore reduces to

$$a/\lambda z_2 \geq b/\lambda\xi_1, \tag{37}$$

whereas Eq. (35) has already been developed for the case of a relatively large illuminating slit image. These relations are best expressed in terms of numerical apertures.

The ratio of the numerical aperture of the influx optics to that of the efflux optics has been assigned various symbols in the literature. Here the designation, ϵ, originally assigned by Kinzly[6] will be used. When Eqs. (20), (21), and (27) are employed, the conditions for linearity become

$$\epsilon \geq 1 \tag{38}$$

and

$$\epsilon \geqq 1+\eta_0/\sigma_0, \tag{39}$$

results cited previously. The first condition is clearly weaker than the second, which contains it as a special case. Therefore, the most-general condition for linearity is given by Eq. (39).

The bearing of this equation on the specific nonlinearity of concern is important. When this condition is not met, the outer portions of one side of the complex sample spectrum are truncated. When the spectrum is subsequently passed through the efflux optics and recombined, the resultant image is not a linear, point-by-point reconstruction of the sample distribution. Then, even when a sampling aperture is placed in the image plane to effect a local measurement of flux transmittance, the output will not be a direct measurement of the corresponding sample transmittance. This, in company with the problem of specularity associated with all microdensitometric measurements, is the fundamental problem of microdensitometry. Indeed, all optical systems with this arrangement of components, illumination, and geometry are similarly susceptible to such nonlinear image formation.

Incoherent illumination is thus no guarantee of instrument linearity. On the other hand, this linear condition was predicated on an incoherently illuminated source slit. But it is physically impossible to realize full incoherence, and therefore it is necessary to investigate the conditions on incident illumination for which the above condition remains valid. To accomplish this, the mutual intensity of Eq. (11) must be generalized.

Generalized Mutual Intensity

The illumination of microdensitometers commonly derives from an extended incandescent source. The source slit is then illuminated through an optical system that ordinarily does not locate the source image in the plane of the source slit. This assures even illumination at an adequate irradiance. In many cases, the mutual intensity of this illumination can be reasonably approximated by the Van Cittert–Zernike theorem.[25] This provides a description of the mutual intensity as a

function of coordinate differences

$$\Gamma_0(\mu_1,\mu_2)=\Gamma_0(\mu_1-\mu_2). \tag{40}$$

This form of generalization is not only useful analytically, but is also measurable, in principle, in real systems. Thus, although this may not be the broadest possible generalization, it characterizes many practical systems and is subject to quantitative observation. When Eq. (40) is substituted in Eq. (11) and the inner integrals are evaluated over μ_1 and μ_2, the image spectrum becomes

$$\begin{aligned}\tilde{I}(\sigma)=\int \tilde{\Gamma}_0\left[\sigma m_1 m_2-\frac{\zeta m_1}{\lambda\xi_1}-m_1\eta-\beta\right]\\ \cdot\tilde{i}(\beta)\tilde{i}^*[\beta+m_1(\eta-\eta')-\sigma m_1 m_2]\\ \cdot\tilde{\varphi}(\eta)\tilde{\varphi}^*(\eta')F_2[\lambda\xi_2\sigma+\zeta]F_2^*[\zeta]\\ \cdot F_1[-\zeta S-\lambda\xi_2\sigma S-\lambda z_2\eta]F_1^*[-\zeta S-\lambda z_2\eta']\\ \cdot d\beta d\eta d\eta' d\zeta.\end{aligned} \tag{41}$$

The exact analytical form of $\tilde{\Gamma}_0$ can easily be ascertained. Because the system illumination originated in an incoherent source (a slit, for the one-dimensional case under consideration), and the source-slit illumination, by the Van Cittert–Zernike theorem, is the Fourier transform of this, the mutual intensity is necessarily of the form $\operatorname{sinc}A$. Its transform, $\tilde{\Gamma}_0$, is therefore rectangular. For all states of partial coherence, $\tilde{\Gamma}_0$ is a rectangular function. In the incoherent limit, it will have infinite width; for the coherent limit, it will be a delta function.

In practice, it will not be necessary to have an infinite width for $\tilde{\Gamma}_0$. To be effectively incoherent, the width need only be large enough so that the amplitude spectrum of the source slit can be autocorrelated over β in Eq. (41). However, a fundamental problem now arises. In a previous consideration of the band limitation on the sample, the low-pass nature of the photographic process provided the necessary physical constraints. The slit-radiance spectrum is an analytic function that will not vanish identically over a finite interval,[26] and its amplitude components necessarily possess related properties. However, since most of the energy in these functions is contained within a small distance of their center, and because some spreading due to truncation can be compensated by an aperture in the image plane, an approximate band limiting can be imposed that, while arbitrary, will subsequently be shown to have some validity.

If the actual width of the source slit is denoted by $1/\beta_0$, the distance between the two first-order zeroes of its transform is $2\beta_0$. Since the location of subsequent zeroes is periodic, at integral multiples of β_0, the width of the truncated spectrum can be denoted by $2N\beta_0$, where N is an integer unspecified at this time. These relations are sketched in Fig. 3.

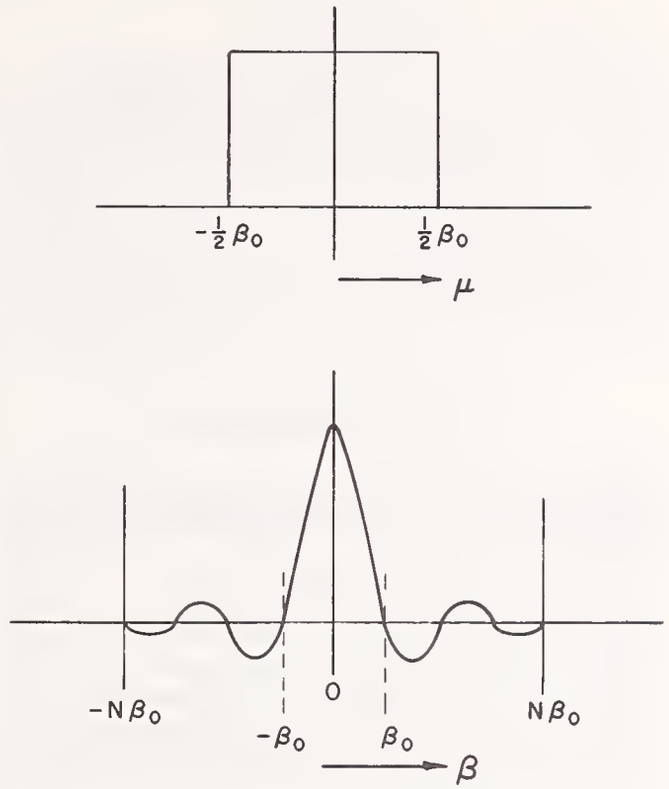


FIG. 3. Sketch of the source slit and its amplitude spectrum, showing the reciprocal dimensional relations. μ and β are the conjugate Fourier variables. Top: Source-slit transmittance; Bottom: Corresponding spectrum.

The rectangular $\tilde{\Gamma}_0$ has a width of $2C$. Its transform has a width of $1/2C$ between its center and first zero. This width is usually taken as a measure of the coherence interval.[1] Thus, the ratio of the slit width to coherence interval is given by

$$\begin{aligned}\rho&=(1/\beta_0)\div(1/2C)\\ &=2C/\beta_0\\ &=\text{slit width/coherence interval.}\end{aligned} \tag{42}$$

This parameter is equivalent to $2R^{-1}$ of Ref. 1.

The integral over β may now be considered. From Eq. (41),

$$\tilde{I}(\sigma,\zeta,\eta)=\int\tilde{\Gamma}_0[a_0-\beta]\tilde{i}(\beta)\tilde{i}^*(\beta+a_1)d\beta_1, \tag{43}$$

where

$$\begin{aligned}a_0&=\sigma m_1 m_2-\zeta m_1/\lambda\xi_1-m_1\eta,\\ a_1&=m_1(\eta-\eta')-\sigma m_1 m_2.\end{aligned} \tag{44}$$

Figure 4 sketches the integrand of Eq. (43), to establish the basis for the condition under which the integral is independent of $\tilde{\Gamma}_0$. This condition is

$$-C+a_0\leq -a_1-N\beta_0. \tag{45}$$

From this and Eqs. (44), we may derive the expression

$$C\geqq N\beta_0-m_1\eta'-\zeta m_1/\lambda\xi_1. \tag{46}$$

For maximum sample frequency, as before, the ampli-

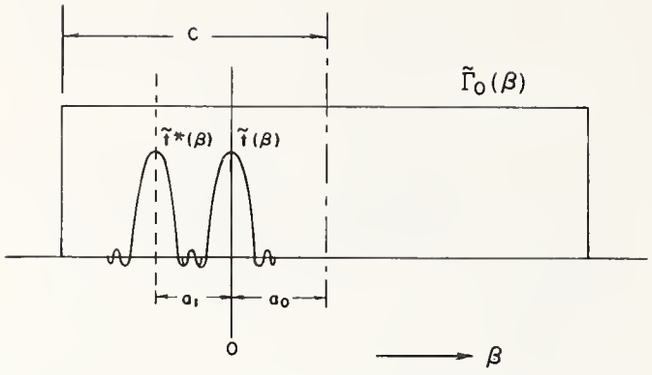


FIG. 4. Sketch of the integrand of Eq. (43).

tude spectrum is limited to $\eta_0/2$. Application of this and Eqs. (1), (27), and (42) to the inequality of Eq. (46) produces

$$\rho \geqq 2N+(|m_1|/\beta_0)(\eta_0+\sigma_0), \tag{47}$$

where the negative sign attached to m_1 to denote inversion during imaging has been dropped.

The quantity $|m_1|/\beta_0$ is the reduced source slit, imaged on the sample. The equation shows a requirement for a larger value of ρ as either system resolution or sample-frequency limit, or both, increases. This is confirmed by experience with normal microdensitometer operation.

Reference 1 shows that system response is essentially incoherent for conditions equivalent to $\rho>4$. It is therefore possible to assign a tentative value of 2 to N and write

$$\rho \geqq 4+(|m_1|/\beta_0)(\eta_0+\sigma_0). \tag{48}$$

This is the minimum condition for which the system may be considered effectively incoherent, and the prerequisite for valid application of Eq. (39). When the sample frequency (η_0) and resolution limit (σ_0) are approximately equal, and the reciprocal of the reduced slit image is on the order of the resolution limit,

$$\rho \geq 6.$$

This is similar to the inequality defining Mielenz's incoherent slit image,[19] and serves to substantiate the choice of 2 for N.

DISCUSSION

The relation of Eq. (39) implies a maximum value of 2 for ϵ, because $\eta_0>\sigma_0$ has no meaning. However, the problem of flare light increases when the numerical apertures are mismatched in the fashion indicated. Further, the numerical apertures of microscope objectives are limited in range. It seems clear, then, that the linearity of the microdensitometer is also a function of the available microscope objectives and the degrading effects of flare light. These limitations make it virtually certain that there will always be sample frequencies for which the system will be nonlinear.

Practical values of ϵ, determined from the references exceed 1.0 for linearity. Reference 11 indicates maximum two-point resolution for $\epsilon=1.5$, and Ref. 12 employs objectives that provide a ϵ of 1.67. All sources show nonlinear behavior for $\epsilon<1.0$.

The condition for effective incoherence is limited by the range of applicability of the Van Cittert–Zernike theorem for a particular microdensitometer. In the absence of specific geometries, it remains a useful approximation. The assigned value of 2 for N in Eq. (48) is arguable, yet a value of, say, 4 appears far too conservative. The value assigned probably represents the lowest practical value. Experimentation will be necessary to establish its correct value.

The purpose of the influx optics is to provide illumination for the sample. When the reduced slit image is too small (the impulse response of the influx optics), the illumination is coherent over its width, and except for cases of constant sample transmittance, results in nonlinear imaging by the efflux optics. When it is too large, the effects of flare light become pronounced. Therefore, practical operation of the microdensitometer must strike a balance between the two conditions.

Ordinarily, the line impulse response of a lens can be obtained as an image when the object slit (accounting for magnification) is less than one-fourth the width of the Airy disk. With diffraction-limited lenses, this width (in micrometers) is approximated by

$$W \cong 1/2(\mathrm{NA}), \tag{49}$$

for $\lambda=500$ nm. Conversely, we can assume that when the reduced slit image is at least four times as wide as the Airy disk, the geometric slit identity has been preserved. This was shown to be generally necessary in the discussion following Eq. (23). Therefore, the recommended minimum reduced slit-image width will be given by

$$(|m_1|/\beta_0)_w \cong 4W \cong 2/(\mathrm{NA})_{\mathrm{inf}}. \tag{50}$$

The various relations developed in this analysis can now be summarized for some microscope objectives typically used in microdensitometers. These are shown in Table I for three specific lens pairs. Further permutations of the four numerical apertures to give higher values of ϵ are not calculated, because experience has shown that flare light becomes objectionable for ϵ greater than 2.

TABLE I. Linearity and coherence parameters for three typical microdensitometer lens pairs ($\lambda=500$ nm).

$(\mathrm{NA})_{\mathrm{inf}}$	$(\mathrm{NA})_{\mathrm{eff}}$	ϵ	σ_0 (mm^{-1})	η_0 (mm^{-1})	$(\lvert m_1\rvert/\beta_0)_w$ (μm)	ρ
0.65	0.40	1.625	1600	1000	3	11.8
0.40	0.25	1.60	1000	600	5	12.0
0.25	0.16	1.56	640	360	8	12.0

TABLE II. Maximum allowable coherence intervals for the influx optics of Table I ($\lambda = 500$ nm).

$(NA)_{inf}$	Magnification (m_1)	Minimum source slit width ($1/\beta_0$) (μm)	Maximum allowable coherence interval ($1/2C$) (μm)
0.65	1/40	120	10.2
0.40	1/20	100	8.3
0.25	1/10	80	6.7

Sample frequencies cannot exceed the cited values of η_0 for linear operation. When the sample frequency range is known, this tabulation can be used to select appropriate lens pairs. A sampling aperture in the image (sensor) plane is necessary, and is assumed to be of sufficiently small width (accounting for efflux-optics magnification) to permit accurate measurements of flux transmittance. Because it undergoes a convolution with the image luminance distribution when scanning, it is necessarily a component of the dynamic system transfer function. This effect is not included in Table I and does not affect the fundamental considerations of this paper.

Some microdensitometers use only single lens groups (e.g., microscope objectives) for influx and efflux optics. Others combine these with eyepieces to alter the magnification. This has a bearing on the magnitude of the required coherence interval, because it allows larger source slit widths. These interrelated parameters are shown in Tables II and III. Here, the typical magnifications of the microscope objectives listed for the influx optics of Table I, and the recommended minimum reduced slit-image widths, are used to calculate the actual minimum source slit size. The appropriate values of ρ are then used to determine the maximum allowable coherence intervals. Table III adds the effect of a 10× eyepiece used for further slit-image reduction.

The required coherence intervals of Table II are marginally realizable with most microdensitometers, whereas those of Table III can be routinely obtained with most sources and slit-illuminating systems. On the basis of these calculations, consistent with the reduction of flare light, the influx optics should employ as large a reduction as practical.

TABLE III. Maximum allowable coherence intervals for the influx optics of Table I, with the addition of a 10× eyepiece ($\lambda = 500$ nm).

$(NA)_{inf}$	Magnification (m_1)	Minimum source slit width ($1/\beta_0$) (μm)	Maximum allowable coherence interval ($1/2C$) (μm)
0.65	1/400	1200	102
0.40	1/200	1000	83
0.25	1/100	800	67

CONCLUSIONS

The requirements for linearity and effective incoherence have been established, and their range of values estimated. When the microdensitometer is used within these constraints, linear, reproducible operation is assured. When these conditions are met, an unambiguous transfer function can be defined and measured. The design of high-numerical-aperture microscope objectives to eliminate or reduce flare light is called for, so that higher values of ϵ may be achieved without loss of image contrast.

ACKNOWLEDGMENTS

The author has benefited immensely from conversations and correspondence with R. Kinzly of Cornell Aeronautical Laboratories who has considered this problem in the spatial domain. Thanks are also due to D. N. Grimes of the National Bureau of Standards for many helpful suggestions made during the course of the investigations. The author is indebted to both referees for suggestions that have served to clarify the underlying concepts and have aided in a more rigorous presentation of the solution.

REFERENCES

* Presented at the 1970 Fall Meeting of the Optical Society of America [J. Opt. Soc. Am. 60, 1551A (1970)].

[1] R. Becherer and G. Parrent, J. Opt. Soc. Am. 57, 1479 (1967).

[2] R. Swing and J. Clay, J. Opt. Soc. Am. 57, 1180 (1967).

[3] D. Canals-Frau and M. Rousseau, Opt. Acta 5, 15 (1958).

[4] S. Slansky, Rev. Opt. 39, 555 (1960).

[5] M. De and S. C. Som, Opt. Acta 4, 17 (1962).

[6] R. Kinzly, J. Opt. Soc. Am. 55, 1002 (1965).

[7] R. Kinzly, J. Opt. Soc. Am. 56, 526 (1966).

[8] R. Kinzly, J. Opt. Soc. Am. 56, 9 (1966).

[9] W. Charman, J. Opt. Soc. Am. 53, 410 (1963).

[10] W. Charman, J. Opt. Soc. Am. 53, 415 (1963).

[11] H. H. Hopkins and P. Barham, Proc. Phys. Soc. (London) 63, 72 (1950).

[12] D. Galburt, R. A. Jones, and J. Bossung, Phot. Sci. Eng. 13, 205 (1969).

[13] A nomenclature for the elements of a microdensitometer has not been standardized, although the need clearly exists. This paper employs a system proposed by C. S. McCamy and is based on the considerations of Phot. Sci. Eng. 10, 314 (1966).

[14] M. J. Beran and G. B. Parrent, Jr., *Theory of Partial Coherence* (Prentice-Hall, Englewood Cliffs, N. J., 1964), p. 53.

[15] E. L. O'Neill, *Introduction to Statistical Optics* (Addison-Wesley, Reading, Mass., 1963), p. 126.

[16] Multiple integrals will be made explicit only when necessary. The number implied corresponds to the number of differentials. Unless otherwise specified, all integrals are to be evaluated between $-\infty$ and ∞.

[17] M. Born and E. Wolf, *Principles of Optics, 2nd ed.* (Pergamon, New York, 1964), p. 526.

[18] R. Swing and D. Rooney, J. Opt. Soc. Am. 58, 629 (1968).

[19] K. Mielenz, J. Opt. Soc. Am. 57, 66 (1967).

[20] At extremely large slit sizes, assuming no flare light, the principal effect is the introduction of a quadratic phase error across the sample.

[21] Reference 15, p. 57.

[22] J. DeVelis and G. Parrent, J. Opt. Soc. Am. 57, 1486 (1967).

[23] Reference 22, discussion preceding (their) Eq. (24).

[24] Reference 17, p. 522.

[25] Reference 17, p. 508.

[26] R. Barakat, J. Opt. Soc. Am. 54, 920 (1964).

Reprinted from JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, Vol. 61, No. 9, 1263, September 1971
Printed in U. S. A.

Linear Microdensitometry

D. N. GRIMES
National Bureau of Standards, Washington, D. C. 20234
(Received 1 April 1971)
INDEX HEADINGS: Coherence; Microdensitometer.

The optical design of conventional microdensitometers[1] is similar to that of a microscope; that is, the sample is illuminated by a condenser system (influx optics) that focuses a slit on the sample, which is then imaged (by the efflux optics) at some magnification. The image falls upon a slit that, as the sample is moved, effectively scans the image. It has been shown by Swing[2] and Kinzly[3] that such microdensitometers operate in a nonlinear mode at high spatial frequencies. This is unavoidable, owing to the optical design of these instruments.

Swing[2] has shown theoretically that, for a conventional microdensitometer to operate in a linear mode out to the cutoff frequency of the efflux optics, the source slit must be essentially incoherent; that is, the ratio of slit width to coherence interval must be substantially greater than 1.0. It is also necessary that the ratio of numerical apertures of influx to efflux optics satisfy the condition

$$\epsilon \geq 1+(\eta_0/\sigma_0),$$

where η_0 is the sample frequency and σ_0 is the cutoff frequency of the efflux optics. These two conditions arise from the demand that the imaging system (efflux optics) be a linear system in (power) transmittance.

The microdensitometer system proposed here and shown in Fig. 1 eliminates the imaging portion of the system. Instead, the sample is illuminated by the focused image of the slit and approximately all of the transmitted light is collected on the other side. Hence, the system essentially measures singly diffuse density rather than specular density as is currently done.

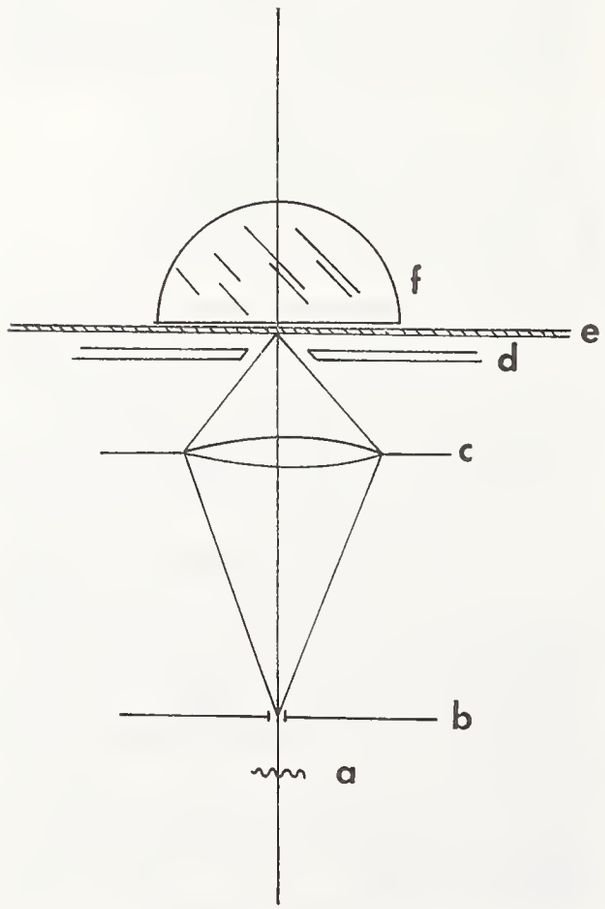


FIG. 1. Linear microdensitometer. (a) source, (b) slit, (c) high-quality condenser, (d) baffles, (e) sample, and (f) light collector.

Mathematically, the (power) transmittance measured by the system for a slit aperture may be described in one dimension by

$$I=\int \Gamma_0(\mu_1,\mu_2)t(\mu_1)t^*(\mu_2)F(\alpha_1)F^*(\alpha_2) \times\exp\left(\frac{ik(\alpha_1-\mu_1)^2}{2z_1}\right)\exp\left(\frac{-ik(\alpha_2-\mu_2)^2}{2z_1}\right)\exp\left(\frac{ik(\alpha_2^2-\alpha_1^2)}{2f}\right) \times\exp\left(\frac{ik(y-\alpha_1)^2}{2z_2}\right)\exp\left(\frac{-ik(y-\alpha_2)^2}{2z_2}\right) \times|\phi(y)|^2 d\mu_1 d\mu_2 d\alpha_1 d\alpha_2 dy, \quad (1)$$

where $\Gamma_0(\mu_1,\mu_2)$ is the coherence at the slit, $t(\mu)$ is the amplitude transmittance of the slit, $F(\alpha)$ is the pupil function of the influx optics, and $\phi(y)$ is the amplitude transmittance of the sample. The coordinates μ, α, and y refer to the planes of the object, influx optics, and sample, respectively, and z_1 and z_2 are the object and image conjugates of the influx optics. This equation may be written in the form

$$I=\int |\phi(y)|^2 S(y)dy, \quad (2)$$

where after using the paraxial approximation and simplifying, $S(y)$ is given by

$$S(y)=\int \Gamma_0(\mu_1,\mu_2)t(\mu_1)t^*(\mu_2)F(\alpha_1)F^*(\alpha_2)\exp\left(\frac{ik(\alpha_2\mu_2-\alpha_1\mu_1)}{z_1}\right) \times\exp\left(\frac{iky(\alpha_2-\alpha_1)}{z_2}\right)d\mu_1 d\mu_2 d\alpha_1 d\alpha_2.$$

As the sample is moved a distance x, the output becomes

$$I(x)=\int |\phi(y-x)|^2 S(y)dy, \quad (3)$$

which is essentially a linear system with impulse response $S(y)$.

If the source is effectively a delta function $\delta(\mu)$, then the impulse response reduces to

$$S(y)=\left|\int F(\alpha)\exp\left(\frac{iky\alpha}{z_2}\right)d\alpha\right|^2 \quad (4)$$

and the system transfer function becomes that of the influx optics. One of the advantages of scanning with the impulse response and collecting all of the transmitted light is that the system is independent of the coherence of the source. Hence, there is no restriction on the type of light source used, other than the spectral requirements of the user and, in fact, a laser illuminating a narrow slit or pinhole could be used in cases where a single wavelength was of interest. Because the system transfer function becomes that of the influx optics, a high-quality condenser would be desirable. Because of this, the resulting high numerical aperture may introduce flare light. It is therefore necessary that baffles be introduced, as shown, and that the barrel of the condenser be blackened. Focusing of the slit image on the sample requires that the collecting system be replaced initially by a viewing system focused on the sample, which may then be replaced by the collector during scanning.

In summary, the advantages of this system are (1) the output is independent of the partial coherence of the source, (2) fewer focusing and alignment problems, (3) improved over-all system transfer function because there is no sampling slit, and (4) measurement of singly diffuse density makes interlaboratory comparisons easier.

[1] J. H. Altman and K. F. Stultz, Rev. Sci. Instr. **27**, 1033 (1956).
[2] R. E. Swing, J. Opt. Soc. Am. **60**, 1551A (1970).
[3] R. E. Kinzly, J. Opt. Soc. Am. **60**, 1551A (1970).

Reprinted from Journal of the Optical Society of America, Vol. 61, No. 7, 870–876, July 1971
Printed in U. S. A.

Imaging of Tri-Bar Targets and the Theoretical Resolution Limit in Partially Coherent Illumination*

D. N. Grimes
National Bureau of Standards, Washington, D. C. 20234
(Received 30 November 1970)

In viewing systems that employ critical, Kohler, or collimated illumination, the illumination may frequently be characterized by the complex degree of coherence in the form of a first-order Bessel function divided by its argument. The anomalies that occur in the image of a tri-bar target viewed in partially coherent illumination of this form are discussed for a two-dimensional circular diffraction-limited imaging system. The Sparrow criterion, which in this case correlates with experimental measurements of resolving power, is applied to determine the resolution limit. The computed resolution-limit curve for a tri-bar target is shown and compared to the previously published two-point resolution-limit curve for a one-dimensional system and to the limiting values obtained in a two-dimensional system for coherent and incoherent illumination. In the latter case, the full curve is shown as a special case of the application of the Sparrow criterion to the partially coherent imaging equations for a circular two-dimensional system. Experimental confirmation of the calculations is given.

Index Headings: Coherence; Diffraction; Resolution.

Tri-bar targets are used extensively in the testing of optical instruments, although it has been realized for some time that methods employing them are subject to error. Unlike sine-wave targets, they do not contain a single spatial-frequency component. Kelly[1] and Charman[2] have calculated the spatial-frequency spectrum of a tri-bar target together with the visibility curve[2] for a diffraction-limited incoherent imaging system with a circular aperture. However, although a tri-bar target does not yield the correct MTF of an optical system, these targets continue to be used because they are easier to make than sine-wave targets and the resolving power so measured is useful for comparing system responses.

In recent years, tri-bar targets have been used frequently for comparing systems employing partially coherent illumination. In this case, the ambiguities significantly increase. The degree of coherence is seldom reproducible from one instrument or instrument mode to another and care must be taken in extending results obtained under one set of conditions to another. Also, as pointed out by Becherer and Parrent[3] and Swing and Clay,[4] partially coherent imaging systems are inherently nonlinear and measurements made on them are necessarily object dependent. Barakat[5] and Offner and Meiron[6] have pointed out the nonlinear effects of partially coherent imaging with one- and two-bar targets. Little attempt has been made to date to examine images or the theoretical resolution limit of the more commonly used tri-bar target viewed in partially coherent illumination, primarily because of the complexity and length of the calculations.

In dealing with the resolution of such a target, we have to define a criterion for judging the theoretical resolution limit. Several different criteria have been applied to the two-point resolution problem, including the Rayleigh criterion[7] (which presupposes incoherent illumination), the midpoint criterion of Wolf[8] used by Hopkins and Barham,[9] and the Sparrow criterion[10] used by Rojak[11,12] and Barakat.[13,14] Of these criteria, the Sparrow criterion is most easily extended from point images to bar images. It states that two point sources are just resolved if the second derivative of the resultant image illuminance vanishes at the midpoint of the images. Extended to tri-bar objects, we shall define the resolution limit as the value of the normalized bar spacing for which the second derivative vanishes at the center of the target image. It will be shown that this corresponds either to the occurrence of spurious resolution or complete loss of contrast and that the imagery is sufficiently well behaved to apply this criterion without confusion. Furthermore, this criterion correlates with experimental measurements of resolving power.

TRI-BAR IMAGES

The amplitude transmittance of the tri-bar target shown in Fig. 1 is

$$t(x)=\begin{cases}1 & \text{when } |x|<P/4\\ 0 & P/4<|x|<3P/4\\ 1 & 3P/4<|x|<5P/4\\ 0 & \text{elsewhere,}\end{cases} \tag{1}$$

where P is the spatial period and $1/P$ is the fundamental spatial frequency associated with a given target. The

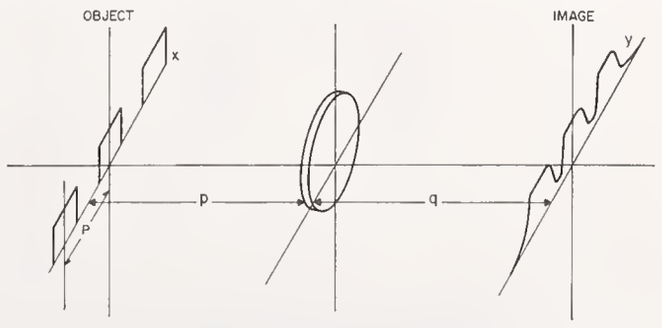


Fig. 1. Coordinates and constants for the image of a tri-bar object.

518-743 O - 73 - 41

general imaging equation in one dimension for a spatially stationary system[12] is

$$\Gamma_{\rm im}(y_1,y_2)=\int_{\rm object}\int\Gamma_{\rm ob}(x_1,x_2)K\left(\frac{y_1}{q}-\frac{x_1}{p}\right)\times K^*\left(\frac{y_2}{q}-\frac{x_2}{p}\right)dx_1dx_2, \quad (2)$$

where $K(y/q-x/p)$ is the amplitude impulse response of the imaging system, $\Gamma_{\rm im}$ is the mutual intensity between the points y_1 and y_2 in the image, and $\Gamma_{\rm ob}$ is similarly the mutual intensity between the points x_1 and x_2 in the object. The coordinates are as shown in Fig. 1.

The degree of coherence of the illumination between the points x_1 and x_2 will be taken as

$$\gamma(x_1,x_2)=\{2J_1[\alpha(x_1-x_2)]\}/\alpha(x_1-x_2)=\Lambda_1[\alpha(x_1-x_2)] \quad (3)$$

in which α is a constant that determines the scale of the coherence function and J_1 is the first-order Bessel function. If the Van Cittert–Zernike theorem is applied to a primary incoherent circular source of radius r at the focus of a collimator of focal length F, then $\alpha=\bar{k}r/F$. Quasimonochromatic illumination is assumed, with mean wavenumber $\bar{k}$. Equation (3) for the degree of coherence is applicable to viewing systems that employ critical, Kohler,[8] or collimated illumination. For critical illumination, α is equal to $\bar{k}$ times the numerical aperture of the condenser.

The mutual intensity in the object is then

$$\Gamma_{\rm ob}(x_1,x_2)=\Lambda_1[\alpha(x_1-x_2)]t(x_1)t^*(x_2), \quad (4)$$

where * denotes the complex conjugate. The amplitude impulse response of a two-dimensional diffraction-limited imaging system with aperture diameter $2a$ is

$$K(y/q)=2a[2J_1(\bar{k}ay/q)]/(\bar{k}ay/q)=2a\Lambda_1(\bar{k}ay/q), \quad (5)$$

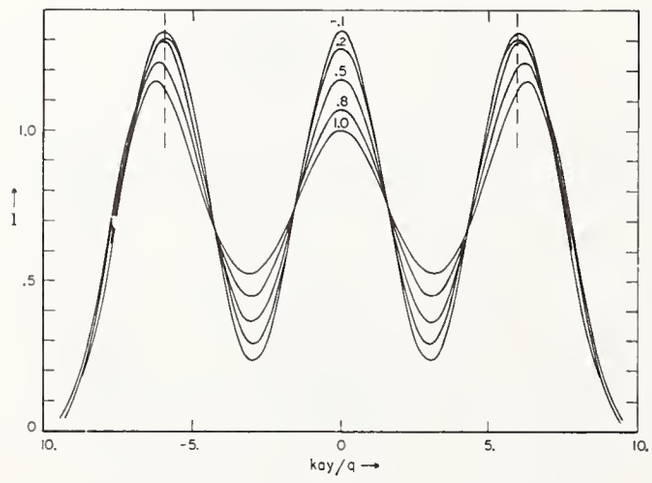


FIG. 2. Image illuminance distribution for δ equal to 6.0 and varying values of $\gamma(\delta)$.

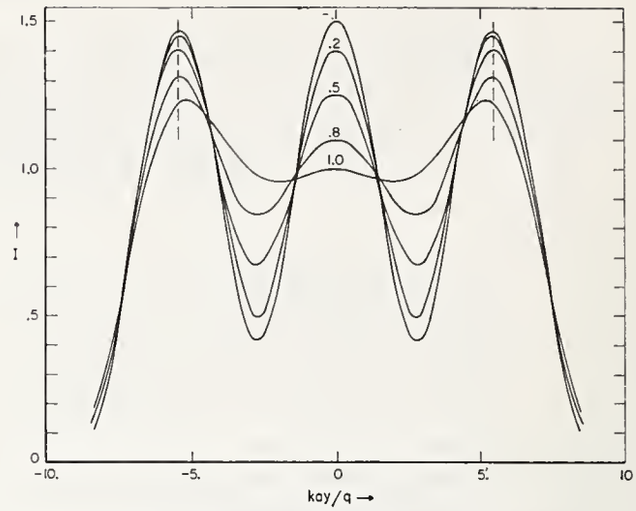


FIG. 3. Image illuminance distribution for δ equal to 5.4 and varying values of $\gamma(\delta)$.

where q is the image distance. The irradiance limit is reached analytically by letting $y_1=y_2=y$. Hence, the illuminance in the image [after substitution of Eqs. (4) and (5) into Eq. (2)] is

$$I(y)=4a^2\int\int\Lambda_1[\alpha(x_1-x_2)]t(x_1)t^*(x_2)\Lambda_1\left[\bar{k}a\left(\frac{y}{q}-\frac{x_1}{p}\right)\right]\times\Lambda_1\left[ka\left(\frac{y}{q}-\frac{x_2}{p}\right)\right]dx_1dx_2. \quad (6)$$

The images of bar targets near the resolution limit of the imaging system were computed from Eq. (6) by use of an Univac 1108 computer. The computation is lengthy. By working with high-contrast targets, and suitable packing of the stored arrays, computation time per image was kept below a few minutes. Some of the images from Eq. (6) are shown in Figs. 2–5. In each case, the image illuminance is plotted as a func-

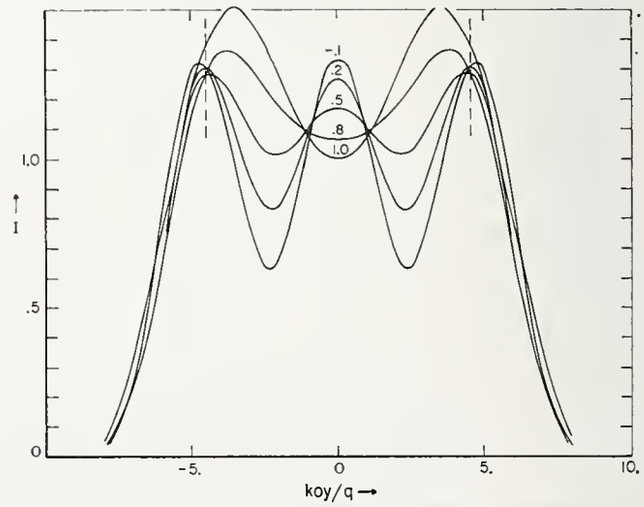


FIG. 4. Image illuminance distribution for δ equal to 4.6 and varying values of $\gamma(\delta)$.

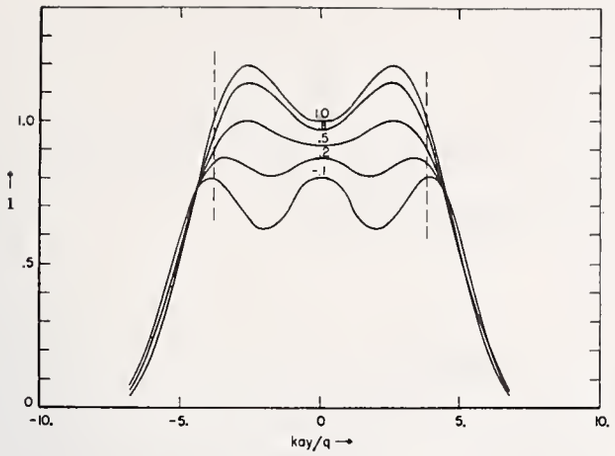


FIG. 5. Image illuminance distribution for δ equal to 3.8 and varying values of $\gamma(\delta)$.

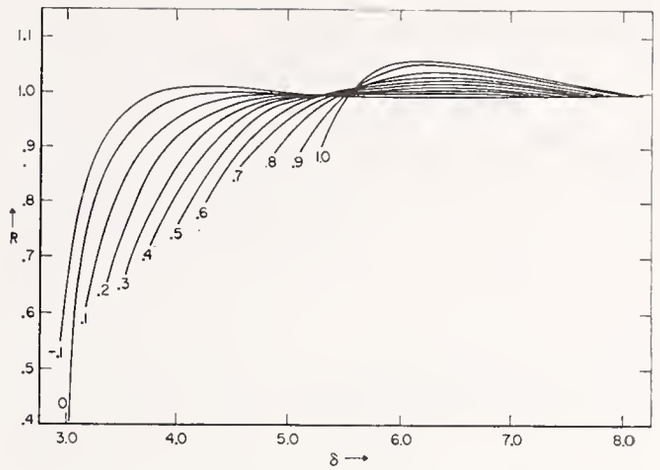


FIG. 6. Ratio, R, of the apparent to real period of a tri-bar object as a function of the real normalized period for varying values of $\gamma(\delta)$.

tion of $\bar{k}ay/q$ for fixed values of the normalized period ($\delta=\bar{k}aP'/q$ where $P'=qP/p$) of the tri-bar target. The curve-to-curve parameter in each case is the scale of the degree of coherence that is denoted by $\gamma(\delta)=\Lambda_1(\alpha\delta)$. This method of plotting allows comparison with two-point image data calculated by Grimes and Thompson.[15] Besides the spurious resolution that occurs in coherent illumination, a change of the apparent spacing of the bars occurs, as in the case of two points. Figure 6 shows a plot of the ratio (R) of the apparent to real period as a function of the normalized period (δ). The curves are terminated for lower values of δ at the value for which the target is no longer resolvable by the Sparrow criterion. In the partially coherent and coherent cases, this corresponds to the value for which the second derivative at the center of the pattern becomes zero as the central maximum inverts to a minimum. In the incoherent limit, the second derivative vanishes as the contrast of the three bars vanishes and a single broad bar remains.

APPLICATION OF THE SPARROW CRITERION

Because of the discrete values of δ for which the images were calculated, it was not possible to determine accurately the resolution limit from the calculated images of the tri-bar targets. However, a rough approximation could be made, which was helpful in checking the following calculations, as well as in limiting their useful range. To apply the mathematical representation of the Sparrow criterion

$$(d^2/dy^2)[I(y)]|_{y=0}=0, \tag{7}$$

it is necessary to take the second derivative of Eq. (6),

$$\frac{d^2}{dy^2}[I(y)]=4a^2\int\int\Lambda_1[\alpha(x_1-x_2)]t(x_1)t^*(x_2)\frac{d^2}{dy^2}$$
$$\times\left\{\Lambda_1\left[\bar{k}a\left(\frac{y}{q}-\frac{x_1}{p}\right)\right]\cdot\Lambda_1\left[ka\left(\frac{y}{q}-\frac{x_2}{p}\right)\right]\right\}dx_1dx_2. \tag{8}$$

Using the following relationships for the derivatives of the functions,

$$(d/dz)\{[J_1(z)]/z\}=-[J_2(z)]/z$$
$$(d^2/dz^2)\{[J_1(z)]/z\}=[J_3(z)]/z-[J_2(z)]/z^2, \tag{9}$$

substituting into Eq. (8), and evaluating at $y=0$ yields

$$\int\int\frac{J_1[\alpha(x_1-x_2)]}{[\alpha(x_1-x_2)]}t(x_1)t^*(x_2)\left(\frac{J_1(kax_1/p)}{(\bar{k}ax_1/p)}\cdot\frac{J_3(\bar{k}ax_2/p)}{(\bar{k}ax_2/p)}+\frac{J_3(\bar{k}ax_1/p)}{(\bar{k}ax_1/p)}\cdot\frac{J_1(kax_2/p)}{(\bar{k}ax_2/p)}-\frac{J_1(kax_1/p)}{(\bar{k}ax_1/p)}\cdot\frac{J_2(\bar{k}ax_2/p)}{(\bar{k}ax_2/p)^2}\right.$$
$$\left.-\frac{J_1(kax_2/p)}{(\bar{k}ax_2/p)}\cdot\frac{J_2(kax_1/p)}{(\bar{k}ax_1/p)^2}+2\frac{J_2(\bar{k}ax_1/p)}{(\bar{k}ax_1/p)}\cdot\frac{J_2(\bar{k}ax_2/p)}{(kax_2/p)}\right)\cdot dx_1dx_2=0. \tag{10}$$

TWO-POINT RESOLUTION LIMIT

Since Eq. (10) does not specify the amplitude transmittance of the target, it may be used to apply the Sparrow criterion to any object including that of two points. Letting

$$t(x)=\delta(x+b)+\delta(x-b) \tag{11}$$

and integrating over x_1 and x_2 yields an equation for the resolution limit of two points in a two-dimensional circular

diffraction-limited imaging system,

$$(1+\gamma)J_1(\bar{k}ab/p)J_3(\bar{k}ab/p)-(1+\gamma)J_1(\bar{k}ab/p)\frac{J_2(\bar{k}ab/p)}{\bar{k}ab/p}+(1-\gamma)[J_2(\bar{k}ab/p)]^2=0, \tag{12}$$

where γ is given by $\Lambda_1(2l\alpha ka/p)$.

In solving this equation for b by computer, it is more convenient to plot the function on the right side of

$$\frac{1-\gamma}{1+\gamma}=\frac{[J_1(\bar{k}ab/p)/(kab/p)]\cdot J_2(\bar{k}ab/p)-J_1(\bar{k}ab/p)J_3(\bar{k}ab/p)}{[J_2(\bar{k}ab/p)]^2}. \tag{13}$$

Then, for a given value of γ, the value of b can be found for which the function takes on the value $(1-\gamma)/(1+\gamma)$. In the limiting cases for $\gamma=0$ (incoherent) and $\gamma=1$ (coherent), we get the values calculated by Barakat.[13,14] The resultant curve for the two-dimensional two-point resolution limit is plotted in Fig. 7.

TRI-BAR RESOLUTION LIMIT

The general solution to Eq. (11) for a tri-bar object was found by substituting Eq. (1) for $t(x)$ and finding the roots of the double integral for varying values of α. Because of the information previously obtained from the images of tri-bar objects of differing period P, the easiest method of solution was to plot the values of the integral in the vicinity of the estimated roots and determine graphically the zero crossing, as shown in Fig. 8. For all but the incoherent limit, there are two roots. They correspond to the inversion of the maximum to a minimum, with the transition from three bars to two bars and then a second inversion as the two bars merge into one broad bar. We have defined the resolution limit as the former, because at that point the target is no longer recognizable as a tri-bar object.

Even with this curve-fitting method, it took approximately 2 h of computer time on the Univac 1108 to do the required computations on a 22 500-member array. The resultant resolution-limit curve is shown in Fig. 7 along with the two-point resolution limits for comparison. As expected, because of the nonlinearity of systems employing partially coherent illumination, the curves are object dependent.

In Fig. 7, the resolution limit is plotted as a function of $\gamma(\delta)$, where δ is either the normalized two-point separation $(2bka/p)$ or the normalized period of the tri-bar object. For two points this suffices. However, for finite-width bars, some functional form of the coherence must be given. Here $\gamma(\delta)=\Lambda_1(\alpha\delta)$. Because this function is multiple valued, the resolution-limit curve is also multiple valued, in this case looping back on itself, as shown. For this reason, it might be more convenient to plot this curve in the form δ_0 vs $\alpha\delta$, where α is $\bar{k}r/F$ for collimated illumination or $\bar{k}$ times the numerical aperture of the condenser for critical illumination. The resulting curve is shown in Fig. 9 together with the two-dimensional two-point resolution limit for the same coherence function. Extended to higher values of $\alpha\delta$, the two-point resolution limit will continue to oscillate while the tri-bar limit remains constant.

From this latter curve it is also possible to derive the relationship of the resolution limit to the ratio of the

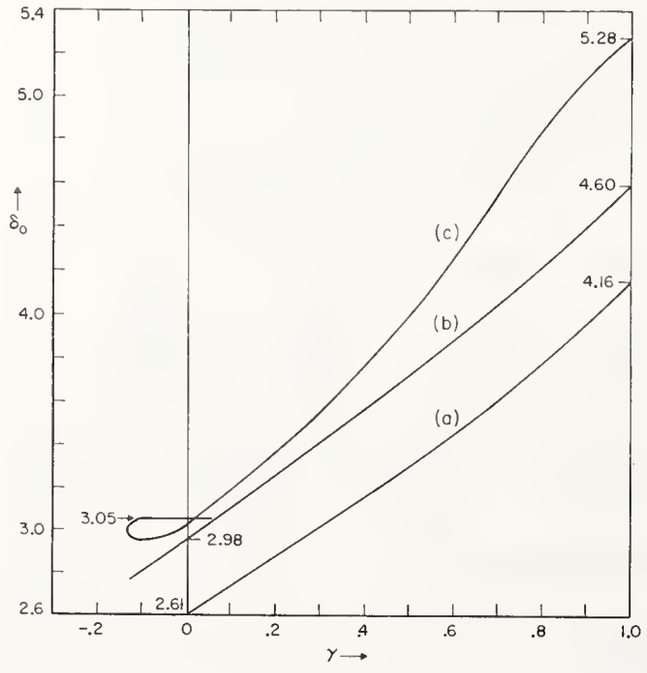


Fig. 7. (a) Two-point resolution limit for a one-dimensional imaging system, (b) two-point resolution limit for a two-dimensional circular imaging system, and (c) tri-bar resolution limit for a two-dimensional circular imaging system.

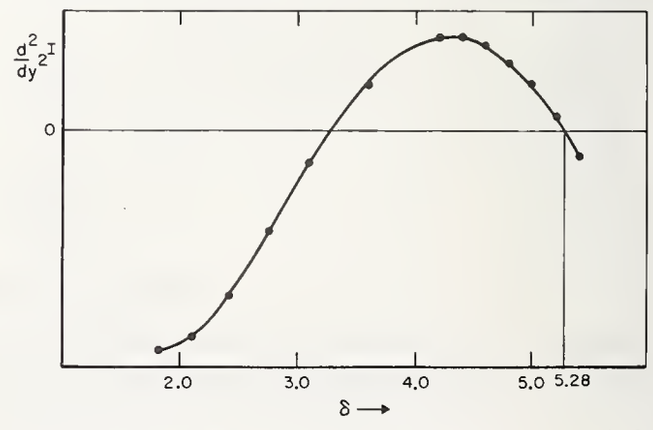


Fig. 8. Illustrating the method of finding the roots of the second derivative of the image illuminance to determine the resolution limit of a tri-bar target using the Sparrow criterion.

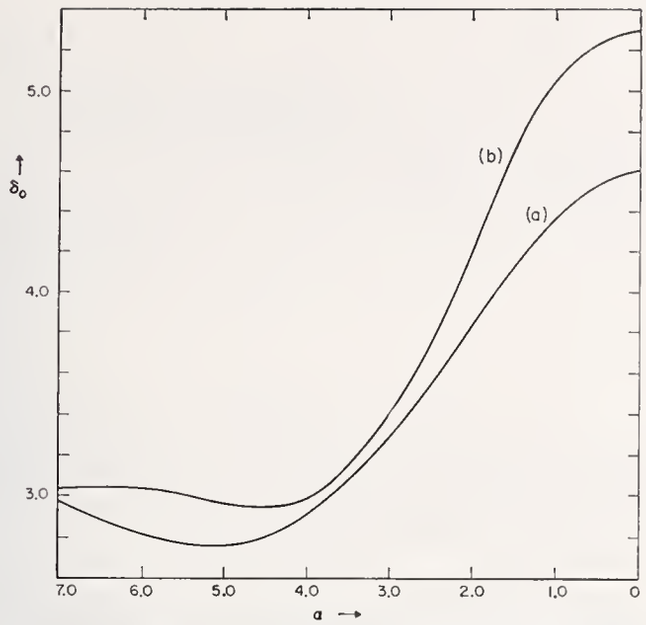


FIG. 9. (a) Two-point resolution limit and (b) tri-bar resolution limit as a function of the argument of the coherence function $\Lambda_1(\alpha\delta)$.

numerical apertures of the condenser and objective for a microscope viewing system. The curves derived are shown in Fig. 10 for differing values of the numerical aperture of the condenser. The shaded areas under the curves correspond to regions where partial coherence introduces significant distortion in the image of the tri-bar object as well as a change in the resolution limit. The cross-hatched areas correspond to the region where the greatest changes of the resolution limit occur, whereas the dotted areas correspond to the region where changes on the order of a few percent or less occur.

It is also possible, from the calculated-image data (Figs. 2–5), to plot visibility curves for varying values of $\gamma(\delta)$ as functions of the fundamental frequency associated with the tri-bar target.

Because in partially coherent illumination, the contrast of the center bar differs from that of the outer bars, a definition of visibility must be given. One possible definition is an average visibility that takes into account the luminances of all the bars and spaces. However, in order to be consistent with our previous definition of the theoretical resolution limit, we shall define the visibility by using the luminance of the central bar for the maximum and the luminance of the adjacent spaces as the minimum. By this definition, the visibility vanishes at the theoretical resolution limit. Figure 11 shows the visibility (for available data) so calculated for varying values of $\gamma(\delta)$ as function a of the fundamental frequency associated with a given tri-bar target.

EXPERIMENTAL RESULTS

Experimental measurements were made to verify the data of Fig. 6. The experimental setup is shown in Fig. 12. The coherence function at the object plane is $\Lambda_1(\bar{k}r/F)$, where r is the radius of the pinhole used as an incoherent source. The radius r may be varied to change the scale of the coherence function. The object, a tri-bar target, is then imaged at approximately 1:1, as shown by a 300-mm Schneider lens that has been stopped down to approximately $f/150$ by placement of special apertures in front of the lens. This yields a system with a cutoff frequency of approximately 10 cycles/mm and may be considered diffraction limited for our purposes. Because of measurement difficulties, it is necessary to work with relatively low-frequency tri-bar objects whose images may be photographed and then scanned on a microddensitometer without noticeable image degradation by either film or instrument.

The system is very sensitive to alignment errors and target quality. Very slight alignment errors or a poor-quality collimator introduce asymmetry and other anomalies in the images. With partially coherent illumi-

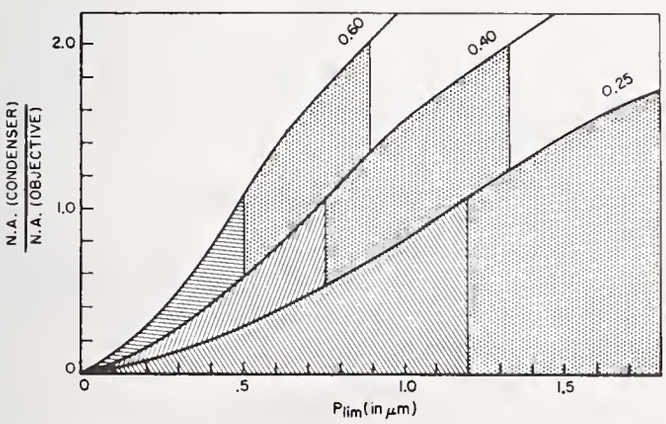


FIG. 10. The resolution limit of a tri-bar object as a function of the ratio of numerical apertures of condenser and objective for $\bar{\lambda} = 550$ nm. The numerical aperture of the condenser is shown as the parameter, on each curve.

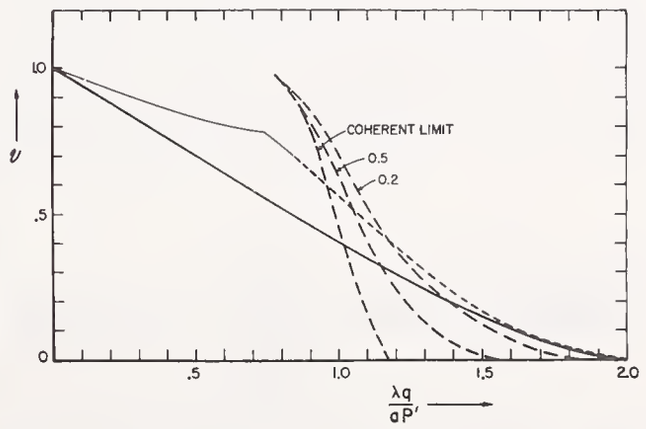


FIG. 11. Visibility curves for the image of a tri-bar target for indicated values of $\gamma(\delta)$. The modulation-transfer curve of an ideal lens is shown by the heavy solid curve. The incoherent limit, according to Charman,[2] is shown by the lighter and partially broken curve.

nation, the surrounding tri-bar groups in an Air Force Bar Target had to be blocked out, leaving only one set of three bars; coherence effects from the adjacent bars distorted the image.

The coherence function was checked by the method used by Thompson and Wolf,[16] and it was found that the scale of the coherence function was such that $\gamma(\delta)=0.72$. The photographic images of the tri-bar object ($1/p=8.98$ cycles/mm) for different aperture sizes ranging from 4 to 6 mm were then scanned and the normalized apparent and real periods were determined. The results are shown in Fig. 13, together with the nearest calculated curve [$\gamma(\delta)=0.70$ rather than 0.72]. The resolution limit and values agree surprisingly well with the calculated values. This was surprising in view of the fact that the target was not in a liquid gate and hence some phase variation was present in the object, due to the relief image in the photographic emulsion. Other sources of error include system alignment (pin-mounted components had to be used in the optical system) and the effects of photographic grain in the AHU microfilm used.

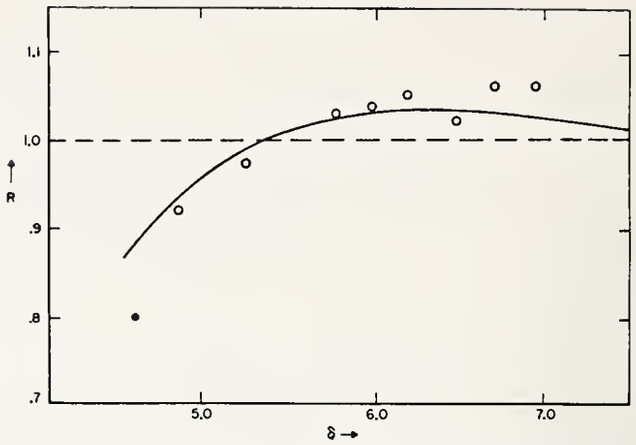


FIG. 13. Experimentally measured values of R for $\gamma(\delta)=0.72$ (●). and nearest calculated curve for $\gamma(\delta)=0.70$ from Fig. 6. The solid point indicates an unresolved image.

CONCLUSIONS

The theoretical resolution limit of a conventional tri-bar target varies with the degree of coherence of the illumination. Specifically, for a coherence function of the form $\Lambda_1(\alpha\delta)$, the predicted limit differs considerably from the two-point resolution limit. The greatest significance of this bears upon the instrumental evaluation of optical instruments with tri-bar targets (including microdensitometer evaluation). Significant losses (up to 75% at the coherent limit) may occur in the estimation of the resolution limit. Figures 7 and 10 are guides to the magnitude of the errors. For example, a microscope system utilizing critical illumination with a condenser numerical aperture of 0.25 and an objective numerical aperture[17] of 0.40 has a resolution limit of 1200 cycles/mm (at $\bar{\lambda}=5500$ Å) due to partial coherence. Because the system in incoherent illumination can theoretically resolve 1450 cycles/mm, there is a 17% loss of resolution when it is operated in this mode. Furthermore, coherence effects (distortion and contrast changes) will be present for tri-bar objects down to approximately 500 cycles/mm. [If the numerical apertures had been matched (0.40), the loss would have been only down to 1400 cycles/mm or 3.4%, with effects observed down to 750 cycles/mm.]

The results discussed here have all assumed a high-contrast 1:1 line/space ratio for the tri-bar object. It is clear that because of the object dependence, different values for the resolution limit will be obtained if the line/space ratio or contrast is changed.

The main difficulty in applying these results to visual evaluation of optical instruments is that the Sparrow criterion does not take into account the minimum modulation detectable by the eye. Although this difference is relatively small compared to the errors due to partial coherence, it can be taken into account in conjunction with the data of Fig. 13.

It is tempting to apply these data to the problem of threshold modulation of photographic film as well. However, because nonideal tri-bar objects are being viewed near the resolution limit of the film and these objects are therefore on the order of the grain size, the effect of the imaging of the grain as a function of coherence level must be considered. Although we expect the visual resolving power to vary with the degree of coherence, it will not correlate with the theoretical values given here for ideal tri-bar objects. Coherence may well be one of the factors causing the spread generally observed in threshold-modulation data from different laboratories.

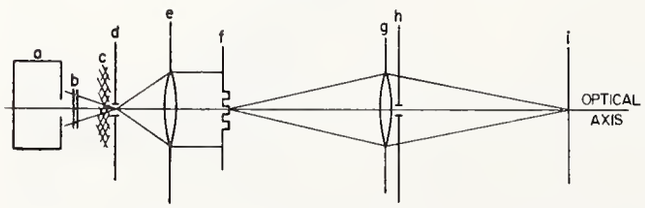


FIG. 12. Experimental system for tri-bar investigation. Elements are (a) mercury arc, (b) interference filter ($\lambda=5461$ Å), (c) diffuser, (d) pinhole, (e) collimator ($f=105$ mm), (f) tri-bar target (8.98 cycles/mm), (g) imaging lens, (h) apertures (4–6 mm), and (i) film plane.

ACKNOWLEDGMENTS

The author is indebted to R. Swing for many helpful discussions, to J. Doyle for help with the experiment, and to B. Peavy for the Bessel-function subroutines used in the calculations.

REFERENCES

* Paper presented at the 1970 Fall Meeting of the Optical Society of America [J. Opt. Soc. Am. **60**, 1563A (1970)].
[1] D. H. Kelly, Appl. Opt. **4**, 435 (1965).
[2] W. N. Charman, Phot. Sci. Eng. **8**, 253 (1964).

[3] R. Becherer and G. B. Parrent, Jr., J. Opt. Soc. Am. **57**, 1479 (1967).
[4] R. Swing and J. Clay, J. Opt. Soc. Am. **57**, 1180 (1967).
[5] R. Barakat, Opt. Acta **17**, 337 (1970).
[6] A. Offner and J. Meiron, Appl. Opt. **8**, 183 (1969).
[7] Lord Rayleigh, *Collected Papers* (Cambridge U. P., Cambridge, England, 1902), Vol. 3 (also Dover, New York, 1964).
[8] M. Born and E. Wolf, *Principles of Optics, 3rd ed.* (Pergamon, New York, 1965), p. 524.
[9] H. H. Hopkins and P. M. Barham, Proc. Phys. Soc. (London) **B63**, 737 (1950).
[10] C. Sparrow, Astrophys. J. **44**, 76 (1916).
[11] F. Rojak, M. S. thesis, Lowell Technological Institute, Lowell, Massachusetts, 1961.
[12] M. Beran and G. B. Parrent, Jr., *Theory of Partial Coherence* (Prentice-Hall, Englewood Cliffs, N. J., 1964), p. 123.
[13] R. Barakat, J. Opt. Soc. Am. **52**, 276 (1962).
[14] R. Barakat, J. Opt. Soc. Am. **53**, 415 (1963).
[15] D. N. Grimes and B. J. Thompson, J. Opt. Soc. Am. **57**, 1330 (1967).
[16] B. J. Thompson and E. Wolf, J. Opt. Soc. Am. **47**, 895 (1957).
[17] On the surface, these values may not seem realistic. However, the ANSI preliminary draft of the standard on MTF of photographic emulsions specifies these values.

nation, the surrounding tri-bar groups in an Air Force Bar Target had to be blocked out, leaving only one set of three bars; coherence effects from the adjacent bars distorted the image.

The coherence function was checked by the method used by Thompson and Wolf,[16] and it was found that the scale of the coherence function was such that $\gamma(\delta)=0.72$. The photographic images of the tri-bar object ($1/p=8.98$ cycles/mm) for different aperture sizes ranging from 4 to 6 mm were then scanned and the normalized apparent and real periods were determined. The results are shown in Fig. 13, together with the nearest calculated curve [$\gamma(\delta)=0.70$ rather than 0.72]. The resolution limit and values agree surprisingly well with the calculated values. This was surprising in view of the fact that the target was not in a liquid gate and hence some phase variation was present in the object, due to the relief image in the photographic emulsion. Other sources of error include system alignment (pin-mounted components had to be used in the optical system) and the effects of photographic grain in the AHU microfilm used.

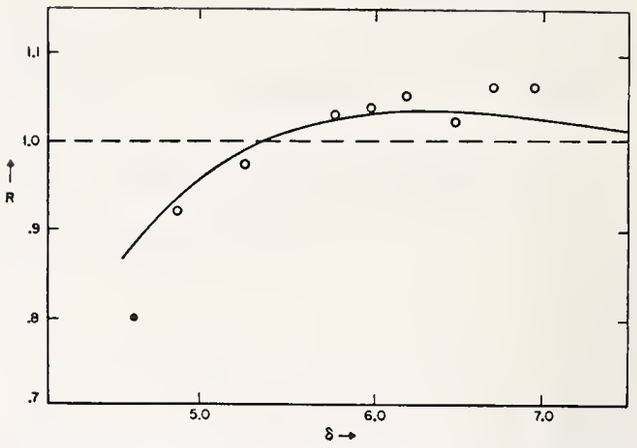


FIG. 13. Experimentally measured values of R for $\gamma(\delta)=0.72$ (●). and nearest calculated curve for $\gamma(\delta)=0.70$ from Fig. 6. The solid point indicates an unresolved image.

CONCLUSIONS

The theoretical resolution limit of a conventional tri-bar target varies with the degree of coherence of the illumination. Specifically, for a coherence function of the form $\Lambda_1(\alpha\delta)$, the predicted limit differs considerably from the two-point resolution limit. The greatest significance of this bears upon the instrumental evaluation of optical instruments with tri-bar targets (including microdensitometer evaluation). Significant losses (up to 75% at the coherent limit) may occur in the estimation of the resolution limit. Figures 7 and 10 are guides to the magnitude of the errors. For example, a microscope system utilizing critical illumination with a condenser numerical aperture of 0.25 and an objective numerical aperture[17] of 0.40 has a resolution limit of 1200 cycles/mm (at $\bar{\lambda}=5500$ Å) due to partial coherence. Because the system in incoherent illumination can theoretically resolve 1450 cycles/mm, there is a 17% loss of resolution when it is operated in this mode. Furthermore, coherence effects (distortion and contrast changes) will be present for tri-bar objects down to approximately 500 cycles/mm. [If the numerical apertures had been matched (0.40), the loss would have been only down to 1400 cycles/mm or 3.4%, with effects observed down to 750 cycles/mm.]

The results discussed here have all assumed a high-contrast 1:1 line/space ratio for the tri-bar object. It is clear that because of the object dependence, different values for the resolution limit will be obtained if the line/space ratio or contrast is changed.

The main difficulty in applying these results to visual evaluation of optical instruments is that the Sparrow criterion does not take into account the minimum modulation detectable by the eye. Although this difference is relatively small compared to the errors due to partial coherence, it can be taken into account in conjunction with the data of Fig. 13.

It is tempting to apply these data to the problem of threshold modulation of photographic film as well. However, because nonideal tri-bar objects are being viewed near the resolution limit of the film and these objects are therefore on the order of the grain size, the effect of the imaging of the grain as a function of coherence level must be considered. Although we expect the visual resolving power to vary with the degree of coherence, it will not correlate with the theoretical values given here for ideal tri-bar objects. Coherence may well be one of the factors causing the spread generally observed in threshold-modulation data from different laboratories.

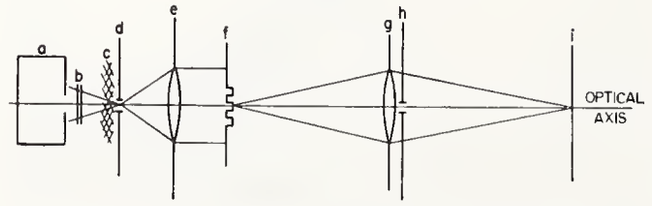


FIG. 12. Experimental system for tri-bar investigation. Elements are (a) mercury arc, (b) interference filter ($\lambda=5461$ Å), (c) diffuser, (d) pinhole, (e) collimator ($f=105$ mm), (f) tri-bar target (8.98 cycles/mm), (g) imaging lens, (h) apertures (4–6 mm), and (i) film plane.

ACKNOWLEDGMENTS

The author is indebted to R. Swing for many helpful discussions, to J. Doyle for help with the experiment, and to B. Peavy for the Bessel-function subroutines used in the calculations.

REFERENCES

* Paper presented at the 1970 Fall Meeting of the Optical Society of America [J. Opt. Soc. Am. 60, 1563A (1970)].
[1] D. H. Kelly, Appl. Opt. 4, 435 (1965).
[2] W. N. Charman, Phot. Sci. Eng. 8, 253 (1964).

[3] R. Becherer and G. B. Parrent, Jr., J. Opt. Soc. Am. **57**, 1479 (1967).
[4] R. Swing and J. Clay, J. Opt. Soc. Am. **57**, 1180 (1967).
[5] R. Barakat, Opt. Acta **17**, 337 (1970).
[6] A. Offner and J. Meiron, Appl. Opt. **8**, 183 (1969).
[7] Lord Rayleigh, *Collected Papers* (Cambridge U. P., Cambridge, England, 1902), Vol. 3 (also Dover, New York, 1964).
[8] M. Born and E. Wolf, *Principles of Optics, 3rd ed.* (Pergamon, New York, 1965), p. 524.
[9] H. H. Hopkins and P. M. Barham, Proc. Phys. Soc. (London) **B63**, 737 (1950).
[10] C. Sparrow, Astrophys. J. **44**, 76 (1916).
[11] F. Rojak, M. S. thesis, Lowell Technological Institute, Lowell, Massachusetts, 1961.
[12] M. Beran and G. B. Parrent, Jr., *Theory of Partial Coherence* (Prentice-Hall, Englewood Cliffs, N. J., 1964), p. 123.
[13] R. Barakat, J. Opt. Soc. Am. **52**, 276 (1962).
[14] R. Barakat, J. Opt. Soc. Am. **53**, 415 (1963).
[15] D. N. Grimes and B. J. Thompson, J. Opt. Soc. Am. **57**, 1330 (1967).
[16] B. J. Thompson and E. Wolf, J. Opt. Soc. Am. **47**, 895 (1957).
[17] On the surface, these values may not seem realistic. However, the ANSI preliminary draft of the standard on MTF of photographic emulsions specifies these values.

Reprinted from **APPLIED OPTICS**, Vol. 11, page 914, April 1972

Optical Autocorrelator with Special Application to MTF Measurement

D. N. Grimes

An optical autocorrelator is described which is based on a modified Sagnac two-mirror interferometer that operates in the zero-fringe mode. The real-time radiant flux output is the square of the autocorrelation of the aperture function. In a particular application, when the aperture function is the pupil function of a lens under test, the autocorrelator output is the MTF of the lens. The system is not limited in focal length; aperature size is limited only by the sizes of the beam splitters and mirrors. Experimental results are given with comparison data for both infinite-conjugate MTF and the autocorrelation of an annular aperture. Methods of adapting the system for measurement of the phase of the transfer function and for finite conjugate testing are also described.

Introduction

Real-time optical autocorrelation has thus far generally required the use of two identical transparencies simultaneously placed in an optical system. The product of the function and its shifted duplicate is formed by transmitting a uniform beam of light through the two transparencies successively and collecting the transmitted light. The technique is cumbersome and impractical when an exact duplicate cannot be made.

Kovasznay and Arman[1] eliminated the need for the duplicate by employing a broad source and a single mirror. However, they found that this system suffered from parasitic effects as the shift variable increased. This was due to the use of off-axis illumination to achieve the desired shift and was therefore not suitable for modulation transfer function (MTF) measurement.

Interferometers have been used for MTF measurement by Hopkins,[2] Kelsall,[3] Hariharan and Sen,[4] and others. However, these interferometers, including the Sagnac interferometer used by Hariharan and Sen, do not directly perform the optical autocorrelation of the pupil function. The method employed by these interferometers based on the addition of sheared wavefronts as described by Hopkins[2] is conceptually quite different from the method of optical autocorrelation described here.

Description

A schematic of the system discussed here is shown in Fig. 1. The portion of the system shown on the right, consisting of two mirrors (M_1 and M_2) and a beam splitter P_2 forming a closed loop, is a Sagnac interferometer. A laser irradiates a pinhole A_1 at the focus of the lens. The collimated beam is the input to the Sagnac[5] interferometer, which is adjusted initially for equal paths in the two beams. The beams return through the lens along the same path forming the product of the aperture function with itself. The beam splitter P_1 separates the return beam from the source, and a second pinhole A_2 at the focus of the lens serves as the collector for the photomultiplier. Displacement of the mirror M_2 perpendicular to its face shears the two arms of the interferometer in opposite directions. The output as a function of the mirror motion is the square of the autocorrelation of the aperture function. Hence, in this form, only the modulus of the autocorrelation function is recovered. If the aperture function is the pupil function of the lens, the autocorrelation yields the MTF. On the other hand, if a transparency is placed in front of the lens and the lens is diffraction limited, the aperture function is the transmittance of the transparency, and the output is the autocorrelation of the transmittance function. If the three-mirror version of the Sagnac interferometer is used instead, the output in this case will be the autoconvolution of the aperture function.

Theoretical Analysis

To describe the system mathematically, it is convenient to use the analog system shown in Fig. 2 where $t(\bar{x})$ is the amplitude transmittance of the aperture A_1, $\gamma(\bar{x}_1,\bar{x}_2)$ the degree of coherence between the points $\bar{x}_1$ and $\bar{x}_2$ in the plane of the aperture A_1, and $\Psi(\bar{y})$ the pupil function of the lens. Then the intensity in the plane of the aperture A_2 is given in vector notation by

The author is with the National Bureau of Standards, Washington, D.C. 20234.

Received 29 July 1971.

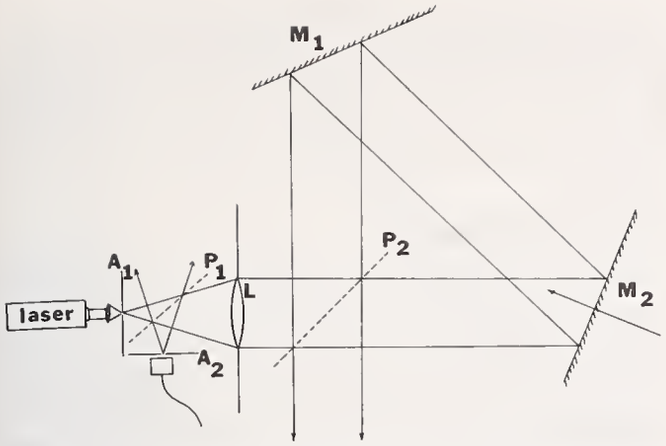


Fig. 1. Schematic of the optical autocorrelator system.

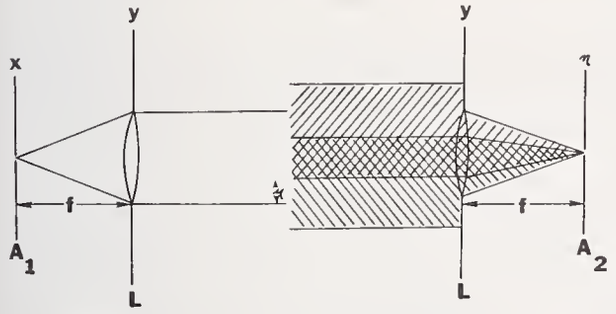


Fig. 2. Analog system for analysis of the optical autocorrelator.

$$I(\mu,\bar{\eta}) = \int \gamma(\bar{x}_1,\bar{x}_2)t(\bar{x}_1)t^*(\bar{x}_2)\exp[ik(\bar{y}_1-\bar{x}_1)^2/2f]\cdot\exp[-ik[\bar{y}_2-\bar{x}_2)^2/2f]\exp[-ik(\bar{y}_2{}^2-\bar{y}_1{}^2)/2f]\cdot\Psi(\bar{y}_1)\Psi(\bar{y}_2)\tfrac{1}{2}[\Psi(\bar{y}_1+\mu)\Psi^*(\bar{y}_2+\mu)+\Psi(\bar{y}_1-\mu)\Psi^*(\bar{y}_2-\mu)]\cdot\exp[ik(\bar{y}_2{}^2-\bar{y}_1{}^2)/2f]\exp[-ik(\bar{\eta}-\bar{y}_1)^2/2f]\cdot\exp[ik(\bar{\eta}-\bar{y}_2)^2/2f]d\bar{x}_1d\bar{x}_2d\bar{y}_1d\bar{y}_2, \quad (1)$$

where the expression in braces represents the two beams sheared in opposite directions on the second pass through the lens. By making the usual paraxial approximations and assuming coherent illumination, where $\gamma(\bar{x}_1,\bar{x}_2) = 1$, we reduce Eq. (1) to

$$I(\mu,\bar{\eta}) = \int \tilde{t}(k\bar{y}_1/f)\tilde{t}^*(k\bar{y}_2/f)\Psi(\bar{y}_1)\Psi^*(\bar{y}_2)\cdot\tfrac{1}{2}[\Psi(\bar{y}_1+\mu)\Psi^*(\bar{y}_2+\mu)+\Psi(\bar{y}_1-\mu)\Psi^*(\bar{y}_2-\mu)]\exp[ik\bar{\eta}(\bar{y}_1-\bar{y}_2)/f]d\bar{y}_1d\bar{y}_2, \quad (2)$$

where $\sim$ denotes the Fourier transform. If we demand that the source be a delta function [i.e., $\tilde{t}(k\bar{y}/f)$ = constant] and evaluate this integral at $\bar{\eta} = 0$, corresponding to an unresolved pinhole source and a pinhole collecting aperture, the output reduces to the square of the MTF of the lens,

$$I(\mu) = I_0\left|\int \Psi(\bar{y})\Psi^*(\bar{y}+\mu)d\bar{y}\right|^2 = I_0|\tau(\mu)|^2. \quad (3)$$

Off-axis MTF measurements may be made by rotating the lens about its nodal point.

If a transparency with transmission $f(y)$ had been placed in front of the lens, the output would be given by

$$I(\mu) = I_0\left|\int \Psi(\bar{y})f(\bar{y})\Psi^*(\bar{y}+\mu)f^*(\bar{y}+\mu)d\bar{y}\right|^2, \quad (4)$$

which for a diffraction-limited lens reduces to the autocorrelation of the transparency

$$I(\mu) = I_0\left|\int f(\bar{y})f^*(\bar{y}+\mu)d\bar{y}\right|^2. \quad (5)$$

Alignment

As shown in Fig. 1, two beams are thrown away at the beam splitter. These overlapping beams serve as an extremely sensitive monitor of system alignment. Only when the paths are identical will the overlapping area be free of fringes. This assures a linear phase error of less than $\lambda/4$ across the face of the aperture due to tilt (assuming negligible error introduced by the mirror or beam splitter surfaces). The linear phase error will not affect the MTF. For a discussion of this point, see Ref. 6. The only affect of this linear phase error is a displacement of the return beam at the focus, causing the beam to wander about the collecting aperture. Calculations show that an error of 1 sec in mirror tilt causes a displacement of 2 μm at the aperture A_2 for a focal length of 180 mm. The angles of the beam splitter and mirrors in the interferometer, while not fixed to specific values, must bear a fixed relationship to one another as shown in Fig. 3. For both beams to return along the same path, we demand that the expression for the angle between the incoming and return beams be equal to 0°. A convenient choice of these angles is $\alpha = 45°$, $\beta = 22.5°$, and $\gamma = 67.5°$. The relationship of the frequency scale of the output to the mirror displacement is given by

$$\omega = (2d\cos\gamma)/\lambda f, \quad (6)$$

where d is the mirror displacement, f the focal length of the lens, and λ the wavelength. As expected, the frequency depends only on the angle of the mirror being moved.

Focus and Autocollimation

When the collecting aperture A_2 is not at the focus of the lens and shift is present, the return images separate. This image separation is a very accurate focus indicator.

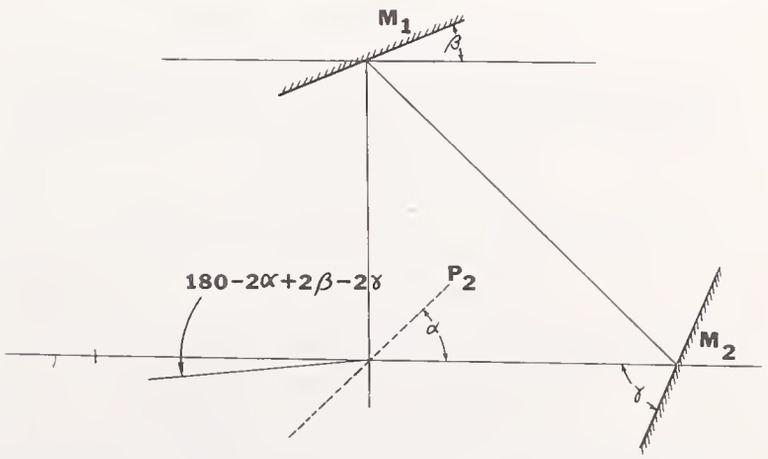


Fig. 3. Angles of components in interferometer portion of the system.

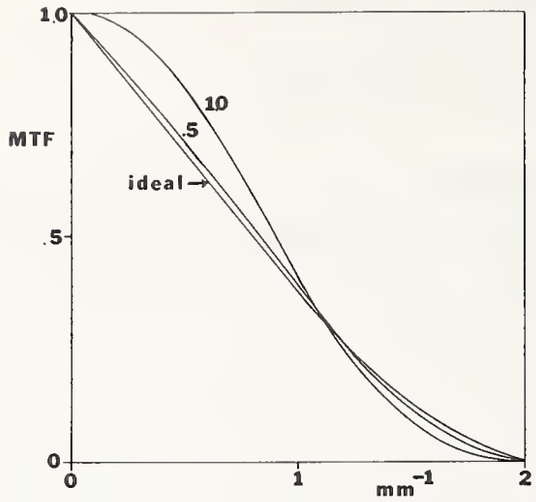


Fig. 4. The effect of source size on the MTF of a diffraction-limited lens for a circular source and indicated values of ρ.

Methods utilizing this phenomenon have been proposed for focus sensing in high angular resolution systems.[7,8] (Vyce[7] indicates this method to be theoretically capable of an accuracy of better than 10^{-2} μm for an $f/4$ system.) However, focus errors at the aperture A_2 are not the only problem. If the aperture A_1 is not at the focus of the lens as well, the resulting beam in the interferometer will not be collimated. The principal manifestation of this kind of error is a shift in focus at the aperture A_2 as the mirror is moved. Again this is easily observable from the resulting image separation in the plane of the aperture A_2. In the initial alignment both focus positions may be adjusted simultaneously until no image separation occurs with mirror motion.

Source Size

In the preceding analysis we assumed a delta function source and collector. The delta function source produced uniform irradiance (amplitude and phase) across the aperture of the lens. The question of concern then is how large a source can be used and still satisfy this criterion. If we go back to Eq. (2) and allow a finite source size while still evaluating the result at $\eta = 0$, we get

$$I(\mu) = \left| \int \bar{l}\left(\frac{k\bar{y}}{f}\right) \Psi(\bar{y})\Psi(\bar{y} + \mu) d\bar{y} \right|^2. \tag{7}$$

If we now assume a uniform circular source with radius ϵ and write the ratio of source diameter to Airy disk diameter as

$$\rho = k\epsilon a/3.83f, \tag{8}$$

the amplitude distribution across the lens becomes

$$\bar{l}(kr/f) = \Lambda_1(3.83\rho r), \tag{9}$$

where Λ_1 is the first-order Bessel function divided by its argument, the output is given by

$$I(\mu) = \left| \int \Lambda_1(3.83\rho r)\Psi(r)\Psi(r + \mu) r dr d\theta \right|^2, \tag{10}$$

where the vector y has been converted to polar coordinates. The results of calculations of the resulting MTF for a diffraction-limited lens are shown in Fig. 4 for various values of the ratio (ρ) of source size to Airy disk diameter. For $\rho = \frac{1}{4}$, there is less than 0.5% observable error in the output. This corresponds to about a 20% observable falloff in intensity at the edges of the lens.

Experimental Results

The experimental system consisted of a 2.5-mW He–Ne laser illuminating a 5-μm pinhole. Interferometer quality pellicles were used for beam splitters to avoid secondary reflections. While it is necessary for the beam splitter near the source to be a pellicle because of aberrations introduced by a thick beam splitter there, the beam splitter in the interferometer need not be a pellicle. Although care must be taken to insure that any light reflected from the second surface does not fall on the detector, a sufficiently thick or wedged beam splitter will assure this and at the same time allow testing of larger aperture optical systems than is currently possible with pellicles. Pellicles, of course, create a problem if acoustical noise is present. However, if the person running the system is alone in a fairly isolated laboratory, the acoustical noise is no more of a problem than laser fluctuations or other vibrations in the system. The movable mirror was mounted on ways with the mirror tilt adjustable with piezoelectric micrometers, which were also used to position the aperture A_2. With no lens or aperture A_1 in the system, the interferometer may be grossly adjusted. Finer adjustment is made with the lens and aperture in place utilizing the fringe field discussed earlier. The angles are adjusted so that there are no fringes present with and without shift. The apertures A_1 and A_2 were both 5-μm pinholes for a lens at $f/16$. Best focus for the aperture A_2 is then determined with no shift present by searching for the maximum intensity on-axis. Shift is then introduced and errors in autocollimation detected if image separa-

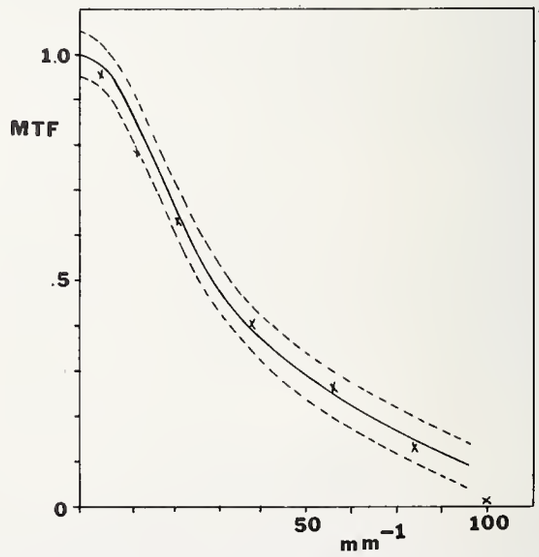


Fig. 5. MTF of an $f/16$ lens measured on the autocorrelator (smooth curve) and the sine wave bench (X). The broken lines represent the σ bounds for the data used.

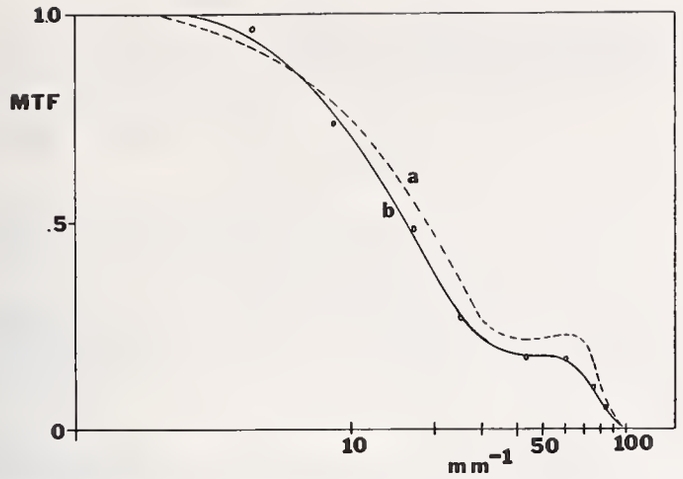


Fig. 6. MTF of an annular aperture (b) for the same lens at $f/16$ together with the curve for a diffraction-limited lens (a).

tion occurs and fringes appear. Repetition of these last two steps as the lens is moved in small increments allows accurate determination of best focus for both apertures. As the mirror is then moved to plot the MTF, small errors in the ways may necessitate small adjustments in the mirror tilt. Although in an automated system this may be done by servo techniques, the data described here were produced with manual adjustments. Figure 5 shows the resulting MTF data. Sine wave data are shown for comparison. The autocorrelator data are based on a least-mean-squares curve fit of ten data sets (194 data points) to a curve of the form $f(x) = a_1 \exp(-a_2x^2) + a_3 \exp(-a_4x^2)$, which has been found to fit MTF data extremely well. The major contribution to the rms error of 5% is due to inability to align accurately the optical elements in the test system. This is a fault of the experimental equipment used and not inherent in the method. Figure 6 shows data for the same lens with a circular obstruction at its center forming an annular aperture. Sine wave data were not available for comparison, therefore only the ideal MTF for an annular aperture of this size is shown.

Although data are shown for a lens with a 100-cycles/mm frequency cutoff, the system is presently being adapted for higher frequencies by insertion of 1-μm pinhole apertures and improved sensitivity in the photomultiplier.

Measurement of the Phase of the MTF and Adaptation to Finite Conjugates

It is desirable in an optical testing device to be able to measure the phase as well as the amplitude of the transfer function. The method described here for measuring the phase is easily implemented in the autocorrelator already described. Basically, the coherent beam thrown away at the first beam splitter is reflected back directly onto the detector. This beam is a diverging spherical wavefront; but because the detector is extremely small (less than 5 μm), the phase of the beam is essentially constant. Hence, a new function is measured by the radiometer,

$$B(\mu) = |\tau(\mu) + C|^2, \tag{11}$$

where $\tau(\mu)$ is the OTF of the lens under test and C is the magnitude of the coherent beam reflected back on the photomultiplier. If we write the optical transfer function in the form,

$$\tau(\mu) = A(\mu) \exp[i\phi(\mu)], \tag{12}$$

where $A(\mu)$ is the MTF already measured, the function $B(\mu)$ becomes

$$B(\mu) = A^2(\mu) + C^2 + 2CA(\mu) \cos\phi(\mu). \tag{13}$$

Solving this equation for the phase of the OTF yields

$$\phi(\mu) = \cos^{-1}\left[\frac{B(\mu) - A^2(\mu) - C^2}{2CA(\mu)}\right]. \tag{14}$$

The constant C may be readily determined by blocking the interferometer beam and measuring the flux from the reference beam incident on the detector. Unlike holography, C need not be very large because we know $A(\mu)$, which is the MTF of the lens. Hence, with the lens set up for an MTF measurement, measurement of $B(\mu)$ may be speedily performed without realigning the system. All that is required is the positioning of the mirror, which is not critical.

The adaptation of the autocorrelator to finite conjugates is shown in Fig. 7. The basic changes required are the addition of another mirror forming the three-mirror version of the Sagnac interferometer and the variation of the pathlength in the interferometer loop so that it equals twice the image conjugate of the lens under test. Imaging the source within the interferometer causes a folding of the beam on its return path through the lens under test. Addition of the third mirror reverses this effect by folding the beam again. In setting up this system, it would be best to make the interferometer rectangular and detach the two mirrors as a unit from the remainder of the interferometer. Moving these two mirrors as a unit allows the variation of the pathlength of the interferometer for testing different conjugate distances without readjusting the remainder of the system. As the mirror is moved to introduce the necessary shear, the pathlength in the interferometer changes thereby causing a shift of focus

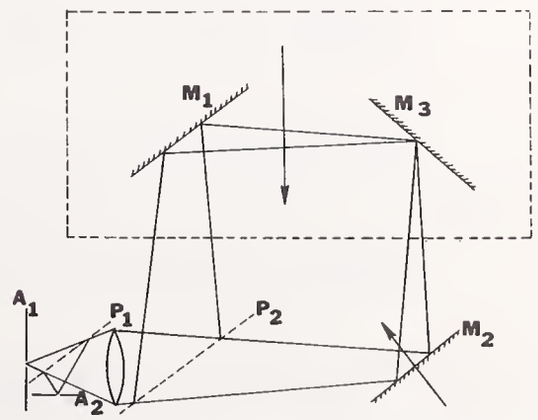


Fig. 7. Adaptation of autocorrelator for finite conjugate MTF testing.

for any but a collimated beam. Therefore, a compensated lengthening of the interferometer pathlength must be made to maintain the proper conjugate distance. This may be accomplished by moving the two-mirror unit simultaneously, as the third mirror is moved.

Conclusions

The autocorrelation of transparencies may be done accurately since diffraction-limited optics are easily available on axis. For MTF measurements the system will in principle take any focal length lens, the aperture size being limited only by the surface area of the beam splitters and mirrors. At present, only infinite conjugate testing may be done and only the modulus recovered. Methods of adapting the system to recover the phase of the OTF and adaptation for finite conjugate testing are presently being tested. We found the principal limitation for both autocorrelation and MTF measurement to be source size. Based on the calculations shown, a lens with 500-cycles/mm frequency cutoff may be measured to 2% accuracy. Higher frequency cutoff systems will exhibit larger errors at all frequencies unless a means of uniformly irradiating larger lens apertures is found.

However, in spite of these limitations, the autocorrelator offers a fast method of MTF measurement. The method is not prone to object- or coherent-dependence in the way that current MTF measurement techniques utilizing slits, edges, or grating objects are.[9] It is therefore theoretically capable of testing higher frequency optics with better accuracy.

The author is indebted to R. E. Swing and K. Mielenz for helpful discussions and to F. Rosberry and R. Hartsock for much of the experimental work.

This paper was given at the Spring 1971 Optical Society of America meeting [J. Opt. Soc. Am. **61,** 687A (1971)].

References

1. L. Kovasznay and A. Arman, Rev. Sci. Instrum. **28,** 793 (1957).
2. H. H. Hopkins, Optica Acta **2,** 23 (1955).
3. D. Kelsall, Proc. Phys. Soc. (London) **73,** 465 (1959).
4. P. Hariharan and D. Sen, Proc. Phys. Soc. (London) **75,** 434 (1960).
5. W. H. Steel, *Interferometry* (Cambridge U. P., London, 1967), p. 94.
6. E. L. O'Neill, *Introduction to Statistical Optics* (Addison-Wesley, Reading, Mass., 1963), p. 92.
7. J. R. Vyce, Opt. Laser Tech. **3,** 18 (1971).
8. J. W. Evans, Appl. Opt. **3,** 1387 (1964).
9. G. C. Brock, *Image Evaluation* (Focal Press, London, 1970).

4. Photographic Science

Foreword

The first reprint in this section describes the method of making the so-called Davis-Gibson filters that are used to reproduce sunlight and daylight in the laboratory. The use of these filters was adopted as part of a standard method of sensitometry by the Seventh International Congress of Photography, in 1928, and as part of the standard sources for photometry and colorimetry by the International Commission on Illumination, in 1931. They have been cited in national and international standards since that time. The publication has been out of print for many years, so its republication will be welcomed by standards writers everywhere. There is a trend toward the adoption of specified spectral power distributions rather than detailed physical and chemical specifications of the sources themselves. The development of glass filters to satisfy the requirements has been slow and not as yet completely satisfactory. The more general availability of Davis and Gibson's work should be particularly valuable during this transition period.

The work on photoresists and high-resolution microphotography played an important role in introducing techniques that were to revolutionize the electronics industry and make much of our space technology possible. Scientists and engineers from every segment of the electronics industry came to the Bureau in the late fifties and early sixties to discuss problems of photoetching with Mr. Pope and to see the new extremely high-resolution camera we had developed to test the resolving powers of photographic materials. (See paper No. 4.8 in this volume.) These techniques, much improved in industry, are the basis of ultramicrominiaturization of electronics.

The NBS Microcopy Resolution Test Chart has provided the means of specifying and testing the resolving power of microfilming systems for over thirty years. Initially intended for specification of Federal Government contract work, these charts have been adopted throughout the industry, both here and abroad.

Most of the concepts, terminology, and notation for optical modulation proposed for standardization in paper No. 4.10 in this volume are in the process of standardization in the American National Standards Institute.

Archivists know NBS primarily for work on the long-term stability of paper and photographs. They will find most of the recent publications on photographs brought together here for the first time.

U. S. DEPARTMENT OF COMMERCE
R. P. LAMONT, Secretary
BUREAU OF STANDARDS
GEORGE K. BURGESS, Director

MISCELLANEOUS PUBLICATION, BUREAU OF STANDARDS, No. 114

FILTERS FOR THE REPRODUCTION OF SUNLIGHT AND DAYLIGHT AND THE DETERMINATION OF COLOR TEMPERATURE

By

RAYMOND DAVIS
K. S. GIBSON

January 21, 1931

UNITED STATES
GOVERNMENT PRINTING OFFICE
WASHINGTON : 1931

FILTERS FOR THE REPRODUCTION OF SUNLIGHT AND DAYLIGHT AND THE DETERMINATION OF COLOR TEMPERATURE [1]

By Raymond Davis and K. S. Gibson

ABSTRACT

A large number of filters, as listed in the introduction of the paper, have been developed for use in photographic sensitometry, colorimetry, and photometry. They are reproducible from specification and may be prepared in the laboratory. They consist of a 2-compartment cell with three borosilicate crown glass windows, the two compartments being filled, respectively, with solutions A and B, having compositions as follows:

Solution A

Copper sulphate ($CuSO_4.5H_2O$)	g	c
Mannite ($C_6H_8(OH)_6$)	g	c
Pyridine (C_5H_5N)	cc	30.0
Water (distilled) to make	cc	1,000

Solution B

Copper sulphate ($CuSO_4.5H_2O$)	g	c_1
Cobalt ammonium sulphate ($CoSO_4.(NH_4)_2SO_4.6H_2O$)	g	c_2
Sulphuric acid (specific gravity 1.835)	cc	10.0
Water (distilled) to make	cc	1,000

The values of c, c_1, and c_2 vary from one filter to another.

The design of the filters has been based upon extensive spectrophotometric measurements of the component solutions and cell, including a study of Beer's law over the ranges of concentrations used, and upon methods of computation giving accurate color matches and relatively good spectral energy matches between the source-and-filter combination and the energy distribution which it was sought to duplicate. Detailed studies have been made of the reliability of the data, the spectrophotometric reproducibility of the chemicals, the permanence of the solutions, their temperature coefficients, and various other factors of importance in the preparation and use of the filters.

The paper contains 33 figures and 26 tables, giving among other things the energy distributions and fundamental spectrophotometric and colorimetric data used in the computations, auxiliary spectrophotometric and colorimetric information obtained during the investigation, and the trilinear coordinates of the various source-and-filter combinations and of the Planckian radiator from 1,600° to 20,000° K. computed on the basis of the adopted mean sun as the "neutral" stimulus. The filters are further described in 38 charts, each chart giving the chemical formula, the spectral transmission, and the light transmission of the filter, as well as the spectral energy distribution of the source, that of the source and filter combined, and the energy distribution which serves as the ideal in that particular case.

[1] Partial reports of this investigation have been published as follows:

1. Davis and Gibson, Reproducible Liquid Filters for the Production of "White Light" (abstract), J. Opt. Soc. Am. and Rev. Sci. Inst., **14**, p. 135; February, 1927.
2. Davis and Gibson, Reproducible Liquid Filters for the Determination of the Color Temperatures of Incandescent Lamps (abstract), Phys. Rev. (2), **29**, p. 916; June, 1927.
3. Davis and Gibson, Filters for the Reproduction of Sunlight and the Determination of Color Temperature (abstract), J. Opt. Soc. Am. and Rev. Sci. Inst., **16**, p. 332; May, 1928.
4. Gibson and Davis, Methods for Determining the Color of Sunlight and Daylight (abstract), J. Opt. Soc. Am. and Rev. Sci. Inst., **18**, p. 442; June, 1929.
5. Davis and Gibson, Artificial Sunlight for Photographic Sensitometry, Trans. Soc. Motion-Picture Engineers, **12**, No. 33, pp. 225–236; 1928; Proc., Seventh International Congress of Photography, July, 1928, pp. 161–173.
6. A Filter for Changing the Color Temperature of 2,848° K. to Average Daylight, B. S. Tech. News Bull., No. 138, pp. 143–144; No. 139, p. 156; 1928.
7. Filters for the Reproduction of Sunlight and the Determination of Color Temperature. (Proposed Working Standards for Sensitometry, Colorimetry, and Photometry.) Standards Yearbook, pp. 188–191; 1928; also Standards Yearbook, pp. 138–141; 1930.

518-743 O - 73 - 42

CONTENTS

I. INTRODUCTION

Both in photographic sensitometry and in colorimetry the need for a practical and reproducible working standard of "white light" is well recognized, although there is as yet no universal agreement regarding a precise definition of white light itself. It is generally conceded, however, that the spectral energy distribution of such white light must approximate that of daylight, noon sunlight, or a Planc-

kian radiator or "black body" at some color temperature between 4,000° and 8,000° K. (For the sake of brevity in the present paper, the authors have used the expression "white light" where it is desired not to be specific, but to refer to radiant energy having a spectral distribution approximating these.)

In the meantime the need is being met by the use of arbitrary working standards. In the description of a color in terms of dominant wave length, purity, and brightness, or in terms of the three excitations, whether such description is by way of computation from spectral distribution or by direct observation, a standard "white," often termed a "neutral" stimulus, is necessary. For such computations, Ives (1) [2] has proposed and used the spectral energy distribution of a Planckian radiator at 5,000° K. Likewise, Priest (2, 3) has devised, both for computation and direct experiment, an artificial sunlight which has been extensively used as a "neutral" stimulus in various investigations at the National Bureau of Standards. Further information regarding the need and uses of standard white light may be found in a report (4) of the Committee on Colorimetry for 1920–21, Optical Society of America, and also in a paper by Guild (5). It may be emphasized that, while the establishment and general adoption of a standard "neutral" stimulus satisfactory in all respects to all concerned is of fundamental importance, valuable contributions have been and are being made to the science of colorimetry by the use of standards such as those noted above, which are not in universal use. It may be further noted that, while any reasonable energy distribution may be readily used in colorimetric computations as a standard white, the accurate reproduction of such distribution in the laboratory may not be possible.

In photographic sensitometry the most important factor in specifying the characteristics of photographic negative materials is that referred to as speed. This has been emphasized in a report (6, 7) of the Committee on the Unit of Photographic Intensity, Optical Society of America. It was there pointed out that relative speeds measured by incandescent light in the laboratory were radically different from the relative speeds obtained under the sunlight illumination to which the great majority of photographic negative materials are exposed. This is a result, of course, of the widely varying spectral sensitivities of the different emulsions. The ordinary emulsions are sensitive only to the blue and violet (besides, of course, the ultra-violet), the orthochromatic to the violet, blue, and yellow, and the panchromatic in varying degrees to the whole visible spectrum.

In 1925 the Sixth International Congress of Photography tentatively adopted (6, 8) two standards of energy distribution for use in sensitometry; one a color temperature of 2,360° K. (nominally the acetylene flame), and the other a color temperature of 5,000° K. (representing sunlight) to be obtained by means of a specified light filter and a 2,360° K. source as recommended by the American committee. The recommendations [3] were as follows (6, 8):

[2] Numbers in parentheses, followed occasionally by page references, refer to the Bibliography at the end of the paper.

[3] The action taken by the Seventh International Congress of Photography in 1928 is considered later in Section VI, 1.

1. It is recommended to use, for sensitometric work, a light source of which the radiation is identical with that of a black body at 2,360° (absolute).

The intensity of this source shall be expressed in candles. This intensity shall be measured visually by ordinary photometric procedure in comparison with the usual photometric standards.

2. In case it is considered necessary to use radiation representing approximately the radiation of a black body at 5,000° absolute (average daylight day) there shall be placed before the light source provided in the preceding paragraph, a colored filter suitably chosen.

3. It is provisionally recommended to use for this purpose a filter composed of two distinct liquids, each of a thickness of 10.0 mm. One of these liquids shall be an aqueous solution of which 100 cubic centimeters shall contain 3.00 grams of pure crystallized copper sulphate $CuSO_4.5H_2O$ and 5 cubic centimeters of a solution of sulphuric acid containing per liter 50 grams of pure sulphuric acid reading 66° Baumé. The other liquid shall consist of an aqueous solution of copper ammoniacal sulphate containing three times the quantity of ammonia necessary to give the compound $Cu(NH_3)SO_4$; 100 cubic centimeters of this solution should contain a quantity of copper corresponding to 0.50 gram of pure crystallized copper sulphate $CuSO_4.5H_2O$; this solution should not be prepared before use for less than a day nor more than a month.

The cell shall be made of three sheets of white polished glass, the thickness of each sheet being not less than 2 mm nor greater than 4 mm. The cell used should be closed.

This compensating filter was hurriedly devised (at the Research Laboratory of the Eastman Kodak Co.), as the time was very limited and the American committee was unanimous in its opinion that daylight or sunlight quality should be used in determining the speeds of photographic emulsions. While this filter did not give the 5,000° K. energy distribution with the desired accuracy, it was distinctly better for the purpose than the 2,360° K. source alone. It was suitable for temporary use, but was not acceptable as a permanent International Standard.

The problem of devising suitable filters for converting the color and spectral energy distribution of incandescent light sources to those of sunlight, 5,000° K., or other possible white light, was of interest to the authors not only as members of the American committee noted above, but also in connection with the work of those sections of the National Bureau of Standards devoted to photographic and colorimetric research. The investigation has developed greatly since its inception, and the scope of the material herein presented may be outlined as follows:

1. A series of filters has been developed by which the spectral energy distribution of a Planckian radiator at any color temperature from 2,300° to 4,000° K., inclusive, may be converted into a close approximation of the spectral energy distribution of mean noon sunlight at Washington. The data cover the range from 350mμ to 720mμ. Light transmissions have been computed for all the filters.

2. By designing these filters so that they all give an accurate color match with the adopted mean sunlight, in addition to a close energy match, they may be used to calibrate the intensity and color temperature of incandescent lamps at any color temperature above 2,300° K., provided a source is available calibrated for intensity and color temperature at some one color temperature within this range. The comparison of intensities of incandescent lamps at different color temperatures may thus be made under conditions of exact color match.

3. A second series of filters has been developed for converting the color of a Planckian radiator at 2,848° K. to that of any Planckian radiator between 3,500° and 10,000° K., inclusive, keeping the energy distribution as closely matched as possible.

4. Seven additional filters, not in the above series, are also given, viz: (*a*) A filter for converting 2,848° K. to the color of the sun outside the earth's atmosphere, which satisfactorily represents the color of overcast sky, this filter thus affording a reproducible daylight standard. By varying the color temperature of the light source from 2,300° to 3,000° K. the filter will closely match the colors given by daylight and noon sunlight; (*b*) a filter for converting 2,848° K. to Abbot-Priest sunlight; (*c*) a filter for converting acetylene to mean noon sunlight at Washington; (*d*) and (*e*) two filters, one for converting 2,848° K. to noon sunlight at the summer solstice at Washington and the other for converting 2,848° K. to noon sunlight at the winter solstice at Washington; (*f*) and (*g*) two filters designed to cover a wide range of black-body color temperatures by varying the color temperature of the light source between the limits of 2,000° and 3,100° K. With this variation one of the filters covers the black-body scale from 2,650° to 5,000° K., and the other from 4,000° to 19,000° K.

5. Spectral transmissive data are given on the standard component solutions which should be of value in checking and calibrating spectrophotometric apparatus. Various other detailed information is given concerning these solutions which is of permanent value independent of their use in the filters.

6. Tables of energy and visibility data are presented which are convenient for general reference, viz: (*a*) The relative spectral energy of various phases of sunlight, from Abbot's data; (*b*) the relative spectral energy of a Planckian radiator, from 350mμ to 720mμ, at various intervals from 2,000° to 20,000° K., the values being taken or computed from the tables and curves (9) of B. S. Miscellaneous Publication, No. 56; (*c*) the internationally adopted (10) visibility data as recommended by Gibson and Tyndall (11, 12).

7. Tables are presented from which the trilinear coordinates (*r*, *g*, *b*) may be obtained for a Planckian radiator at any temperature from 1,600° to 20,000° K., on the basis of the adopted mean sunlight as standard "white." The excitations and the method used in these computations are given so that the *r*, *g*, and *b* coordinates may be computed for any desired energy distribution.

The division of work between the photographic and colorimetry sections of the bureau has been, in general, as follows: The former has been responsible for the selection of the materials best suited for the purpose, for devising methods for the determination of the correct proportions of the ingredients (filter design), and for the very extensive computations necessary to the final solution of the problem; the latter has been responsible for the energy distributions, and for the fundamental spectrophotometric data on the solutions, including the miscellaneous studies of Beer's law, temperature effect, permanence, etc.

The authors take pleasure in making the following acknowledgments:

To Hugh K. Clark for general assistance throughout the investigation, this consisting of extensive computations and reduction of data, preparation of graphs for publication, and preparation of chemicals and solutions; to E. Wichers for advice regarding copper and cobalt

salts; to A. Isaacs for the chemical analysis of such salts and solutions; to Burt H. Carroll for advice, analysis, and titrations used in the study of pyridine and picoline; to Mabel E. Brown for assistance in spectrophotometric measurements and computations; to Irwin G. Priest for his interest and encouragement in the development of the filters; and to others, as noted at various places in the paper.

II. CHARACTERISTICS OF WHITE-LIGHT FILTERS

1. PREVIOUS INVESTIGATIONS

The problem of absorbing the excess energy of the longer wave lengths in common incandescent illuminants has been a subject of investigation for many years. In general, such research and development have had one of two objects in view.

In one case the problem has been the commercial production and utilization of white light. "Artificial daylight" is a term used to advertise a wide variety of illuminants varying in the color of their light at normal operation from yellower than noon sunlight to the blue of north skylight and produced by bluish "daylight" glass in one form or another used as a filter for the yellow light of the incandescent source. Those interested in the development of this kind of glass should refer to papers by Ives (13), Luckiesh and Cady (14), Brady (15), and Gage (16) and to the bibliographies given in the first two of these papers. Combinations of glass and dyed gelatin were tried in some cases, but the daylight glasses finally evolved have proved satisfactory for general illumination purposes without the use of any additional material. A recent survey of the present status of artificial white-light illumination may be found in a paper by Macbeth (17). The filters described in the present paper were not designed for general illumination but for scientific purposes. They will, however, serve in an excellent manner as standards for the testing and calibration of daylight glass and artificial white-light illumination.

In the other case the problem has been the development of permanent and reproducible filters for laboratory use in photographic sensitometry, colorimetry, or photometry. Lacking more suitable filters, daylight glass has often been used for scientific purposes. It is not sufficiently reproducible, however, for many kinds of work and the spectral distribution which it gives often deviates more widely from the ideal than is desirable. Various other attempts have been made to devise filters which could be prepared in the laboratory from specification, and which would be satisfactory as regards permanence, convenience, and spectral distribution. The following brief résumé will illustrate what has been attempted and what is available along this line.

One of the earliest attempts, apparently, is that of Mees and Sheppard (18, p. 292; 19, p. 314) who devised filters for converting the energy distribution of the acetylene flame to that of daylight. Their method for designing the filters was to take photographs on panchromatic plates in a Tallent spectroscope, altering the composition of the filter placed in front of the acetylene burner until the

resulting energy gave, as nearly as possible, the same curve as daylight. Their final filter was 1 cm each of the following two solutions:

A.	Gentian violet	g	0.002
	Acid green	g	.001
	Mandarin orange	g	.001
	Rose bengal	g	.004
	Water to make	cc	100
B.	Copper acetate	g	1.500
	Water to make	cc	100

This filter was apparently unsatisfactory as regards stability and the accuracy with which it reproduced the ideal spectral distribution. A later filter (20) is noted in which only dyes are used and which was considered permanent. The composition of the filter is not given, however, and the practical difficulties of preparing the filter were said to be considerable.

The present Wratten filters (21), made of dyed gelatin—No. 78 "Tungsten to daylight (visual)," No. 78AA "Tungsten (Mazda C) to daylight," and No. 79, "Acetylene to daylight"—are doubtless the result of this earlier work. They are classed as "stable" or "moderately stable." Illustrations of the energy distributions obtainable with these filters are given below. (Fig. 1.)

A dyed gelatin filter for converting 2,200° to 5,000° K. has been described by Naumann (22). In its final form, superseding previously published formulas, the filter is composed as follows:

	Grams
Thionine blue	0.18
Methylene blue	.08
Crystal violet	.12
Fast red D	.03
Rapid filter red I	.04
Orange II	.14
Tartrazine	.03

These dyes are dissolved separately and the solutions mixed and prepared for coating by addition of gelatin. The formula is designed to cover 1 m^2 of filter surface. Small scale density curves of this and the author's previous filters are given, but, otherwise, no quantitative data are supplied from which one could derive the spectral energy distribution or color of this filter. It is said to be somewhat bluer than daylight in spite of the fact that it was intended to duplicate 5,000° K. Questions of reproducibilityand permanence are not adequately considered.

A different type of filter, developed by Pfund, is described by Sinden (23) as a liquid-cell having two distinct compartments, one of which contained an aqueous solution of ammoniacal copper sulphate and the other an aqueous solution of the sulphates of copper and cobalt. The filter was used with a lamp operated at 3,000° K., and an actual comparison showed that this combination gave no appreciable color difference from noon sunlight at Baltimore, in April. Its energy distribution compared with that of the O. S. A. average noon sunlight is briefly illustrated. No details regarding proportions of the ingredients were given. It would appear that this filter was designed by trial and error procedure and the exact proportions have apparently never been published.

More recently Guild (24) has described two tentative filters devised by his colleague Young, using materials similar to those in Pfund's filter. One of these is for converting 3,260° to 3,000° K. and the

other for converting 3,000° to 5,000° K. for use as a standard white for colorimetry. Their compositions are given as follows:

2,360° to 3,000° K:			
Solution A—	Copper sulphate	g	1. 41
	Ammonia (density 0.90)	cc	141
	Distilled water to	cc	1, 000
Solution B—	Copper sulphate	g	11
	Cobalt sulphate	g	8. 5
	Distilled water to	cc	1, 000
3,000° to 5,000° K:			
Solution A—	Copper sulphate	g	2. 30
	Ammonia (density 0.90)	cc	230
	Distilled water to	cc	1, 000
Solution B—	Copper sulphate	g	13
	Cobalt sulphate	g	12
	Distilled water to	cc	1, 000

Solutions A and B are contained in separate compartments of a double cell, of which the walls are of optical glass having very high and as nearly as possible nonselective transmission. The thickness of each solution is 10.00 mm. It is emphasized that the figures given are provisional and that some modification must be looked for in a subsequent publication, when "full particulars of the actual energy distribution, etc., will be given." Such publication has apparently not yet appeared.[4] Guild's filter is, however, now being sold by Hilger (25) under the name of "Standard artificial daylight."

In 1918 Priest (2) proposed the use of a rotatory dispersion filter for obtaining artificial noon sunlight from the gas-filled tungsten lamp operated at 15.6 lumens per watt. He shortly after this changed the specification of the source to a lamp at a color temperature 2,848° K., this being a more precise form of specification having the same intent. The resulting energy distribution has been designated as "Abbot-Priest sunlight." Since this proposal the filter has been constructed and in service at the National Bureau of Standards for some time. The filter consists (3) "of a crystalline quartz plate between 'crossed nicols,' the plate being 0.500 mm thick and placed so that the path of light is parallel to the optic axis. These parts are built in a cell which is inserted as a unit near the eye in the optical train of the colorimeter. The effect is that of a blue filter having a spectral transmission such that the spectral distribution of energy delivered to the observer's eye approximates very closely to average natural noon sun at Washington." The desired reproduction was considered more precise than had previously been obtained by blue glass, dyed gelatin filters, or solutions of selective absorbents.

Artificial white-light filters available commercially include:

1. The Wratten No. 78, No. 78AA, and No. 79, as noted above, manufactured by the Eastman Kodak Co. These filters are composed of dyed gelatin cemented between optical glass flats.

2. Guild's filter as sold by Hilger.

3. Daylight glass, particularly the Luckiesh "Trutint" and the Corning "Daylite" glass. The thickness of glass necessary to obtain the nearest approximation to the desired energy distribution

[4] A partial revision of the specification has, however, been made and is given in the Appendix to a paper by W. D. Wright, A Redetermination of the Trichromatic Coefficients of the Spectral Colors, Trans. Opt. Soc., 30, pp. 141-164; 1928-29. The auxiliary filter for rating the gas-filled lamp is unchanged, but the color temperature obtained when it is used with a 2,360° K. source is now "believed to be very close to 2,900° K." The "white light" filter has been changed in that the copper and cobalt sulphates in solution B are now given as 15 g each instead of the 13 and 12 g above given. The color temperature obtained with this filter (with source calibrated by means of the auxiliary filter) is now considered to be "approximately 4,800° K."

depends, of course, upon the color temperature of the source to be used with the glass as well as upon the spectral transmission characteristics of the glass itself. Corning "Daylite" glass has been used by Davis in his work on photographic sensitometry (26). The particular sample used converted the color of a gas-filled lamp operated at 2,810° K. to that of noon sunlight.

The spectral distributions which may be obtained by some of the foregoing filters are shown in the accompanying figures. Figure 1 shows the computed energy distributions of the three Wratten filters, based on the published spectral transmission data (21) for them. With Wratten No. 78 the energy for 2,360° K. was used, with No. 78AA that for 2,900° K., and with No. 79 the energy data of Coblentz (27) for acetylene. Mean sunlight,[5] daylight, and 5,000° K. are also included for comparison.

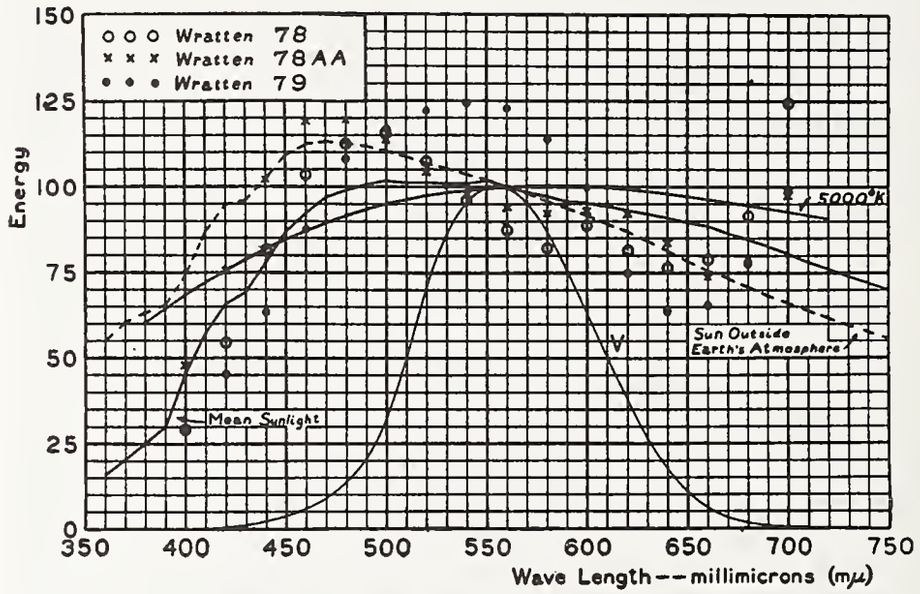


FIGURE 1.—*Relative energy distributions obtained with Wratten filters as indicated, compared to 5,000° K., mean sunlight, and sun outside earth's atmosphere (which approximates average daylight)*

With No. 78, the energy for 2,360° K. was used; with No. 78AA, that for 2,900° K; and with No. 79, the energy distribution of acetylene was used.

In Figure 2 is shown the relative energy distribution obtained with the filter tentatively adopted by the Sixth International Congress of Photography in 1925. The data were computed using a 2,360° K. source (black-body distribution) and a spectral transmission obtained by the writers on a solution made up in accordance with the formula already given using 36 cc of ammonia. Also on the graph are plotted 5,000° K. (the ideal in this case) and mean sunlight for comparison.

Figure 3 illustrates Abbot-Priest sunlight as calculated (by Priest) for the rotatory dispersion filter. For comparison mean sunlight and 5,000° K. energy curves are included.

In Figure 4 are shown the energy distributions obtained with two samples of Corning "Daylite" glass from different melts. The thick-

[5] The expression "mean sunlight" (or "mean sun") refers specifically in all cases to a certain energy distribution, or its resulting color, as defined in Sec. III, 1.

nesses of the glass and the color temperatures of the source are noted. The spectral transmission measurements were made at the bureau.

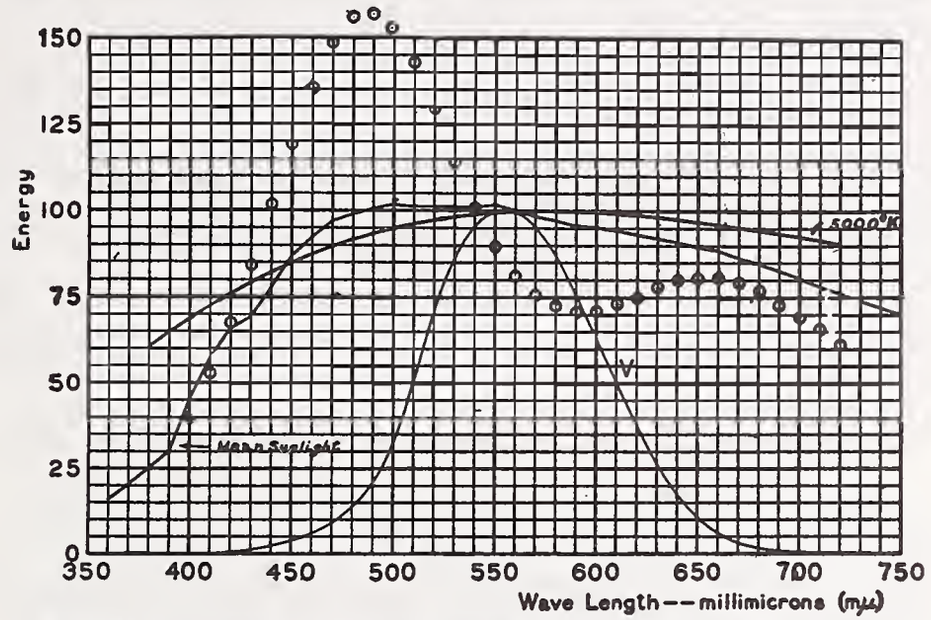


FIGURE 2.—*Relative energy distribution obtained with 2,360° K. source and filter tentatively proposed by the Sixth International Congress of Photography (shown by circles) compared with mean sunlight and 5,000° K.*

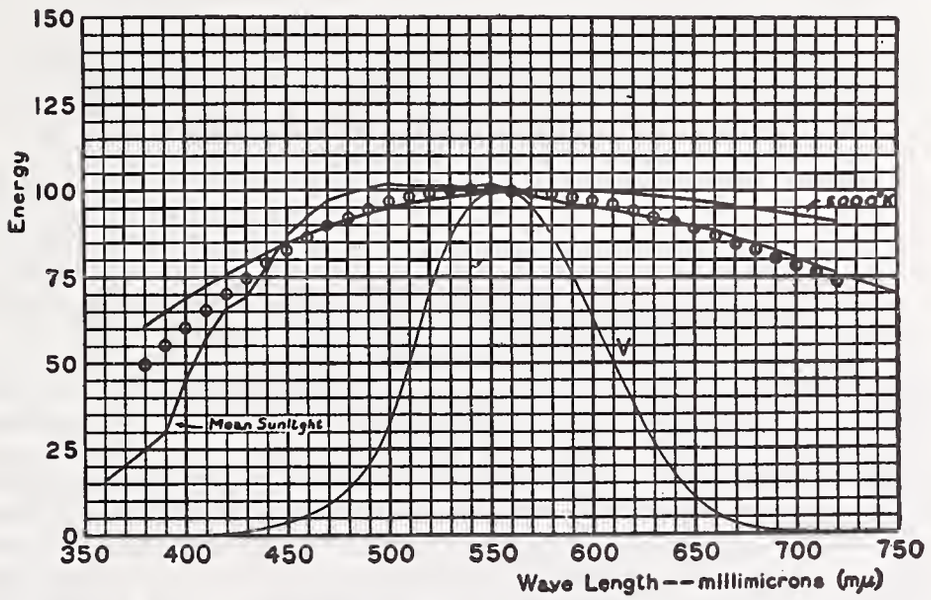


FIGURE 3.—*Relative energy distribution of Abbot-Priest sunlight (shown by circles) compared with mean sunlight and 5,000° K.*

For comparison the curves for mean sunlight and 5,325° K. are also given.

2. DESIRABLE FEATURES

In designing filters for scientific and technical purposes, such as the production of standards of light quality, the following properties would be desirable: (1) Accurate reproducibility from specification, (2) spectral transmissions which are unaffected by time and temperature, (3) spectral transmissions such that the energy distribution obtained is (*a*) exactly that sought and (*b*) an exact color match therewith regardless of whether or not (*a*) is fully satisfied, (4) the highest transmission of light consistent with (3), and (5) other physical properties (size, shape, durability, etc.) which facilitate use of the filters.

It is not to be expected that filters can be constructed which will embody all these properties. No known absorbing materials exist, for example, by means of which it would be possible to get (3 (*a*)) a

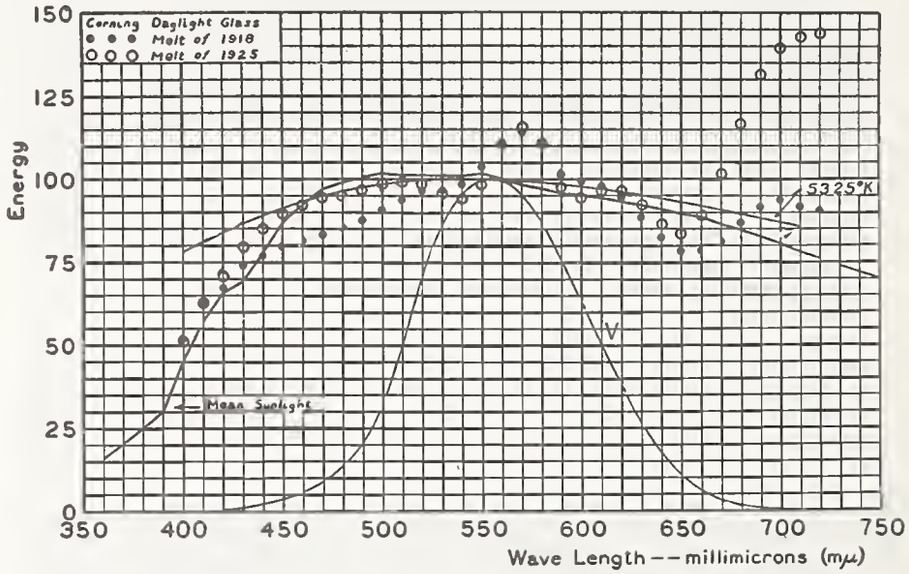


FIGURE 4.—*Relative energy distribution obtained with Corning daylight glass as indicated compared with mean sunlight and 5,325° K.*

The thicknesses of glass and color temperatures of source were: Melt of 1918, 3.59 mm and 2,900° K.; melt of 1925, 4.75 mm and 2,810° K.

perfect match in energy distribution. Careful consideration must be paid to the special purposes of a given filter before the sort of compromise which is most advantageous can be determined. A compromise which would ordinarily be advantageous is to adjust the unavoidable deviations in energy distribution so as to obtain (3 (*b*)) an accurate color match and, at the same time, (4) a reasonably high transmission of light. Likewise, the lack of an absorbing medium, strictly permanent and completely independent of temperature, is not practically important if a reasonably permanent medium is available having known temperature coefficients of spectral transmission, because such variations in spectral transmission as may occur can be taken into account. It should be obvious that the minimum requirement (in the past often unfulfilled) in this respect is that the spectral transmission of the filter be accurately known, including the effects of age and temperature.

From the standpoint of convenience and permanence probably nothing is superior to the various daylight glasses. However, the resulting energy distributions depart, in general, to an undesirable extent from the ideal sought, and, as already illustrated in Figure 4, glass from different formulas or different melts from the same formula vary considerably in this respect. Even the best samples show excess transmission of energy in the yellow-green and usually also in the red. There seem to be no data available regarding changes in transmission with temperature, although cobalt blue glasses do not have very large temperature coefficients over most of the visible spectrum. As ordinarily used with incandescent light, the transmission is reasonably high (10 to 20 per cent). While satisfactory in many respects then, the principal objection to daylight glass is that it can not be accurately reproduced from specification. This outstanding difficulty, together with its spectral irregularities in the yellow-green and red, would seem sufficient to prevent the general adoption of any of the known daylight glasses as working standard filters for the production of white light.

Dye solutions and dyed gelatin films likewise can not be considered reproducible from specification, the dyes themselves being, in general, not accurately reproducible in their spectral characteristics. Furthermore, the absorbancy of dyes (coloring strength) is, in general, so great that it is difficult to make weighings with the desired precision. When spectrophotometric data have been given to an extent sufficient for reliable computation, the resulting energy distributions have been hardly as good as those obtained with daylight glass. A certain stability is claimed for the Wratten filters, but the effects of temperature variation are not given. The use of dyes can not, in general, be recommended, inasmuch as in solution they are not as satisfactory as solutions of inorganic salts (see below) and in dyed gelatin films are not as satisfactory as daylight glass.

The quartz-nicol rotatory dispersion filter described in the previous section is reproducible from specification (although the spectral selectivity of the nicol prisms has been ignored in the computation) and its spectral transmission is accurately known (subject to the same limitation). Its reproduction of sunlight is very satisfactory from a colorimetric standpoint, but its relatively high transmission in the violet and ultra-violet renders it less suitable for photographic purposes. Its light transmission is very low (about 1½ per cent), and the small dimensions of nicol prisms limit its usefulness practically to the eyepiece of optical instruments. Its temperature coefficient is not given.

Solutions may not at first be considered very convenient to use, but, on the other hand, liquid filters are more readily prepared in the laboratory than any other type. Once prepared and inclosed in a suitable glass cell, they are not inconvenient to handle. The extent to which the desired energy distributions are attained with our liquid filters is brought out in the present paper, together with data regarding reproducibility, permanence, change with temperature, etc. The light transmissions compare very favorably with those of other filters.

It is apparent from the brief discussion given in this and the previous sections, considering the variety of materials and methods which have been used and the degrees of approximation to the desired energy distributions which it has been necessary to accept, that the design

of satisfactory white-light filters is not a simple undertaking. The reason is not hard to find. No single absorbing material has been found which has the desired spectral characteristics and the number of those which even crudely approximate the desired selectivity is apparently limited.

In general, the first approximation has been secured by the use of some blue coloring material; and additional materials have then been added, to the extent considered necessary, to correct for spectral deficiencies of the principal material. Considering the type of change that occurs in the spectral transmission as the concentration or thickness is varied (see Sec. V, 1), it is obvious that the difficulty has been not merely in finding suitable materials but in adjusting the concentrations of these materials so as to secure the most suitable transmission and resulting energy match.

In the present investigation the problem has been not only that of selecting materials and adjusting them so as to secure the best spectral energy match, but also that of securing exact color matches. This has been done, furthermore, not merely with one filter but with a large number of them, as already outlined in Section I. It is hoped that this work will satisfy the demand felt in many lines of research for reproducible white-light filters which have satisfactory spectral characteristics and which may be readily prepared and conveniently used in the laboratory.

III. SPECTRAL ENERGY DISTRIBUTIONS

Before describing the materials and methods of computation which have been used in the development of the present filters, it is necessary to define the various energy distributions which enter into these computations. These include sunlight and daylight energies, various black-body distributions, and that of the acetylene flame.

1. NOON SUNLIGHT AT WASHINGTON

Sunlight distribution is of importance to photographic sensitometry, not because it may approximately represent a colorimetric or subjective white, but because it represents, perhaps better than any other single energy distribution, the average conditions [6] under which the great majority of photographic negative materials are exposed.

(a) ABBOT'S DATA

Extensive data are available on the spectral energy distribution of noon sunlight at Washington. Such data representative of average atmospheric conditions at the two extremes, the summer and winter solstices, were furnished the National Bureau of Standards by Doctor Abbot, in a letter dated November 7, 1917. These data have served as a basis in the development of Abbot-Priest sunlight as already described. The two sets of data—average noon sunlight at Washington on June 21 and on December 21—are given in Table 1 and illus-

[6] Varying from the incandescent sources often used in indoor motion-picture photography through sunlight to daylight or blue sky,

trated in Figure 5. The difference between the two sets is of the same order of magnitude as the variation at noon due to changes in atmospheric transmission at a given time of year, and the data do not, of course, at all represent early morning and late afternoon sunlight, which is relatively yellow.

An arithmetical mean of these two sets of noon sunlight energy data, given in Table 1 and illustrated in Figure 5 and elsewhere, was

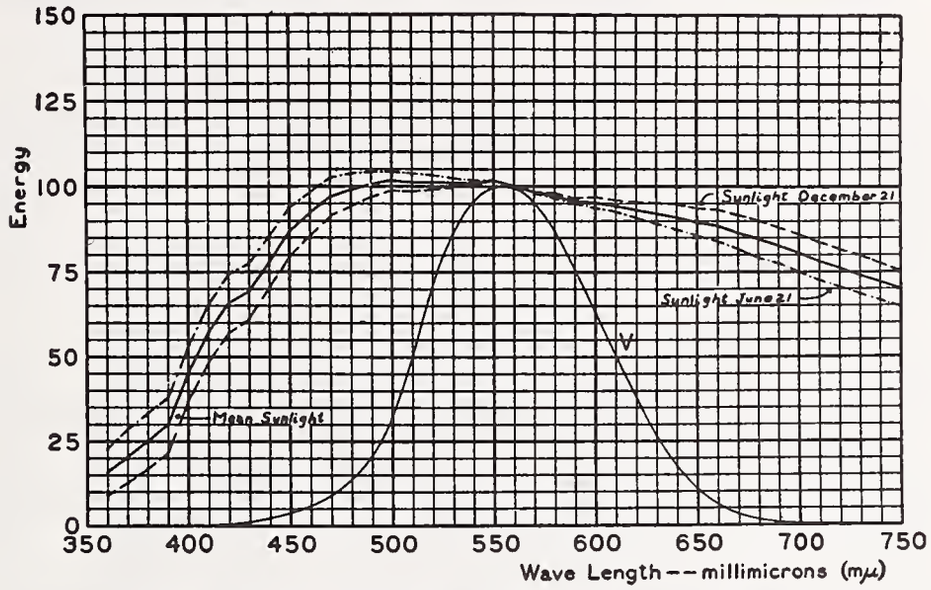


FIGURE 5.—*Relative energy distribution of noon sunlight at Washington as given by Abbot's data of 1917 (Table 1 of the present paper)*

chosen by the writers as the ideal energy distribution, the duplication of which has been attempted in the first series of filters described herein. It is hereinafter often referred to simply as "mean sun" or "mean sunlight"; and whenever these expressions are used in this paper this particular energy distribution (or color) is meant. The average noon sunlight given in the report (4, p. 563) of the Colorimetry Committee of the Optical Society of America is essentially the same as herein adopted, except below 420mμ, where an error has apparently entered in the average data used in that report. The data as used by the writers extend from 360mμ to 720mμ.

Table 1.—Sunlight energy distribution

Wave length in millimicrons	Relative energy distribution of sunlight					
	Outside earth's atmosphere[1]	At Washington, D. C., at noon, average atmospheric transmission		Mean sun[2]		Sun[2] outside atmosphere reduced to 100.00 at 560mμ
		June 21[1]	Dec. 21[1]	Arithmetical mean of data for June 21 and Dec. 21	Reduced to 100.00 at 560mμ	
360	59.1	24.0	9.4	16.70	15.97	55.34
370	64.5	30.0	12.8	21.40	20.47	60.39
380	67.6	35.0	17.4	26.20	25.06	63.30
390	69.7	40.4	22.5	31.45	30.08	65.26
400	79.1	56.0	38.5	47.25	45.19	74.06
410	93.0	69.3	50.6	59.95	57.34	87.08
420	101.5	78.4	59.3	68.85	65.85	95.04
430	102.8	81.5	63.2	72.35	69.20	96.25
440	108.9	89.7	72.9	81.30	77.76	101.97
450	117.0	99.1	82.5	90.80	86.85	109.55
460	120.0	103.9	89.0	96.45	92.25	112.36
470	121.2	107.7	95.0	101.35	96.94	113.48
480	120.6	109.2	97.9	103.55	99.04	112.92
490	119.4	109.8	100.5	105.15	100.57	111.80
500	117.7	110.2	102.6	106.40	101.77	110.21
510	116.0	109.2	102.3	105.75	101.15	108.61
520	114.3	108.7	102.9	105.80	101.20	107.02
530	112.6	107.8	103.5	105.65	101.05	105.43
540	110.8	107.3	103.7	105.50	100.91	103.75
550	109.1	107.3	105.4	106.35	101.72	102.15
560	106.8	105.4	103.7	104.55	100.00	100.00
570	104.6	103.4	102.3	102.85	98.37	97.94
580	102.3	101.8	101.4	101.60	97.18	95.79
590	100.0	100.0	100.0	100.00	95.65	93.63
600	97.7	98.8	100.2	99.50	95.17	91.48
610	95.2	97.5	99.6	98.55	94.26	89.14
620	92.4	95.7	99.1	97.40	93.16	86.52
630	89.7	94.0	98.7	96.35	92.16	83.99
640	86.8	92.1	98.2	95.15	91.01	81.27
650	83.9	90.2	97.3	93.75	89.67	78.56
660	81.2	88.5	96.6	92.55	88.52	76.03
670	78.4	85.9	94.7	90.30	86.37	73.41
680	75.8	83.6	93.5	88.55	84.70	70.97
690	73.2	81.5	91.4	86.45	82.69	68.54
700	70.5	79.1	89.3	84.20	80.54	66.01
710	68.1	76.5	86.7	81.60	78.05	63.76
720	65.8	74.3	84.8	79.55	76.09	61.61
730	63.6	72.0	82.6	77.30	73.94	59.55
740	61.3	70.0	80.5	75.25	71.98	57.40
750	59.4	67.9	78.5	73.20	70.01	55.62

[1] These values were submitted to the National Bureau of Standards by Dr. C. G. Abbot in letter of Nov. 7, 1917.
[2] These values were computed from Abbot's data by the authors for the purposes of this investigation.

(b) A STANDARD FOR SENSITOMETRY

As previously noted, the American Committee on the Unit of Photographic Intensity originally recommended, and the Sixth International Congress of Photography tentatively adopted in 1925, 5,000° K. as the primary standard of energy distribution for use in photographic sensitometry. The color of 5,000° K. is not greatly different from that of average noon sunlight—it was, in fact, chosen for this reason. In addition, it serves as a precise definition and is a good round number.

As a standard for use in photographic sensitometry, however, the authors believe their adopted mean sun is much preferable to 5,000° K. The reasons are as follows:

1. In photographic sensitometry it is not the color that is primarily important but the energy distribution, particularly in the violet and near ultra-violet where all photographic materials are highly sensitive. In this important region the relative energy of 5,000° K. is not representative of sunlight, as is shown in Figure 6.

2. The energy distribution of mean sun can be duplicated much more closely by means of incandescent lights and filters than can that of 5,000° K. This is well illustrated by a comparison of charts 8 and 24 showing the energy distribution curves obtained with a 2,848° K. source and the respective filters for obtaining mean sun and 5,000° K. Not only are the deviations in the latter case much larger,

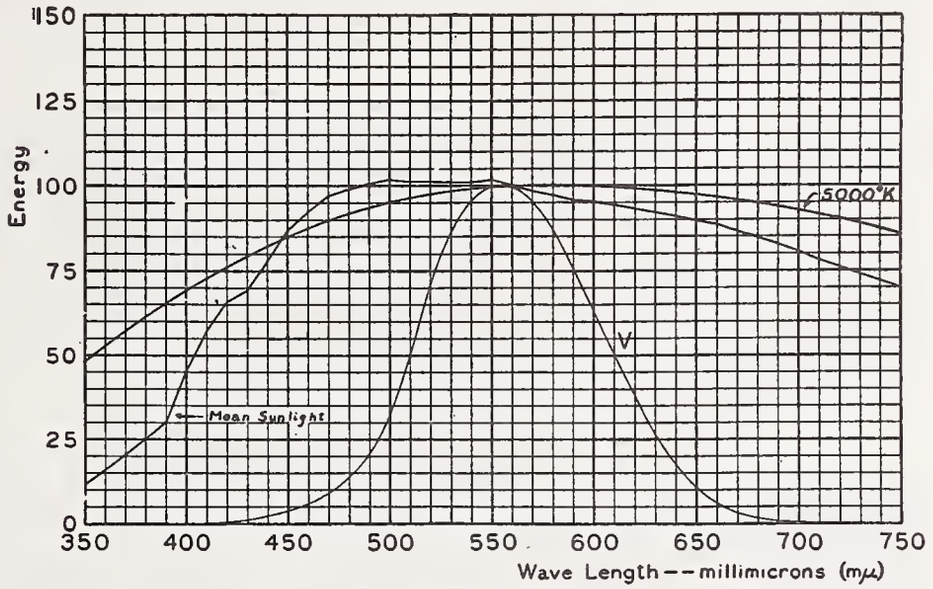


FIGURE 6.—*Relative energy distributions of 5,000° K. and mean sunlight*

This shows that 5,000° K. is not a satisfactory representation of sunlight in the photographically important region below 420mμ.

but the transmission is lower; and this difficulty of reproducing 5,000° K. from any practicable incandescent lamp color temperature, considering both the spectral energy match and the light transmission, is inherent in the differences in the energy distributions, not in the filters employed.

It seems undesirable to choose any ideal standard of energy distribution for use in photographic sensitometry which is less representative of natural sunlight conditions than the adopted mean sun, and which is at the same time less satisfactorily reproducible in the laboratory.

2. SUNLIGHT OUTSIDE THE EARTH'S ATMOSPHERE—A STANDARD OF DAYLIGHT

Certain problems may arise in colorimetry where a daylight filter may be preferred to a sunlight filter. However, the spectral energy distribution of daylight has apparently never been measured as such;

518-743 O - 73 - 43

indeed, there may be question as to the proper definition of daylight, which may show considerable variation in color, depending upon atmospheric conditions, time of day, and other factors.

Priest has previously shown (28; 43, p. 483) that the color temperature (6,000° to 7,000° K.) of the overcast sky is approximately the same as the color temperature of the sun as it would appear outside the earth's atmosphere. Priest's later data (28) also show that, for clear and partly cloudy skies, the color temperature of the combined light of sun and sky received on a horizontal surface exposed to the whole sky lies generally within this same range, 6,000° to 7,000° K., for a wide range of solar altitudes. Taylor's data ("The Color of Daylight," Tab. I, 23d Ann. Con. I. E. S.; 1929) are in substantial accord with Priest's. Kimball (29, p. 514) states (on the basis of Priest's color-match data and other data on spectral distribution): "Therefore, for the total daylight with either clear, cloudy, or partly cloudy skies the energy distribution corresponds closely to the distribution for black-body radiation at a temperature from 5,500° to 6,500° K."

As will be shown later in this paper, we have found by actual trial that the color of the overcast sky may be nearly matched by the color of the sun outside the earth's atmosphere. A filter was designed (Sec. V, 6) to convert 2,848° K. to sunlight outside the atmosphere as specified by Abbot's data of 1917, given in Table 1. At certain times on four different days near the autumnal equinox of 1928 and once during the tests in December, 1928, a perfect color match was observed between this source-and-filter combination and overcast south sky, although on the average the two colors were not identical. Details of these measurements are given in Section VIII, 4.

3. THE PLANCKIAN RADIATOR OR BLACK BODY

(a) AN EQUIVALENT OF INCANDESCENT ILLUMINANTS—COLOR TEMPERATURE

The determination of spectral energy distribution by direct radiometric methods is a matter of considerable experimental difficulty, requiring elaborate apparatus and expert knowledge and technique. The problem is considerably simplified provided one has available a standardized source of accurately known spectral energy distribution; for in this case it is necessary merely to make comparison, wave length by wave length, visually or otherwise, of the unknown source with the standard, and various instrumental calibrations necessary in direct radiometric measurements are eliminated. At best, however, it is a laborious undertaking to make spectral energy determinations by spectrophotometric or spectroradiometric methods.

However, a substitute method is available, involving much less apparatus, time, and skill, in which the spectral energy distribution is derived by determining what is known as the color temperature of the source. The relative energy distribution from a complete radiator or black body is a function only of the temperature of the radiator. The theoretical equation derived by Planck expressing this relation is of the form:

$$E_\lambda = \frac{C_1 \lambda^{-5}}{e^{\frac{C_2}{\lambda\theta}} - 1}$$

where

E_λ is the energy per unit wave length at wave length λ,
θ is the absolute temperature in ° K. (° C.+273),
e is the base of natural logarithms, and
C_1 and C_2 are constants.

It has been found by experiment that the relative energy distribution from many common incandescent light sources in the visible spectrum is practically the same as that from a Planckian radiator, provided the temperature of the latter is taken to give an accurate color match with the source in question. The color temperature of a light source is defined (30, p. 1180) as "the temperature at which a Planckian radiator would emit radiant energy competent to evoke a color of the same quality as that evoked by the radiant energy from the source in question." A light source accurately allocated on the color temperature scale—that is, having a perfect color match with the Planckian radiator at some temperature—may or may not have an exact relative spectral energy match with this ideal radiator.

The practical advantage of the color-temperature method as a substitute for spectral energy determinations lies in the fact, noted above, that many common illuminants, particularly incandescent electric lamps, not only give a good color match with the Planckian radiator but a very close approximation, at least in the visible spectrum, to the relative energy distribution as well. This approximation is so good that a specification of color temperature serves in practically all cases as an adequate specification of the relative spectral energy values.

Of course, there are certain sources, such as the mercury arc, which can not be described in terms of color temperature, and others, such as the Welsbach mantle, which can be placed only crudely on the color temperature scale, not being a match with any black-body color. However, practically all sources whose light is produced solely by temperature may be quite accurately described in terms of color temperature. Passing over the varying red shades of the furnace, we may note as reference points the approximate color temperatures of various illuminants—the candle and kerosene flames at 1,900° K., the carbon lamp at 2,100° K., the acetylene flame [7] and vacuum tungsten lamp at 2,400° K., and the gas-filled tungsten lamps from 2,600° to 3,100° K. In the case of flames, the values vary for different portions of the flame and for different conditions of pressure, humidity, etc. With the incandescent electric lamps, the values differ with the applied voltage and with the type of lamp, the effective portion of the filament, etc. The melting point of tungsten is near 3,600° K., which is approximately the color temperature of the low power carbon arc. For more complete discussions of color temperature the reader is referred to papers from the Nela Research Laboratory (31) and from the National Bureau of Standards (30), and to references contained therein. An illustration

[7] See Sec. III, 4 for discussion of the acetylene flame.

(fig. 7) from one of the bureau papers (Sci. Pap. No. 443, fig. 1, p. 225) will serve to illustrate the sort of deviations which may be expected in the case of incandescent electric lamps between experimental determinations of spectral energy distribution and values computed from Planck's equation.

Such actual experimental determinations of spectral energy distribution would be all too meager for the needs of the present paper. Furthermore, they are subject to experimental uncertainty, especially at the shorter wave lengths where the energy values are relatively small. It has, therefore, been not only preferable, but necessary, to use the theoretical values given by Planck's equation in

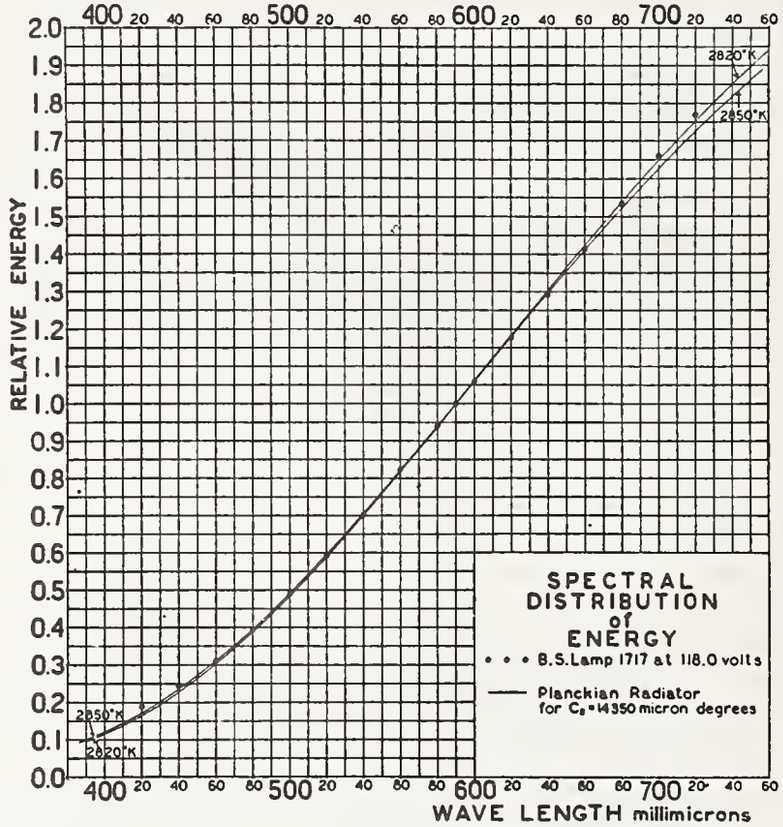


FIGURE 7.—*Relative energy distribution of standard lamp, B. S. 1717, at 118.0 volts, compared with Planckian radiator at 2,820° and 2,850° K.*

This illustration is reproduced from B. S. Sci. paper No. 443, by I. G. Priest.

the computation of the filters to be described, with the exception of the acetylene-to-sunlight filter.

(b) ENERGY DISTRIBUTIONS, 2,000° TO 20,000° K.

Relative spectral energy data computed from Planck's (or Wien's) equation have been published by various investigators. The most extensive of these data [8] comprise Bureau of Standards Miscellaneous

[8] See also a paper by Skogland entitled, "Tables of Spectral Energy Distribution and Luminosity for Use in Computing Light Transmissions and Relative Brightnesses from Spectrophotometric Data," B. S. Misc. Pub. No. 86; 1929. These tables cover the range from 2,000° to 3,120° K., at 20° K. intervals for the energy distributions and at 80° K. intervals for the luminosity data, differences for interpolation being also given. These tables are based on 14,330μ° as the value of C_2.

Publication No. 56 (hereinafter designated as M56), by Frehafer and Snow, entitled "Tables and Graphs for Facilitating the Computation of Spectral Energy Distribution by Planck's Formula." By means of these tables and graphs, Planckian energy distributions over a wave length range from 400mμ to 720mμ may be readily obtained for all color temperatures from 1,000° to 28,000° K. A table is also given by means of which with further computation values may be obtained at wave lengths beyond this range. The value of C_2 used in these computations was 14,350μ°. The graphs are so arranged, however, that values may readily be obtained for any reasonable value of C_2. The accuracy claimed for these computations was in general 0.15 per cent or better for the tables and 0.33 per cent or better for the graphs.

TABLE 2.—*Spectral energy distribution of a Planckian radiator or black body at various temperatures*

Wave length in millimicrons	Relative energy distribution [1] at—							
	2,000° K.	2,100° K.	2,200° K.	2,300° K.	2,360° K.	2,400° K.	2,450° K.	2,500° K.
350	0.48	0.69	0.96	1.31	1.55	1.73	1.97	2.24
360	.74	1.04	1.41	1.86	2.19	2.42	2.73	3.06
370	1.10	1.51	2.00	2.60	3.01	3.31	3.69	4.11
380	1.61	2.15	2.79	3.54	4.06	4.42	4.90	5.41
390	2.29	2.99	3.81	4.74	5.38	5.81	6.40	7.00
400	3.19	4.09	5.10	6.24	6.99	7.51	8.18	8.89
410	4.37	5.45	6.71	8.08	8.94	9.55	10.34	11.15
420	5.88	7.20	8.68	10.26	11.29	12.00	12.91	13.83
430	7.78	9.36	11.08	12.92	14.06	14.85	15.87	16.90
440	10.14	11.98	13.94	16.00	17.28	18.16	19.26	20.38
450	13.01	15.12	17.32	19.60	21.00	21.95	23.13	24.33
460	16.49	18.84	21.26	23.75	25.22	26.25	27.50	28.76
470	20.63	23.22	25.82	28.46	30.03	31.07	32.40	33.71
480	25.52	28.31	31.03	33.79	35.37	36.45	37.80	39.12
490	31.24	34.11	36.92	39.67	41.32	42.39	43.77	45.08
500	37.85	40.76	43.55	46.31	47.88	48.95	50.20	51.45
510	45.40	48.25	50.95	53.60	55.08	56.03	57.25	58.45
520	54.03	56.70	59.13	61.50	62.84	63.67	64.80	65.80
530	63.70	65.90	68.06	70.00	71.18	71.92	72.75	73.60
540	74.49	76.35	77.94	79.35	80.23	80.72	81.35	82.00
550	86.57	87.60	88.52	89.30	89.74	90.04	90.45	90.75
560	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
570	114.5	113.3	112.3	111.3	110.68	110.4	109.9	109.5
580	130.5	127.8	125.4	123.2	122.06	121.2	120.3	119.5
590	147.7	143.2	139.3	135.8	133.78	132.6	131.1	129.7
600	166.3	159.7	153.9	148.8	146.07	144.2	142.1	140.2
610	186.3	177.2	169.5	162.5	158.77	156.4	153.6	151.0
620	207.7	195.8	185.5	176.6	171.96	168.9	165.4	162.1
630	230.2	215.0	202.4	191.3	185.35	181.7	177.3	173.2
640	254.3	235.5	220.0	206.4	199.24	194.7	189.5	184.6
650	279.4	256.8	238.2	221.9	213.36	208.1	201.9	196.2
660	306.0	279.5	256.9	237.7	227.83	221.8	214.6	207.9
670	334.0	302.0	276.0	253.9	242.64	235.5	227.4	219.7
680	362.9	325.5	295.7	270.2	257.36	249.3	239.7	231.2
690	393.1	350.0	316.2	287.1	272.47	263.1	252.3	242.7
700	424.7	376.5	336.8	304.5	287.56	277.2	265.2	254.4
710	456.8	402.0	357.6	321.0	302.64	291.1	277.9	266.2
720	490.1	428.0	378.6	338.0	317.87	305.2	290.7	277.4

(See footnote at end of table.)

TABLE 2.—*Spectral energy distribution of a Planckian radiator or black body at various temperatures*—Continued

Wave length in millimicrons	Relative energy distribution [1] at—							
	2,600° K.	2,700° K.	2,800° K.	2,848° K.	2,900° K.	3,000° K.	3,100° K.	3,200° K.
350	2.84	3.52	4.32	4.75	5.23	6.22	7.36	8.58
360	3.81	4.67	5.64	6.16	6.71	7.92	9.22	10.64
370	5.03	6.07	7.22	7.82	8.50	9.88	11.39	13.00
380	6.52	7.76	9.11	9.79	10.58	12.15	13.86	15.66
390	8.33	9.74	11.30	12.09	12.96	14.74	16.65	18.60
400	10.44	12.10	13.82	14.71	15.68	17.66	19.74	21.87
410	12.91	14.76	16.71	17.69	18.75	20.87	23.10	25.40
420	15.78	17.82	19.94	21.00	22.15	24.44	26.81	29.20
430	19.03	21.24	23.54	24.68	25.91	28.33	30.80	33.30
440	22.71	25.09	27.52	28.73	30.01	32.52	35.05	37.60
450	26.83	29.33	31.86	33.11	34.43	36.99	39.56	42.17
460	31.36	33.97	36.55	37.81	39.16	41.78	44.35	46.89
470	36.38	38.99	41.60	42.87	44.24	46.81	49.35	51.80
480	41.80	44.45	47.03	48.27	49.59	52.08	54.50	56.85
490	47.69	50.25	52.71	53.92	55.15	57.54	59.85	62.15
500	53.99	56.50	58.78	59.90	61.10	63.22	65.35	67.40
510	60.75	63.00	65.08	66.08	67.15	69.11	71.00	72.80
520	67.88	69.85	71.64	72.51	73.40	75.08	76.70	78.20
530	75.37	76.90	78.45	79.16	79.80	81.23	82.45	83.70
540	83.26	84.35	85.45	85.97	86.45	87.43	88.30	89.15
550	91.49	92.10	92.68	92.97	93.25	93.71	94.15	94.60
560	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
570	108.8	108.1	107.4	107.20	106.8	106.3	105.7	105.3
580	117.9	116.4	115.1	114.52	113.8	112.7	111.6	110.6
590	127.1	124.8	122.7	121.80	120.7	119.0	117.3	115.8
600	136.6	133.3	130.4	129.07	127.7	125.3	123.0	120.9
610	146.3	141.9	138.1	136.42	134.6	131.5	128.6	125.8
620	156.1	150.6	145.8	143.64	141.4	137.5	133.8	130.6
630	165.9	159.3	153.5	150.88	148.2	143.4	139.0	135.1
640	175.9	168.0	161.1	158.06	154.9	149.3	144.3	139.8
650	185.9	176.6	168.6	165.08	161.4	155.0	149.2	144.0
660	195.7	185.3	176.1	172.12	167.8	160.5	153.9	148.1
670	205.8	193.6	183.2	178.87	174.0	166.0	158.7	152.1
680	215.6	202.2	190.5	185.52	180.2	171.2	163.2	155.9
690	225.6	210.5	197.5	192.13	186.1	176.3	167.4	159.4
700	235.3	218.6	204.4	198.32	192.1	181.2	171.5	162.8
710	244.9	226.6	211.3	204.49	197.6	185.8	175.3	166.0
720	254.4	234.5	217.7	210.57	203.0	190.1	178.8	168.9

Wave length in millimicrons	Relative energy distribution [1] at—							
	3,250° K.	3,300° K.	3,400° K.	3,500° K.	3,600° K.	3,700° K.	3,750° K.	3,800° K.
350	9.26	9.93	11.40	12.97	14.64	16.43	17.36	18.32
360	11.40	12.18	13.83	15.58	17.44	19.41	20.42	21.45
370	13.86	14.72	16.55	18.49	20.52	22.63	23.73	24.87
380	16.60	17.55	19.56	21.66	23.82	26.13	27.28	28.42
390	19.63	20.66	22.83	25.06	27.42	29.77	31.00	32.23
400	22.97	24.06	26.36	28.73	31.17	33.65	34.90	36.17
410	26.57	27.73	30.14	32.60	35.10	37.67	38.98	40.29
420	30.42	31.65	34.15	36.67	39.23	41.83	43.14	44.45
430	34.56	35.80	38.32	40.90	43.50	46.08	47.41	48.72
440	38.88	40.19	42.73	45.30	47.84	50.35	51.72	53.00
450	43.43	44.73	47.28	49.79	52.30	54.85	56.08	57.40
460	48.15	49.37	51.80	54.38	56.80	59.20	60.45	61.60
470	53.03	54.25	56.65	59.06	61.30	63.65	64.80	65.85
480	58.07	59.20	61.45	63.70	65.85	68.05	69.14	70.15
490	63.21	64.30	66.40	68.47	70.45	72.40	73.34	74.30
500	68.41	69.35	71.25	73.14	74.95	76.70	77.57	78.40
510	73.67	74.50	76.15	77.82	79.30	80.85	81.63	82.30
520	79.01	79.65	81.05	82.39	83.65	84.90	85.56	86.15
530	84.29	84.80	85.90	86.94	87.95	88.90	89.40	89.85
540	89.56	89.90	90.70	91.36	92.10	92.70	93.10	93.35

(See footnote at end of table.)

TABLE 2.—*Spectral energy distribution of a Planckian radiator or black body at various temperatures*—Continued

Wave length in millimicrons	Relative energy distribution [1] at—							
	3,250° K.	3,300° K.	3,400° K.	3,500° K.	3,600° K.	3,700° K.	3,750° K.	3,800° K.
550	94.84	95.00	95.35	95.69	96.10	96.40	96.55	96.70
560	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
570	105.1	104.8	104.4	104.0	103.6	103.2	103.1	102.9
580	110.2	109.7	108.8	108.0	107.2	106.5	106.2	105.8
590	115.1	114.4	113.0	111.8	110.6	109.5	109.0	108.5
600	119.9	118.9	117.1	115.4	113.8	112.4	111.7	111.1
610	124.6	123.4	121.0	118.9	116.9	115.1	114.3	113.4
620	129.0	127.5	124.6	122.1	119.7	117.5	116.5	115.5
630	133.3	131.5	128.2	125.2	122.4	119.9	118.7	117.5
640	137.6	135.6	131.8	128.2	125.0	122.1	120.7	119.3
650	141.6	139.3	135.0	130.9	127.3	124.0	122.4	120.9
660	145.4	142.8	137.9	133.5	129.5	125.7	124.0	122.3
670	149.1	146.2	140.8	135.9	131.4	127.3	125.4	123.6
680	152.6	149.4	143.4	138.1	133.2	128.8	126.7	124.8
690	155.8	152.3	145.9	140.1	134.8	130.1	127.9	125.8
700	158.8	155.0	148.2	142.0	136.3	131.2	128.8	126.6
710	161.8	157.7	150.2	143.6	137.6	132.2	129.7	127.3
720	164.4	160.0	152.2	145.1	138.8	133.0	130.4	127.8

Wave length in millimicrons	Relative energy distribution [1] at—							
	3,900° K.	4,000° K.	4,500° K.	5,000° K.	5,500° K.	6,000° K.	6,500° K.	7,000° K.
350	20.29	22.42	34.32	48.17	63.48	79.83	96.73	113.9
360	23.64	25.84	38.37	52.55	67.85	83.86	100.1	116.5
370	27.14	29.56	42.50	56.84	71.95	87.51	103.1	118.5
380	30.88	33.41	46.69	61.02	75.85	90.82	105.7	120.1
390	34.76	37.36	50.86	65.04	79.43	93.74	107.7	121.2
400	38.72	41.39	54.95	68.91	82.72	96.37	109.4	121.9
410	42.90	45.54	58.97	72.54	85.73	98.58	110.6	122.2
420	47.05	49.70	62.91	76.00	88.50	100.5	111.5	122.1
430	51.30	53.92	66.76	79.23	90.90	102.0	112.2	121.8
440	55.55	58.13	70.48	82.17	93.05	103.3	112.5	121.2
450	59.85	62.26	74.00	84.90	94.93	104.1	112.4	120.1
460	63.90	66.31	77.35	87.42	96.51	104.8	112.2	119.0
470	68.10	70.32	80.47	89.71	97.86	105.1	111.7	117.6
480	72.20	74.20	83.42	91.72	98.87	105.5	111.0	116.1
490	76.15	77.93	86.22	93.49	99.70	105.3	110.0	114.3
500	80.05	81.63	88.78	95.02	100.4	105.0	109.0	112.6
510	83.70	85.11	91.13	96.41	100.7	104.5	107.7	110.6
520	87.25	88.42	93.36	97.50	100.9	103.9	106.3	108.7
530	90.70	91.55	95.25	98.41	100.9	103.1	104.8	106.5
540	93.90	94.50	97.06	99.07	100.7	102.2	103.4	104.5
550	97.00	97.33	98.58	99.65	100.4	101.1	101.7	102.3
560	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
570	102.6	102.4	101.1	100.2	99.33	98.77	98.20	97.75
580	105.2	104.7	102.2	100.3	98.73	97.51	96.41	95.60
590	107.6	106.8	103.0	100.2	97.89	96.03	94.51	93.28
600	109.8	108.7	103.6	100.1	96.90	94.62	92.58	91.05
610	111.8	110.4	104.1	99.57	95.97	93.09	90.75	88.84
620	113.6	111.9	104.6	99.13	94.89	91.63	88.80	86.73
630	115.3	113.2	104.7	98.50	93.68	90.02	86.94	84.50
640	116.8	114.5	104.8	97.92	92.49	88.40	84.96	82.37
650	118.1	115.5	104.8	97.16	91.26	86.71	83.05	80.21
660	119.2	116.3	104.6	96.24	89.98	85.09	81.20	78.15
670	120.2	117.1	104.3	95.35	88.63	83.43	79.33	76.05
680	121.1	117.7	104.0	94.39	87.24	81.83	77.44	74.04
690	121.8	118.2	103.5	93.44	85.88	80.17	75.59	72.11
700	122.4	118.5	103.0	92.33	84.46	78.47	73.80	70.16
710	122.8	118.7	102.5	91.24	82.98	76.82	72.07	68.30
720	123.1	118.7	101.7	90.07	81.54	75.20	70.29	66.43

(See footnote at end of table.)

TABLE 2.—*Spectral energy distribution of a Planckian radiator or black body at various temperatures*—Continued

Wave lengtb in millimicrons	Relative energy distribution [1] at—							
	8,000° K.	9,000° K.	10,000° K.	12,000 °K.	14,000° K.	16,000° K.	18,000° K.	20,000° K.
350	148.1	180.9	211.5	265.6	310.3	347.0	377.6	403.0
360	148.4	178.5	206.2	254.5	293.6	325.7	352.0	373.7
370	148.2	175.7	200.7	243.6	278.1	305.7	328.2	347.1
380	147.6	172.5	195.0	233.0	263.1	287.0	306.3	322.5
390	146.4	169.0	189.1	222.6	248.8	269.4	286.3	300.0
400	145.0	165.3	182.9	212.4	235.1	253.4	267.6	278.9
410	143.1	161.2	177.0	202.7	222.5	237.9	250.4	260.0
420	140.9	157.2	170.9	193.6	210.4	224.0	234.3	242.5
430	138.5	152.9	165.1	184.6	199.1	210.5	219.4	226.2
440	136.0	148.6	159.2	176.0	188.4	198.3	205.6	211.3
450	133.4	144.3	153.4	167.9	178.4	186.6	192.8	197.5
460	130.5	140.0	147.7	160.0	169.0	175.8	181.1	185.0
470	127.5	135.7	142.3	152.7	160.0	165.8	170.1	173.2
480	124.5	131.4	136.8	145.6	151.7	156.3	159.8	162.5
490	121.5	127.2	131.7	138.7	143.7	147.6	150.4	152.5
500	118.3	123.0	126.7	132.3	136.3	139.3	141.7	143.2
510	115.3	119.0	121.8	126.2	129.3	131.6	133.4	134.6
520	112.1	115.0	117.0	120.4	122.7	124.5	125.8	126.6
530	109.0	111.1	112.5	114.9	116.5	117.8	118.7	119.2
540	106.0	107.2	108.1	109.6	110.7	111.5	112.1	112.3
550	103.0	103.6	104.0	104.7	105.1	105.5	105.8	105.9
560	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
570	97.05	96.53	95.98	95.46	95.10	94.81	94.60	94.37
580	94.14	93.14	92.31	91.19	90.49	90.05	89.55	89.22
590	91.36	89.92	88.75	87.16	86.15	85.46	84.91	84.38
600	88.50	86.76	85.21	83.34	82.05	81.14	80.42	79.83
610	85.87	83.66	82.02	79.73	78.17	77.16	76.30	75.64
620	83.24	80.71	78.91	76.20	74.51	73.29	72.40	71.67
630	80.68	77.90	75.78	72.97	71.05	69.74	68.82	67.95
640	78.18	75.21	72.93	69.78	67.74	66.44	65.32	64.49
650	75.73	72.54	70.10	66.83	64.69	63.23	62.09	61.21
660	73.36	69.99	67.51	64.02	61.78	60.26	59.10	58.18
670	71.12	67.55	64.90	61.33	59.01	57.42	56.25	55.29
680	68.83	65.20	62.49	58.77	56.43	54.79	53.61	52.61
690	66.73	62.93	60.12	56.38	53.90	52.28	51.08	50.05
700	64.68	60.75	57.87	54.09	51.61	49.94	48.66	47.68
710	62.67	58.65	55.76	51.89	49.35	47.70	46.41	45.40
720	60.70	56.62	53.70	49.77	47.24	45.55	44.32	43.27

[1] $C_2=14{,}350\mu°$. All values from 350mμ to 390mμ, inclusive, bave been computed from Table 1 of B. S. Misc. Pub. No. 56, Frehafer and Snow, 1925; the remaining values, 400mμ to 720mμ, have been taken from the tables of M56 or read from tbe graphs of M56 when not tabulated, except at 2,360° and 2,848° K., wbich came from previous computation in the colorimetry section. Note, however, tbat the tabulated values of M56 are given consistently to four significant figures, whereas values less than 10 derived from the graphs of M56 can be read only to the second decimal. Since further figures were of no practical importance in tbis investigation, all values less than 10, however derived, were taken only to the second decimal as abulated herein.

The authors of the present paper have used the tables of M56 so far as they were applicable,[9] have read values from the graphs at intermediate temperatures,[9] and by means of the auxiliary table noted above have extended values of relative energy to 350mμ, the region 350mμ to 390mμ being of importance photographically. These values of relative energy are given in Table 2, and serve not only as a record of the actual values used in the present investigation, but as a convenient set of reference tables more extensive than those of M56 from which they were derived.

[9] Slight qualifications to these statements may be noted in the footnote to Table 2.

The following method was used to obtain values of relative energy from 350mμ to 390mμ. A large-scale plot of Table 1 of M56 over the range applicable was made, plotting values of $\frac{E_\lambda}{E_m}$ against the products $\lambda\theta$ as therein tabulated. Values of $\lambda\theta$ were then computed for the values of temperature, θ, given in Table 2 and for values of λ equal to 350mμ, 360mμ, 370mμ, 380mμ, 390mμ, 400mμ, 410mμ, 420mμ, 450mμ, and 560mμ. For any value of θ, thereupon, values of $\frac{E_\lambda}{E_m}$ were read from the large-scale plot for the values of $\lambda\theta$ computed, and then tabulated at the respective wave lengths. This gave a series of relative energy values for a given temperature, θ, which were then reduced to 100.00 at 560mμ by means of the proper factor.

Values thus obtained at 400mμ, 410mμ, 420mμ, and 450mμ were then compared with the values at those wave lengths previously obtained from the tables and graphs of M56. Let E_1 be the values obtained by the authors, as indicated above, from Table 1 of M56, and let E_2 be the values taken from Table 2b of M56 or read from the graphs of M56 which are based directly upon Table 2b. Then $\frac{E_1-E_2}{E_2}\times 100$ represents the percentage deviation of the two sets of energy values. Where the values of E_2 were taken from the tables of M56, the deviations were nowhere greater than 0.17 per cent; and where values of E_2 were obtained from the curves of M56, the deviations were nowhere greater than 0.29 per cent. These figures afford an interesting comparison with the accuracy claimed by the authors of M56, viz, 0.15 and 0.33 per cent, respectively, for the tables and graphs. In table 2 of the present paper the values given at 400mμ, 410mμ, 420mμ, and 450mμ are those obtained from Table 2b and the graphs of M56, not those derived from Table 1 of M56 for comparison purposes.

(c) USE OF HIGH COLOR TEMPERATURES

It has been demonstrated by means of Priest's rotatory-dispersion colorimetric photometer (30) that the various phases of sunlight and daylight ranging from the yellow of morning and afternoon sun to the blue of a clear sky may be approximately matched in color by light having a Planckian spectral distribution. The colors of the stars may also be described in terms of the color temperature scale. Furthermore, the true subjective white may be found to lie at some place along this high color temperature scale.

It has, therefore, seemed desirable to design filters for duplicating black-body colors, beginning at the color temperatures above which bare Mazda C lamps can not be operated and extending up to 10,000° or 20,000° K., the color of blue sky. The energy distributions referred to in the last section have served as ideals in the design of such filters.

4. THE ACETYLENE FLAME

The acetylene flame has been used extensively in photographic sensitometry and a filter has, therefore, been designed (Sec. V, 6) for converting the spectral energy and color of this light to those of mean sunlight. For this purpose the relative spectral energy data of Coblentz (27) have been used. These data were not given regularly at

every 10mμ. They were accordingly plotted and a curve drawn carefully through these experimental values. Values at every 10mμ were read from this curve and then adjusted by an appropriate factor giving 100.00 at 560mμ. In Table 3 are given the values of relative energy as thus derived from Coblentz's data, and the same values reduced to 100.00 at 560mμ.

The color temperature of the acetylene flame with an E. K. standard burner is commonly taken as 2,360° K., as determined by Hyde, Forsythe, and Cady (32). The value computed from Coblentz's energy data is, however, 2,413° K.[10] Some preliminary tests of the color temperature of this flame by the authors also gave a value over 2,400° K. This question is discussed further in Section VIII, 6.

TABLE 3.—*Spectral energy distribution of an acetylene flame*

Wave length in millimicrons	Relative enorgy distribution		Wave length in millimicrons	Relative energy distribution	
	Coblentz's data	Reduced to 100.00 at 560mμ		Coblentz's data	Reduced to 100.00 at 560mμ
380	3.9	9.09	600	[1] 62.5	145.69
390	4.4	10.26	610	67.8	158.04
			620	[1] 73.3	170.86
400	[1] 5.0	11.66	630	79.1	184.38
410	5.7	13.29	640	[1] 85.0	198.14
420	6.5	15.15			
430	7.4	17.25	650	[1] 91.2	212.59
440	[1] 8.5	19.81	660	[1] 97.6	227.51
			670	104.2	242.89
450	[1] 10.0	23.31	680	[1] 110.9	258.51
460	[1] 11.8	27.51	690	117.4	273.66
470	13.8	32.17			
480	16.2	37.76	700	[1] 124.1	289.28
490	18.5	43.12	710	130.7	304.66
			720	[1] 137.5	320.51
500	[1] 20.9	48.72			
510	24.1	56.18	425	[1] 7.0	----------
520	[1] 27.5	64.10	475	[1] 15.0	----------
530	31.0	72.26	525	[1] 29.2	----------
540	[1] 34.6	80.65	575	[1] 49.8	----------
			625	[1] 76.1	----------
550	[1] 38.9	90.68	675	[1] 107.5	----------
560	[1] 42.9	100.00			
570	47.5	110.72			
580	[1] 52.2	121.68			
590	57.3	133.57			

[1] These values are taken from Table 1 of B. S. Sci. Paper No. 362, W. W. Coblentz, "Distribution of Energy in the Spectrum of an Acetylene Flame," 1920. The remaining values were derived by the authors from a large-scale plot of Coblentz's data.

IV. COMPONENT MATERIALS OF FILTERS

The several desirable characteristics of white-light filters have been indicated in Section II, 2, and it was there noted that certain solutions of inorganic salts apparently had none of the serious faults to which the other materials were subject and which would prevent their adoption and use as general working standards. Obviously, filters for converting lower to higher color temperatures should have little absorption in the violet with increasing absorption toward the red end of the spectrum. Such filters will, of course, have a bluish color. Several blue solutions are available, such as aqueous nickel sulphate and ammonia, aqueous copper sulphate and ammonia, and

[10] As computed by Davis' method, see footnote to Table 13.

others, but they only crudely approximate the desired spectral absorption. An important step toward the solution of the problem was made by Pfund (23) in the discovery that the spectral deficiencies in the absorption of an ammoniacal copper-sulphate solution could be corrected to a large extent by an aqueous solution of cobalt sulphate and copper sulphate. Considerable interesting information regarding the characteristics of various solutions and glass cells tested for photometric use is given in papers (33, 34) by Ives and Kingsbury.

1. CHOICE OF MATERIALS

(a) AQUEOUS AMMONIA SOLUTION OF COPPER SULPHATE

At the start of the investigation considerable time was spent designing filters, by trial and error and by computation from spectrophotometric data, with the component materials used in Pfund's filter. It was found that the transmittancy curve of the ammoniacal copper-sulphate solution changed considerably with the ammonia content for a given concentration of copper sulphate. There were shifts in the wave lengths of the maximum and minimum transmittancies with changes in ammonia content, accompanied by increased values of transmittancy at some wave lengths and decreased values at others. From the standpoint of the resulting energy distribution a content of about 100 cc of ammonia water (28 per cent NH_3) per liter of solution was determined as the most suitable.

When the copper sulphate is varied, keeping the ammonia-water content constant at 100 cc per liter of solution, Beer's law was apparently obeyed over a range from 2 to 5 g of copper sulphate, but failed to some extent at 1 g. This question of obedience to Beer's law is, of course, important when a series of filters is designed by computation from standard spectrophotometric data obtained at some one concentration, if it has been chosen to vary concentration rather than thickness. (The data on Beer's law were rather meager, and the measurements were not continued because of the decision to use pyridine instead of ammonia.)

Ammonia water has two well-known characteristics which tend to make its use for filters unsatisfactory, and which led to the search for a substitute. These undesirable features are: (1) Its corrosive action on glass. It attacks the glass of cells and bottles. In addition to this undesirable property, particularly with respect to its effect on the optical glass plates of the filter cells, it was thought that this action might be accompanied by changes in the absorption of the solution. (2) Its tendency to lose ammonia. Ammonia water must be tested for strength before using. Furthermore, if the container is inadequately stoppered or sealed, there is danger of ammonia gas escaping from the solution, with consequent change in absorption as noted above. Ives and Kingsbury (33, p. 799) speak of the change of color caused by the "unpreventable evaporation" of ammonia.

(b) AQUEOUS PYRIDINE SOLUTON OF COPPER SULPHATE WITH MANNITE—SOLUTION A

As a substitute for ammonia, pyridine (C_5H_5N) was tried. This gives a blue solution with copper sulphate resembling that obtained with ammonia. Spectrophotometric tests of the solutions showed

absorption changes with variation of pyridine content, as was the case with ammonia.

However, a very serious defect of the copper-sulphate-pyridine solution soon developed. It was found to be rather unstable. This instability varied considerably with the pyridine content, the solutions being most stable with a large proportion (20 per cent total volume) of pyridine and poorest with small proportions. The best proportion from the standpoint of the final energy distribution was judged to be about 30 cc of pyridine per liter, but with this amount the solution showed signs of precipitation after standing a few days in very warm weather (35° to 40° C.).

Nevertheless, these preliminary tests showed the use of pyridine to be so promising that some method was sought whereby the copper-sulphate-pyridine solution could be made stable. During a series of experiments having this end in view, it was found that the addition of glycerine to the solution retarded the precipitation enormously. When precipitation with glycerine present did start, however, a scum appeared on the surface of the solution and on the sides of the vessel. (In these experiments just barely enough pyridine was used to convert the copper sulphate to the soluble copper-pyridine compound so that precipitation started very shortly after the mixture was made. The temperature has a very marked influence on the rate of precipitation.) This increased stability with glycerine was so striking that a search was made for other materials of similar structure to glycerin, with the result that mannite (*d*-manitol) was suggested. This material proved to be very satisfactory, and solutions over four years old show no precipitation. A very slight change in absorption does take place upon long standing, however, and this feature is discussed in Section VII, 3. The quantity of mannite used is an amount equal to the weight of copper sulphate used. Apparently no change in the transmittancy of the copper-sulphate-pyridine solution is caused by the addition of mannite.

The copper-sulphate-pyridine-mannite solutions showed the same changes in absorption with changes in the pyridine content as were exhibited by the simple copper-sulphate-pyridine solutions, analogous to the changes in the ammonia solutions as already noted. The data illustrated in Figure 8 show the effects of varying the pyridine content over the range from 10 to 100 cc per liter of solution with a constant quantity (3 g) of copper sulphate. Further investigation (Sec. VIII, 3) confirmed the previous conclusion that 30 cc of pyridine per liter gave a more suitable type of absorption than larger or smaller proportions. An aqueous solution containing, unless otherwise stated, 30 cc of pyridine per liter and having varying amounts of copper sulphate (and mannite) has been designated as "solution A" and is usually thus referred to throughout the rest of the paper.

Having determined the most desirable amount of pyridine to use, a study of Beer's law was made over the range of concentrations of copper sulphate (and mannite) necessary for the filters. A slight failure was found which is described in detail in Section VII, 6. The method of making the Beer's law corrections in the computations is illustrated in Section V, 3 and in Table 4.

The comparative advantages and disadvantages in the use of aqueous ammonia or pyridine solutions of copper sulphate may be summarized as follows:

1. In the most suitable proportions as determined spectrophotometrically, 30 cc of pyridine and 100 cc of ammonia per liter, the ammonia solutions will attack glass while the pyridine solutions will not.

2. Ammonia gas readily escapes from ammonia water. This difficulty is not encountered with pyridine, the boiling point of which

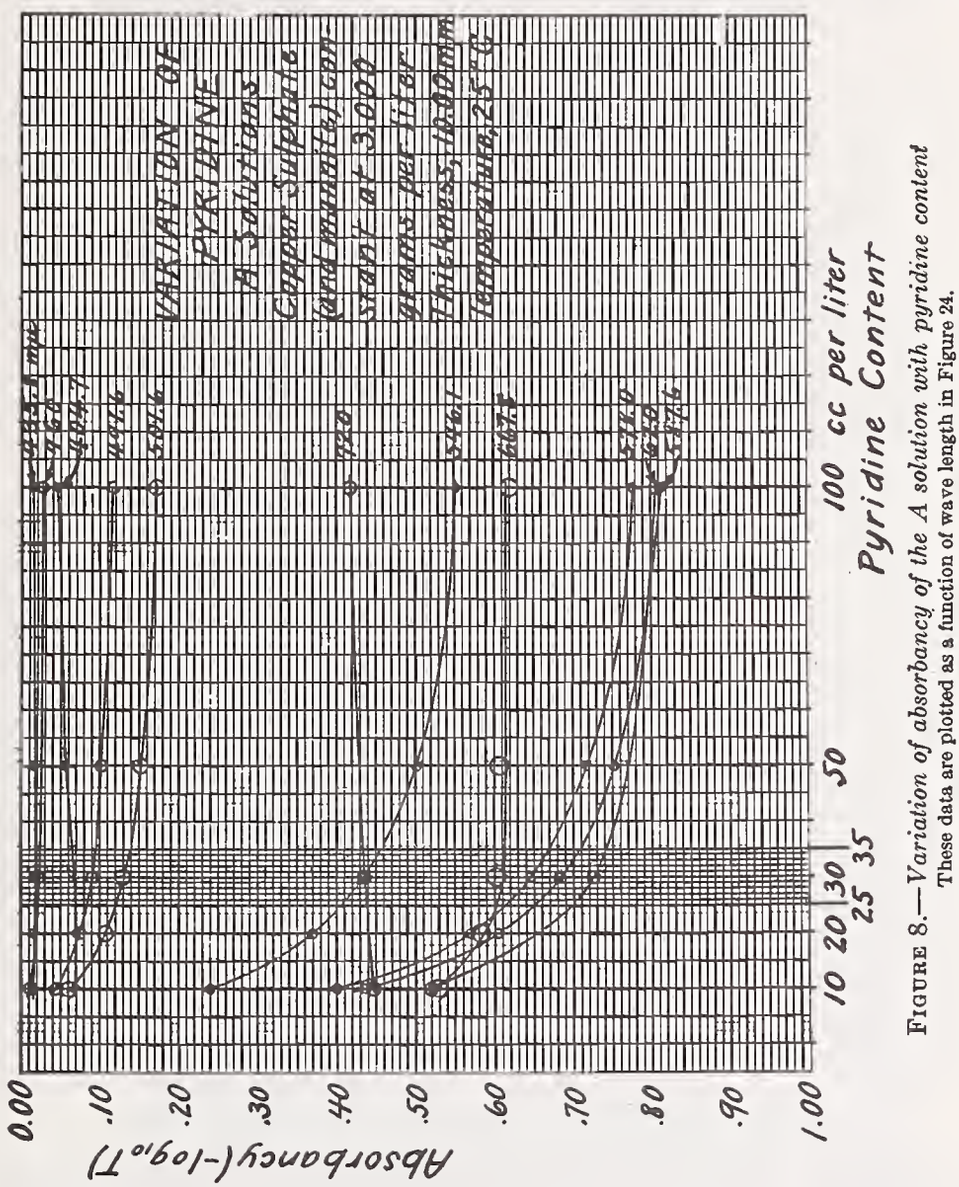


FIGURE 8.—*Variation of absorbancy of the A solution with pyridine content*
These data are plotted as a function of wave length in Figure 24.

is 115.5° C. In view of the change in absorption that takes place with variation in either the ammonia or pyridine content this advantage of the pyridine over the ammonia is important.

3. At the most favorable content of either ammonia or pyridine there is no particular advantage of one over the other so far as spectral transmission is concerned, except that in the violet and ultra-violet the ammonia solutions have the higher transmittancy. If filters were prepared using ammonia instead of pyridine, the agreement between

the computed and ideal energy distributions would be slightly better at these short wave lengths in the majority of cases than was obtained with the pyridine filters as herein described.

4. So far as the evidence goes, the obedience to Beer's law is closer for the ammonia solutions than for the pyridine solutions. This might have saved time and labor in the measurements and computations, but would not result in any greater accuracy in the final filters.

5. By spectrophotometric test, the ammonia solutions, if evaporation is prevented, appear to be somewhat more permanent than the pyridine solutions. Here, as elsewhere, the data for the ammonia solutions is less extensive than for the pyridine solutions, because of the decision to use pyridine in preference to ammonia. Whether the ammonia solutions would remain as permanent under actual working conditions is questionable.

It is believed the importance of the first two points noted above, wherein the pyridine solutions are superior to ammonia solutions, is sufficiently great to outweigh the other points wherein the ammonia solutions seem superior. Although the relative permanence could not be known at the time a decision had to be made, the authors feel justified in their search for and use of pyridine, in preference to ammonia with copper sulphate for securing the principal (blue) component of the filters.

(c) AQUEOUS ACIDIFIED SOLUTION OF COPPER SULPHATE—SOLUTION B_1

With either solution A or the copper-sulphate-ammonia solution the resulting energy distributions obtained with the most favorable concentrations will show a considerable excess in the blue-green and far red, and it has already been noted that Pfund found that a satisfactory correcting solution could be obtained from a mixture of copper and cobalt sulphates in aqueous solution.

Copper sulphate ($CuSO_4.5H_2O$) is a stable compound, readily obtained in relatively pure form. Aqueous solutions containing 10 cc of sulphuric acid showed no failure of Beer's law over a range of concentrations from 5 to 40 g per liter. The standard medium concentration used for the principal spectrophotometric measurements was 20 g per liter. While the sulphuric acid, which is a part of solutions B_1, B_2, and B, does not cause any change in absorption it serves the useful purposes of clarifying the solutions and preventing the formation of basic copper compounds. An aqueous solution containing 10 cc of sulphuric acid (specific gravity 1.84) per liter of solution and having varying amounts of copper sulphate will be designated as "solution B_1."

(d) AQUEOUS ACIDIFIED SOLUTION OF COBALT SULPHATE

Extensive spectrophotometric data were taken with cobalt sulphate ($CoSO_4.7H_2O$). The standard concentration, chosen to be in approximately the middle of the desired range, was 10 g of cobalt sulphate per liter of solution, containing also 10 cc of concentrated sulphuric acid. It was found, however, that the cobalt sulphate had lost some of its water of crystallization and that the standard solution did not, therefore, contain exactly 10 g of $CoSO_4.7H_2O$. Chemical analysis of this solution showed that it had a concentration equivalent to 10.3 g of $CoSO_4.7H_2O$.

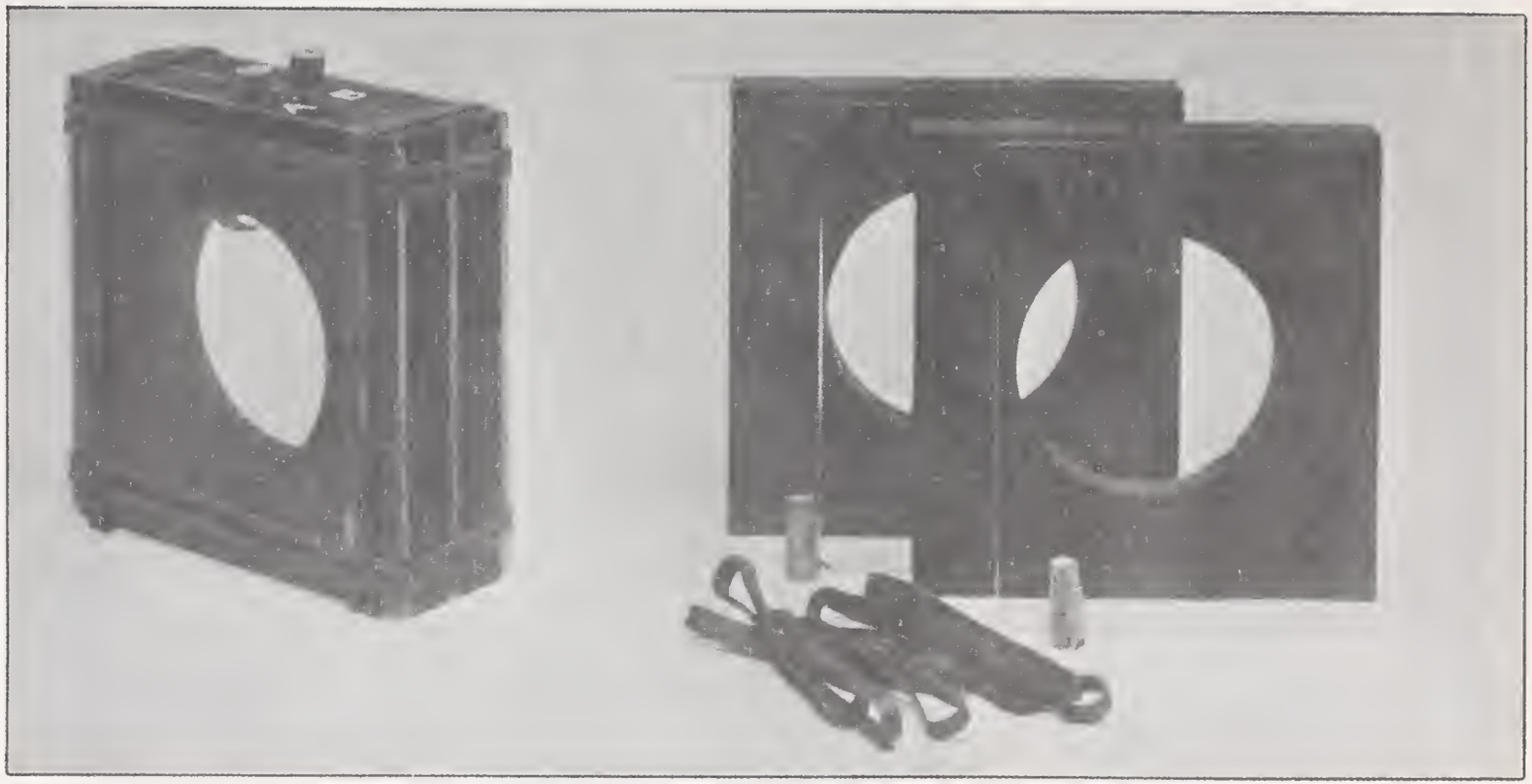


FIGURE 9.—*Type of cell found satisfactory for holding filter solutions*

(e) AQUEOUS ACIDIFIED SOLUTION OF COBALT AMMONIUM SULPHATE — SOLUTION B_2

To avoid the uncertainty which would thus be present in the use of cobalt sulphate due to the uncertainty of the water content, the double salt, cobalt ammonium sulphate ($CoSO_4(NH_4)_2SO_4.6H_2O$), was suggested. This material was also used in the preparation of photometric filters by Ives and Kingsbury (33) for the same reason.

Spectrophotometric measurements proved the practical equivalence of solutions containing 14.481 g of cobalt ammonium sulphate and the standardized solution containing 10.3 g of cobalt sulphate, both alone and in mixture with copper sulphate (See Secs. VII, 2 and VII, 3). Beer's law is obeyed in both cases over the range of concentrations used. An aqueous solution containing 10 cc of sulphuric acid per liter and having varying amounts of cobalt ammonium sulphate will be referred to as "solution B_2."

(f) MIXTURE OF SOLUTIONS B_1 AND B_2—SOLUTION B

It is, of course, desirable to mix the acidified cobalt ammonium sulphate and copper sulphate solutions, as this reduces the number of filter compartments from three to two. Tests on the effect of mixing these two ingredients to form a single solution showed, as expected, that no reaction took place, and that the observed spectral transmittancies agreed perfectly with computed values based on the two standard concentrations. An aqueous solution containing copper sulphate and cobalt ammonium sulphate in varying amounts, together with 10 cc of sulphuric acid per liter is referred to throughout the rest of the paper as "solution B."

(g) CELL FOR HOLDING FILTER SOLUTIONS

The A and B solutions are contained in a 2-compartment cell, such, for example, as is illustrated in Figure 9.[11] The essential specification of this cell is that it consist of three borosilicate crown glass plates (each 2.5 mm thick, refractive index, D line, = 1.51), separated by two glass frames, each 10.00 mm thick, for containing the respective A and B solutions. Borosilicate crown glass was selected for the windows because it is the hardest and most transparent of the optical glasses and probably the most resistant to chemical action.

(h) DEFINITION OF DAVIS-GIBSON FILTERS

Summarizing the selection of materials, a Davis-Gibson filter may be defined as a 2-compartment cell with three borosilicate crown glass windows, the two compartments being filled, respectively, with solutions A and B, having compositions as specified below.

Solution A

Copper sulphate ($CuSO_4.5H_2O$)	g	c
Mannite ($C_6H_8(OH)_6$)	g	c
Pyridine (C_5H_5N)	cc	30.0
Water (distilled) to make	cc	1,000

The value of c, the weight in grams of copper sulphate (and mannite) per liter of solution, varies with the purpose of the filter, as illustrated in the various charts and tables. When c has the value of 3.000 g

[11] This design is similar to that used by Ives and Kingsbury (33).

(the concentration at which the principal spectrophotometric determinations were made) the solution is referred to as solution A′.

Solution B

Copper sulphate ($CuSO_4.5H_2O$)	g	c_1
Cobalt ammonium sulphate ($CoSO_4.(NH_4)_2SO_4.6H_2O$)	g	c_2
Sulphuric acid (specific gravity, 1.84)	cc	10.0
Water (distilled) to make	cc	1,000

The values of c_1 and c_2 depend, again, on the purpose of the filter. When $c_1 = 20.000$ g and $c_2 = 14.481$ g, the mixture is referred to as solution B′. Separate solutions of copper sulphate and cobalt ammonium sulphate having the weights just given, and containing 10 cc each of sulphuric acid per liter, are referred to, respectively, as solution B_1' and solution B_2'. Extensive spectrophotometric measurements were made on solutions B′, B_1', and B_2'.

2. PURITY OF CHEMICALS USED

It was decided early in the investigation to use a grade of chemicals known as "C. P.," because such chemicals are readily available. The term "C. P.," although originally intended to mean "chemically pure," is not at present generally used to designate the purest material available on the market. The terms "analytical reagent quality" or "reagent quality" are often used to designate a higher grade. "C. P." chemicals are, however, superior to "technical" or "commercial" grades. The impurities present in different lots of a given chemical of the "C. P." grade are generally of the same nature, but may differ in amount. It was believed to be the better plan to include the effects of these impurities in the spectrophotometric data, provided such effects were small or negligible, rather than to use specially purified chemicals which would not be generally obtainable. Our experience throughout the investigation has confirmed this opinion. The nature and extent of the most important impurities are indicated below, and their spectrophotometric effects are discussed and illustrated in Section VII.

The analyses given in this section serve to define the materials used in deriving the standard data given in the paper. Questions of reproducibility are discussed further in Section VII, 8. Certain precautions to use in the selection of the materials are considered in Section IX on the preparation of the filters.

To prevent possible contamination from cork stoppers, glass-stoppered bottles were used to hold all solutions.

(a) COPPER SULPHATE

This well-known chemical ($CuSO_4.5H_2O$), is readily obtainable in high purity. The copper sulphate used in deriving the standard data for the filters was of C. P. quality obtained from Mallinckrodt Chemical Works, St. Louis, Mo. Chemical analysis gave the following impurities (memorandum, March 25, 1926, from chemistry division):

518-743 O - 73 - 44

	Per cent
Water insoluble material	0. 000
Iron and aluminum (as R_2O_3)	. 007
Zinc (as Zn)	. 021
Calcium	. 000
Nickel (as Ni)	. 029
Magnesium and alkali salts	. 04

These small quantities of impurities have no practical effect on the spectral transmission of the filters.

The pentahydrate of copper sulphate is ordinarily considered as a stable salt, and for many purposes this is true. It should be noted, however, that the water of crystallization is subject to minor changes under ordinary conditions of keeping. The crystals tend to follow atmospheric changes, gaining or losing water depending on the variation of the aqueous vapor pressure of the air. It may also happen that the salt is not fully air dried, following crystallization before bottling, in which case a given weight of the copper sulphate will contain less than the theoretical copper content. These variations are usually small, seldom amounting to more than ±1 per cent, because of the tenacity with which copper sulphate clings to its water of crystallization. The following chemical analysis (Chem. Lab. No. 551/31079–84, January 3, 1929) made later on the same stock as analyzed above for impurities, illustrates this point:

	Copper content	Percentage of theoretical
	Per cent	
Opened bottle from which sample had been taken over a period of more than 6 months	25. 70	100. 9
Unopened bottle, same stock, on hand about 2 years	25. 24	99. 1

However, copper sulphate which has been kept for long periods of time, particularly in very dry atmosphere, may lose a larger percentage of its water content. When such a change has taken place to a considerable extent, a whitish powder ($CuSO_4 . 3H_2O$) will be localized on or may cover the surface of the crystals. Such material should not be used in preparing light-filter solutions. It may, of course, be made satisfactory for this purpose by recrystallizing or by rinsing with distilled water and drying to equilibrium.

This uncertainty in the copper content of copper sulphate, amounting, in general, to perhaps ±1 per cent of the theoretical content, is apparently of much more importance in the use of the filters than is that resulting from any impurities which are likely to be present. It is, however, of practical importance only in the case of the A solution. (The magnitude of a 1 per cent variation in the copper content of the A and B solutions, both as regards the spectral absorbancy of the solutions and the color of the filters, is apparent from Table 4 and fig. 30.)

The results of the determination of the copper content (Chem. Lab. No. 551/31079–84, above), made on the same lot of copper sulphate as had been used in preparing the standard solutions, rendered somewhat uncertain the exact copper content of solutions for which the standard spectrophotometric data given in Tables 4 and 5 should apply. Accordingly, analyses were made of the A′,

B_1', and B′ solutions which had been given the most weight in adopting the standard data. The following results were obtained (Chem. Lab. Nos. 551/31079–84, January 3, 1929, and 551/31429–32, January 2, 1929):

	Copper content	Percentage of theoretical
	Grams per liter	
A′ (No. 62-15)	2.99	99.7
B_1' (No. 62-18)	19.78	98.9
B′ (No. 65-2)	20.10	100.5

The value 99.7 per cent, falling between the limits obtained on the salt, may be taken as the copper content for which the data of Table 4, solution A′, apply, as stated in the footnote to that table.

Solutions B_1' (No. 62–18) and B′ (No. 65–2) were both supposed to have 20.000 g of copper sulphate per liter. It is shown later on (Sec. VII, 2, (*e*) and Table 21) that, at 667.8mμ in the red where the copper sulphate absorbs strongly, there was a good check in the absorbancy of the two solutions, allowing for the small absorbancy of the cobalt ammonium sulphate in solution B′. This check makes it impossible to choose either of the above analyses in preference to the other; and their mean value of 99.7 per cent (the mean of 98.9 and 100.5, this being the same value as obtained with the A′ solution, No. 62–15) may be taken as the copper content for which the data of Table 5, Solution B_1', apply.

(b) COBALT AMMONIUM SULPHATE

This is a double salt ($CoSO_4(NH_4)_2SO_4.6H_2O$) of cobalt sulphate and ammonium sulphate. Nickel is practically always present in the salts of cobalt and it is extremely difficult to remove completely. Nickel sulphate has absorption bands in the violet and far red and is, therefore, very undesirable to have present in any considerable quantity. Because of the cost and difficulty of obtaining cobalt ammonium sulphate free from nickel, which would make it prohibitive for general use in the filters, we chose as a standard a nickel content of 0.075 per cent (ratio of nickel to cobalt = 1:200). At this proportion the absorption of light due to nickel is practically negligible, but, such as it is, is included in the spectrophotometric data. A ratio of nickel to cobalt not exceeding 1:100 is satisfactory for use in the filters. The cobalt ammonium sulphate used in this work was prepared by crystallization from cobalt sulphate and ammonium sulphate. The following extracts from Doctor Wichers's memorandum (chemistry division, March 17, 1927) on tests of these salts will be of interest:

> All samples of cobalt sulphate and cobalt ammonium sulphate tested were found to contain nickel. Two samples of cobalt sulphate made by American manufacturers were found to have 0.10 per cent and 0.09 per cent Ni, respectively. This is a ratio of Ni to Co of about 1:200. A sample of cobalt ammonium sulphate of foreign make held 1.37 per cent Ni (ratio of Ni to Co of 1:11). A sample of American-made cobalt ammonium sulphate held 0.20 per cent nickel. These results indicate that nickel may be present in excessive amounts in these salts. Experience shows also that the amount of nickel indicated in the analysis on the label is often in error.

There seems to be no reason why a sample of cobalt ammonium sulphate can not be made as free of nickel (ratio of Ni to Co) as the corresponding cobalt sulphate. The sample of cobalt ammonium sulphate made from the cobalt sulphate containing 0.10 per cent Ni had a nickel content of about 0.05 per cent Ni. The ratio of nickel to cobalt here is 1:300. The nickel was determined in a sample that had been twice crystallized by you and twice recrystallized in this laboratory. This indicates that a salt contaminated with nickel can not be easily improved by successive recrystallizations as the double sulphate.

Cobalt ammonium sulphate crystallizes as $CoSO_4.(NH_4)_2SO_4.6H_2O$ and is stable under ordinary atmospheric conditions. Exposed in a balance case for a period of two months it showed practically no change in weight. As was to be expected, the weight changed slightly with changes in humidity. Weights taken at different times and under different conditions covered a maximum range of 0.0016 g on a 7.5 g sample. The final weight was 0.0004 g less than the initial.

The theoretical cobalt content in cobalt ammonium sulphate is 14.91 per cent. The cobalt content of the foreign sample was 14.58 per cent; that supplied by an American maker was 14.86 per cent, while the material made by you held 14.60 per cent cobalt. In every case the cobalt is determined by electrolysis so that the figure reported is really the combined cobalt and nickel content. Your sample was recrystallized twice in this laboratory to see if a salt with the theoretical cobalt content could be easily produced. The crystals after being separated from the mother liquor were air dried. Upon analysis they were found to contain the theoretical cobalt content. The values found in two determinations were 14.92 and 14.88 per cent Co.

The following is the method regularly used in the chemical laboratories of the National Bureau of Standards for determinations of the nickel content of cobalt ammonium sulphate (memorandum of January 6, 1928):

Dissolve 1.00 g of cobalt ammonium sulphate in 300 cc of water. Add 1 g of tartaric acid and heat the solution to about 80° C. Then add 100 cc of a 1 per cent alcoholic solution of dimethylglyoxime, and ammonium hydroxide solution until the solution is distinctly alkaline to litmus paper (an excess is to be avoided). Cover the beaker and place on the steam bath for 3 hours and then keep in a warm place (about 40° C.) for 36 to 48 hours longer. Filter off the precipitate. If little or no precipitate formed allow the filtrate to stand an additional 24 hours to be sure that all the nickel has been precipitated. Add any precipitate which forms to the first precipitate. Dissolve the combined precipitates in 10 cc of hot 1:1 nitric and dilute to 100 cc. Heat to about 80° C. Then add 10 cc of dimethylglyoxime solution and ammonium hydroxide solution until the solution is alkaline to litmus paper, but avoiding an excess. Digest on the steam bath to coagulate the precipitate and filter through a tared, asbestos-covered Gooch crucible. Wash the precipitate thoroughly with hot water and dry at 110° to 120° C. for 45 minutes. Multiply the weight of the precipitate by 0.2031 to get the weight of nickel (Ni).

It was stated in Section IV, 1, (*d*) that the cobalt-sulphate solution (No. 62–17) was found by chemical analysis to have 10.3 g of $CoSO_4.7H_2O$ instead of the 10.000 g which it was supposed to have. This value was the mean of three analyses made by the chemistry division of the bureau, the individual values being as follows: (1) 10.30 g per liter, November 19, 1926; (2) 10.39 g per liter, December 7, 1926; (3) 10.25 g per liter, December 9, 1926. The second determination was possibly somewhat too high and less accurate than the other two. At any rate the value of 10.3 g of cobalt sulphate per liter (equivalent to 14.481 g of cobalt ammonium sulphate per liter) may be accepted as accurate to a fraction of 1 per cent, and it is this value for which the standard spectrophotometric data of Table 6 apply.

(c) PYRIDINE

This is a stable colorless liquid organic compound (C_5H_5N). It has a strong unpleasant odor which gives one the false impression of high volatility. Its boiling point is 115.5° C. at 760 mm; its specific gravity is 0.979 ± 0.001 at 25° C. The change in specific gravity is 0.001 per °C. The pyridine used throughout the present investigation was J. T. Baker Chemical Co.'s C. P. grade.

Pyridine, like most organic compounds, is very difficult to obtain in a pure state. The material which is available commercially in the C. P. grade always contains a small quantity of the higher homologues, the principal one being α-picoline. A small amount of water is also usually present. Complete chemical analysis was not made, because a satisfactory procedure to accomplish this result is not available. Since α-picoline is considered to be the major impurity likely to be present in pyridine, and because of the difficulty of separating the higher homologues, we have expressed the total effect of these impurities in terms of the equivalent α-picoline. Samples of commercial pyridine and α-picoline were purified both by fractional distillation and fractional crystallization[12] of the mercuric chloride compounds of the hydrochlorides, and A solutions prepared containing various amounts of the picoline. The spectrophotometric analysis of these solutions is described in Section VII, 9 (*b*). The amount of picoline present in our standard pyridine was shown by the spectrophotometric analysis to be 2.4 per cent by weight. This analysis, combined with titration tests and specific gravity determinations gave 1.2 per cent water and 96.4 per cent pyridine, by weight.

Pyridine is ordinarily tested for strength by titration methods (as described below), and values thus obtained show the true percentage of pyridine only in the absence of the higher homologues or other impurities which also react with the sulphuric acid. With the presence of the higher homologues the combined strength, in terms of pyridine, is obtained.

Our standard pyridine, as analyzed above and as used in deriving the standard spectrophotometric data of Table 4, was thus titrated for strength with half normal sulphuric acid, using methyl orange for an indicator. This work, based on a series of runs by two observers gave a value of 98.4 per cent, which is equivalent to the analysis given above in terms of pyridine, picoline, and water. Since the end point is not particularly sharp, the following procedure was followed:

The pH value of the end point (3.2 ± 0.1) was determined by electrometric titration with the hydrogen electrode. In the colorimetric titrations a buffered solution, pH 3.2, was used as a comparison standard for the end-point color. Making use of this end point a sample of pyridine which had been fractionated several times and was without doubt practically pure, gave a value of 100 per cent strength.

We have examined, altogether, 17 samples of pyridine obtained from various sources, and, with the exception of 2 samples which were colored (yellow), all were suitable for use in the filter solutions. The titration tests gave values for these 17 samples ranging from 97 to 100 per cent strength. As shown in Section VIII, 2 (*c*) a departure

[12] We are indebted to I. M. Jacobsohn, of the chemistry division, for the fractional crystallization of pyridine and picoline.

of a least 6 per cent from our standard of 98.4 per cent strength is necessary to affect the color of the filters as much as is caused by a 1 per cent error in the copper sulphate. It thus appears that for general use it is unnecessary to test pyridine by titrating for strength. Independent tests on Davis-Gibson filters made at the National Physical Laboratory of Great Britain have led to the same conclusion. (See quotation at end of Sec. VII, 8.)

The presence of the homologues might be indicated by treatment with potassium permanganate. However, the reaction is slow at room temperatures, and other reducing agents which may be present in pyridine react with the permanganate more rapidly than do the homologues. Although this makes the test of little value as a means of detecting the higher homologues, it is nevertheless useful in detecting the presence of other undesirable impurities. We recommend that the pyridine used in the filter solutions pass the following simple test: To 5 cc of pyridine add 2 drops of $N/10$ $KMnO_4$. This should remain red for at least an hour.

Pyridine should be kept in glass-stoppered bottles because it reacts with cork, becoming yellow. Pyridine so contaminated will not pass the permanganate test. We would strongly recommend that in purchasing pyridine, glass-stoppered bottles be specified.

(d) MANNITE

This material ($C_6H_8(OH)_6$), also referred to as *d*-mannitol is one of the sugar alcohols. It is commercially obtainable as fine white needles which melt at 166° C. The mannite used in this work was prepared from crude manna and purified by recrystallization from water solution by the polarimetry section of the bureau. This mannite was compared by C. F. Snyder, polarimetry section, with a sample from the Eastman Kodak Co. in the following test:

Optical rotation.—Ten grams of *d*-mannitol and 10 g of borax were dissolved in water and the volume made up to 100 ml at 20° C. The solutions were filtered and observed on a Bates type saccharimeter using white-light illumination. From the values so obtained the specific rotations for sodium light were calculated. The same value was obtained for both samples, viz, +20.9° (with borax) for the D line at 20° C. (The specific rotation of mannite is greatly increased by borax. In simple aqueous solution the specific rotation is very small.)

Melting point.—Determinations on the two samples showed for the Eastman sample, M. P. = +166° C., and for the Bureau of Standards sample, M. P. = +165° C.

The use of mannite in the filter is that of a stabilizer. Any rotatory power possessed by the solution containing mannite need have no influence on the color and the spectral energy distribution of the lamp and filter combination. Rotatory effects could be important only if the filter were placed between a polarizer and analyzer, and such an arrangement need never be used. In cases where the intensity of light is modified by polarization methods, any rotatory powers that the filter might have could be nullified by placing the filter elsewhere than between the polarizing elements.

Furthermore, even these precautions are entirely unnecessary as is shown by the following test, made by F. P. Phelps of the polarimetry

section of the bureau, on the A solution (containing mannite) of the 2,360° K-to-mean-sun filter:

The rotation was measured in a 20-mm tube.
The estimated experimental error is about ±0.005°.

1. The observed rotation for the band (blue) transmitted by the tube of solution when illuminated by white light from a 100-watt tungsten stereoptican lamp was	+0.001°
2. The observed rotation for λ=546 was	+.005°
3. The observed rotation for λ=578 was	+.002°
4. When a red screen was interposed under the conditions of (1), so much light was absorbed that no reading could be made.	

It will be seen that while all the readings obtained are positive, their magnitude is well within the experimental error. The rotation, therefore, if any, is so small as to be entirely negligible.

(e) AMMONIA

Ammonium hydroxide (ammonia water) is easily obtained in high purity. The strength of the usual C. P. grade is about 28 per cent NH_3. It is well known that ammonia water is unstable at ordinary temperatures, decomposing into water and ammonia gas, and it is thus necessary to determine the NH_3 content before using. In our work Baker and Adamson ammonia water of C. P. quality was used. The NH_3 content was tested by standard titration methods immediately before using, and the quantity used in preparing the solutions was adjusted in all cases so as to obtain the equivalent of 28 per cent ammonia water.

V. METHODS OF COMPUTATION

1. SPECTROPHOTOMETRIC TERMINOLOGY

Before describing the various computations required in the design of the filters, it is necessary to define precisely the various spectrophotometric terms which are used. Throughout the paper, both before and following the present section, these terms are used in strict accord with these definitions.

Let radiant energy, homogeneous with respect to wave length, be incident perpendicularly upon the first surface of a homogeneous, isotropic, nondiffusing medium or series of such media in contact with one another, having plane, nondiffusing, parallel surfaces, and after passage through the medium or media emerge perpendicularly from the last surface. If E_1 be the energy thus incident upon the first surface and E_2 the energy thus leaving the last surface in the same direction as E_1, then transmission, T, is defined as:

$$T=\frac{E_2}{E_1}$$

Thus, we may refer to the transmission of a plate of glass, or a cell with glass end plates containing a solution, or of the two-compartment filters herein described, in all cases referring to the ratio of the energies transmitted by and incident (normally) upon the material or group of materials under consideration.

Density, D, is defined as the negative logarithm of the transmission, thus:

$$D=-\log_{10}T$$

Consider further a substance in homogeneous solution in a solvent contained in a cell with plane, parallel sides, perpendicular to the direction of propagation of energy. Let T_{sol} be the transmission of a cell containing the solution, and let T_{sov} be the transmission of the same or a similar cell filled with the solvent. Then transmittancy, T, is defined as:

$$\mathsf{T} = T_{sol}/T_{sov}$$

Throughout the present paper T_{sov} refers to a cell containing distilled water only.

Absorbancy, A, is defined as the negative logarithm of the transmittancy, thus:

$$\mathsf{A} = -\log_{10}\mathsf{T}$$

Let b and c refer to the thickness and concentration, respectively, of a solution. The quantity $\mathbf{k}$, equal to A/bc, is known as the specific absorptive index. Its magnitude depends, of course, upon the particular units chosen for thickness and concentration. The relation, $\mathsf{A} = \mathbf{k}bc$, is known as Beer's law. For a given concentration, $\mathsf{A} = \mathbf{k}'b$. This relation is known as Lambert's law and may be derived from theoretical considerations. The analogous relation, with thickness constant, viz, $\mathsf{A} = \mathbf{k}''c$, may or may not be valid, depending upon the particular solution. Failure of Beer's law always refers to a failure of this last relation.

The terms transmission, transmittancy, and density, as defined above, are in accordance with the terminology of the O. S. A. committee report on spectrophotometry (35, p. 177). The primary meaning of density as used in photographic sensitometry has more nearly the meaning of absorbancy than of density as defined in this paper. In the present case, it is thought best to use the terminology developed for spectrophotometric work. The quantity $\mathbf{k}$ $(=\mathsf{A}/bc)$ was defined in the O. S. A. committee report as the specific transmissive index, but it seems more appropriate to name it the specific absorptive index, as has been done by Peters and Phelps (36, p. 265) and Brode (37, p. 503).

The term absorbancy was used by Peters and Phelps to represent the quantity $1-\mathsf{T}$ instead of $-\log_{10}\mathsf{T}$ as defined above; however, the need for a word or symbol to designate $1-\mathsf{T}$ is not so pronounced. In discussions of the absorptive properties of solutions, on the other hand, the negative logarithm of the transmittancy is one of the most commonly used quantities. Hitherto, the terms density and extinction coefficient have often been used in this connection, but there are objections to both expressions. Density is widely used in photographic work where it usually has reference to transmittancy that is more or less diffused, in addition to its use in the consideration of rectilinear transmission as defined above. Extinction coefficient is an unsatisfactory expression, for, when used in this sense, it is not a coefficient and in its derivation does not accord with the usual meaning of extinction. The use of the word absorbancy eliminates the need for these expressions and with its symbol A makes unnecessary the continual repetition of the cumbersome expressions $-\log_{10}$ transmittancy and $-\log_{10}\mathsf{T}$.

2. STANDARD SPECTROPHOTOMETRIC DATA

With solutions of medium concentration selected as noted in Section IV, 1, and by spectrophotometric methods described in detail in Section VII, standard absorbancy values were adopted for the component solutions and standard density values for the double cell filled with water. These fundamental values are given in Tables 4,

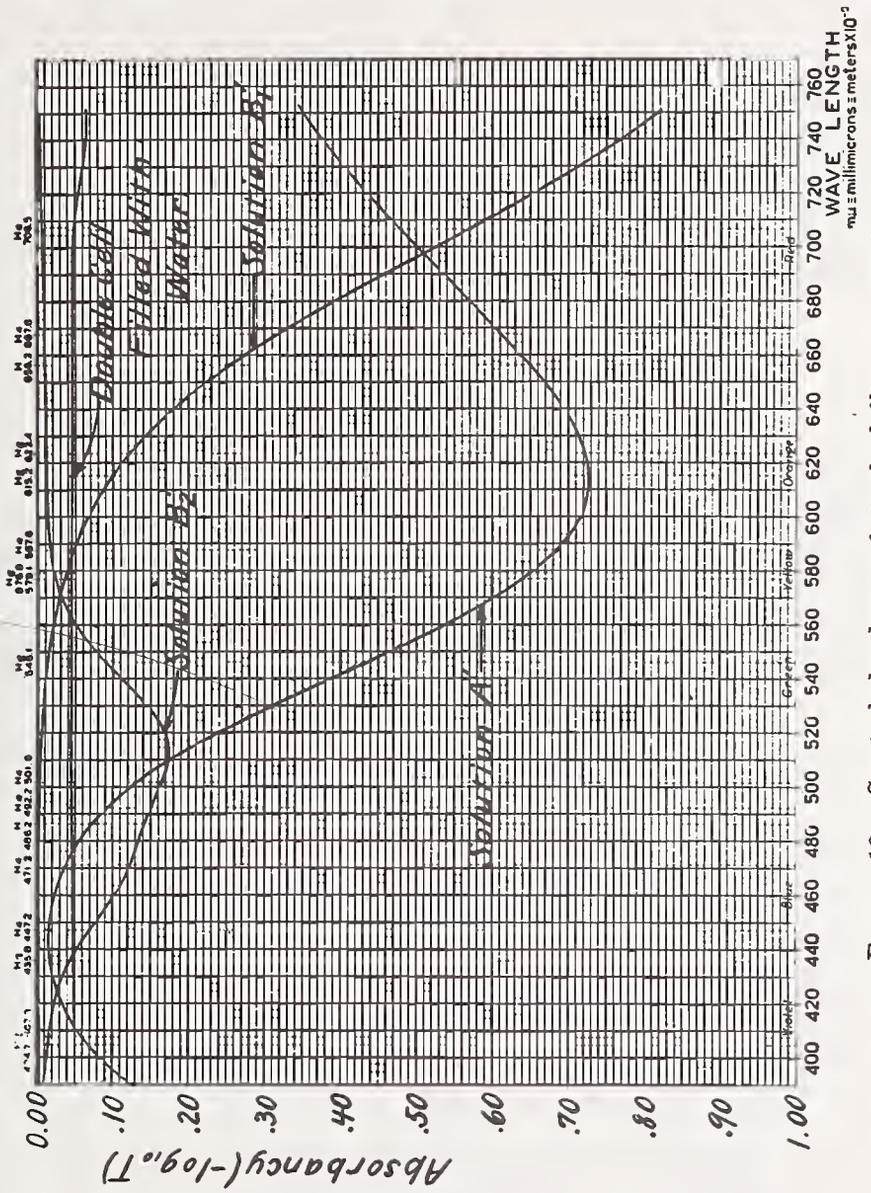


FIGURE 10.—*Spectral absorbancy of standard filter components*
These curves are plotted from the data of Tables 4, 5, 6, and 7, which see for composition of solutions and cell.

5, 6, and 7, and illustrated in Figure 10; in the tables are also given values of absorbancy on a 1 g basis as used in the computations, values of the Beer's law correction factor for the A solutions, and values of transmittancy and transmission derived, respectively, from the absorbancy and density values. Values of absorbancy for any concentration may be computed by multiplying the 1 g absorbancy values by the concentration in grams, provided that any Beer's law failure is allowed for in such computation. As noted elsewhere, a

Beer's law correction was necessary only with the A solutions and the method of applying it is described in the next section.

The advantage of using the negative logarithms of spectral transmission and transmittancy in computations involving the combination of two or more materials may be noted here. Let D_W represent the density of a double cell, such as used in the present filters, when filled with distilled water, and D_F the density of the same cell when filled with the A and B solutions. Furthermore, let $\mathbf{A}_A$, $\mathbf{A}_{B1}$, $\mathbf{A}_{B2}$, and $\mathbf{A}_B$ refer, respectively, to the absorbancies of the A, B_1, B_2, and B solutions. Then, at any wave length,

$$D_F = D_W + \mathbf{A}_A + \mathbf{A}_B$$

or

$$D_F = D_W + \mathbf{A}_A + \mathbf{A}_{B_1} + \mathbf{A}_{B_2}$$

For the computations in which we are interested these are simpler formulas than the corresponding ones,

$$T_F = T_W \times \mathbf{T}_A \times \mathbf{T}_B$$

and

$$T_F = T_W \times \mathbf{T}_A \times \mathbf{T}_{B_1} \times \mathbf{T}_{B_2}$$

TABLE 4.—*Standard spectrophotometric data for solution A'*

Unfiltered, thickness 10.00 mm, temperature 25° C., having the following composition:

Copper sulphate ($CuSO_4 \cdot 5H_2O$)[1]	g	3.000
Mannite ($C_6H_8(OH)_6$)	g	3.000
Pyridine (C_5H_5N)[2]	cc	30.0
Water (distilled) to make	cc	1,000

Wave length in millimicrons	Absorbancy A_c ($c_c=3$)	$\frac{A_c}{c_c}$	Beer's law correction factor[3] B	Transmittancy[4] T_c	Wave length in millimicrons	Absorbancy A_c $c_c=3$	$\frac{A_c}{c_c}$	Beer's law correction factor[3] B	Transmittancy[4] T_c
350	0.470	0.156667	−0.120	0.339	600	0.717	0.239000	−0.010	0.192
360	.360	.120000	.120	.437	610	.726	.242000	.010	.188
370	.267	.089000	.118	.541	620	.725	.241667	.010	.188
380	.190	.063333	.114	.646	630	.714	.238000	.010	.193
390	.127	.042333	.105	.746	640	.696	.232000	.010	.201
400	.0835	.027833	.080	.825	650	.668	.222667	.010	.215
410	.053	.017667	.050	.885	660	.637	.212333	.010	.231
420	.033	.011000	.032	.927	670	.603	.201000	.010	.249
430	.0215	.007167	.020	.952	680	.570	.190000	.010	.269
440	.0165	.005500	.013	.963	690	.536	.178667	.010	.291
450	.017	.005667	.010	.962	700	.501	.167000	.010	.316
460	.023	.007667	.010	.948	710	.468	.156000	.010	.340
470	.0355	.011833	.010	.922	720	.435	.145000	.010	.367
480	.0555	.018500	.010	.880	730	.405			.394
490	.085	.028333	.010	.822	740	.377			.420
500	.123	.041000	.010	.753	750	.351			.446
510	.175	.058333	.010	.668					
520	.240	.080000	.010	.575	Hg 404.7	.068		.065	.855
530	.311	.103667	.010	.489	Hg 435.8	.018		.016	.959
540	.387	.129000	.010	.410	Hg 491.6	.091		.010	.811
					He 501.6	.131		.010	.740
550	.464	.154667	.010	.344					
560	.538	.179333	.010	.290	Hg 546.1	.434		.010	.368
570	.603	.201000	.010	.249	Hg 578.0	.646		.010	.226
580	.655	.218333	.010	.221	He 587.6	.685		.010	.207
590	.693	.231000	.010	.203	He 667.8	.610		.010	.245

[1] Chemical analysis showed the copper sulphate to have 99.7 per cent of the theoretical copper content.

[2] Titration showed the pyridine to have an effective strength of 98.4 per cent, this including 2½ per cent picoline impurity.

[3] The Beer's law correction factor is applied as follows (see Sec. V, 3):

$$A = \frac{A_c}{c_c} c[1 + B(c - c_c)]$$

Note that B is a negative quantity.

[4] These values of transmittancy are derived from the adopted values of absorbancy. $A_c = -\log_{10} T_c$.

TABLE 5.—*Standard spectrophotometric data for solution B_1'*

Unfiltered, thickness 10.00 mm, temperature 25° C., having the following composition:

Copper sulphate ($CuSO_4.5H_2O$)[1]	g	20.000
Sulphuric acid (specific gravity 1.835)	cc	10.0
Water (distilled) to make	cc	1,000

Wave length in millimicrons	Absorbancy A_ϵ	$\frac{A_\epsilon}{20}$	Transmittancy[2] T_ϵ	Wave length in millimicrons	Absorbancy A_ϵ	$\frac{A_\epsilon}{20}$	Transmittancy[2] T_ϵ
350	0.0090	0.000450	0.979	600	0.0680	0.003400	0.855
360	.0063	.000315	.986	610	.0885	.004425	.816
370	.0046	.000230	.989	620	.1125	.005625	.772
380	.0035	.000175	.992	630	.143	.007150	.719
390	.0028	.000140	.994	640	.180	.009000	.661
400	.0023	.000115	.995	650	.224	.011200	.597
410	.0019	.000095	.996	660	.274	.013700	.532
420	.0016	.000080	.996	670	.332	.016600	.466
430	.0014	.000070	.997	680	.392	.019600	.406
440	.0012	.000060	.997	690	.459	.022950	.348
450	.0011	.000055	.997	700	.527	.026350	.297
460	.0011	.000055	.997	710	.592	.029600	.256
470	.0012	.000060	.997	720	.656	.032800	.221
480	.0014	.000070	.997	730	.715		.193
490	.0018	.000090	.996	740	.768		.171
500	.0026	.000130	.994	750	.817		.152
510	.0038	.000190	.991				
520	.0055	.000275	.987	Hg 404.7	.0021		.995
530	.0079	.000395	.982	Hg 435.8	.0013		.997
540	.0111	.000555	.975	Hg 491.6	.0019		.996
550	.0155	.000775	.965	He 501.6	.0028		.994
560	.0216	.001080	.951	Hg 546.1	.0135		.969
570	.0292	.001460	.935	Hg 578.0	.0368		.919
580	.0390	.001950	.914	He 587.6	.0487		.894
590	.0518	.002590	.888	He 667.8	.319		.480

[1] Analysis showed the copper sulphate to have 99.7 per cent of the theoretical copper content.
[2] These values of transmittancy are derived from the adopted values of absorbancy. $A_\epsilon = -\log_{10} T_\epsilon$.

TABLE 6.—*Standard spectrophotometric data for solution B'_2*

Unfiltered, thickness 10.00 mm, temperature 25° C., having the following composition:[1]

Cobalt ammonium sulphate ($CoSO_4.(NH_4)_4SO_4.6H_2O$)[2]	g	14.481
Sulphuric acid (specific gravity 1.835)	cc	10.0
Water (distilled) to make	cc	1000

Wave length in millimicrons	Absorbancy A_ϵ	$\frac{A_\epsilon}{14.481}$	Transmittancy[3] T_ϵ	Wave length in millimicrons	Absorbancy A_ϵ	$\frac{A_\epsilon}{14.481}$	Transmittancy[3] T_ϵ
350	0.0038	0.0002624	0.991	600	0.0137	0.0009461	0.969
360	.0040	.0002762	.991	610	.0124	.0008563	.972
370	.0050	.0003453	.989	620	.0115	.0007941	.974
380	.0065	.0004489	.985	630	.0112	.0007734	.975
390	.0088	.0006077	.980	640	.0110	.0007596	.975
400	.0125	.0008632	.972	650	.0105	.0007251	.976
410	.0168	.0011601	.962	660	.0097	.0006698	.978
420	.0224	.0015469	.950	670	.0087	.0006008	.980
430	.0340	.0023479	.925	680	.0076	.0005248	.983
440	.0522	.0036047	.887	690	.0066	.0004558	.985
450	.0773	.0053380	.837	700	.0054	.0003729	.988
460	.1031	.0071197	.789	710	.0046	.0003177	.989
470	.1213	.0083765	.756	720	.0038	.0002624	.991
480	.1349	.0093157	.733	730	.0032		.993
490	.1472	.0101650	.713	740	.0030		.993
500	.1635	.0112907	.686	750	.0028		.994
510	.1742	.0120296	.670				
520	.1689	.0116636	.678	Hg 404.7	.0144		.967
530	.1452	.0100269	.716	Hg 435.8	.0437		.904
540	.1113	.0076859	.774	Hg 491.6	.1497		.708
550	.0775	.0053518	.837	He 501.6	.1661		.682
560	.0496	.0034252	.892	Hg 546.1	.0901		.813
570	.0308	.0021269	.932	Hg 578.0	.0219		.951
580	.0207	.0014295	.953	He 587.6	.0167		.962
590	.0158	.0010911	.964	He 667.8	.0089		.980

[1] These data apply accurately also throughout the visible spectrum to a similar solution made up with 10.3 g. of cobalt sulphate ($CoSO_4.7H_2O$).
[2] Chemical analysis showed a cobalt (plus nickel) content of 100.0 per cent of the theoretical, the ratio of nickel to cobalt (metals) being 1 to 200.
[3] These values of transmittancy are derived from the adopted values of absorbancy. $A_\epsilon = -\log_{10} T_\epsilon$.

TABLE 7.—*Standard spectrophotometric data on double cell filled with water*

Three glass plates, each 2.5 mm thick (borosilicate crown, refractive index, D line = 1.51), two 1.000 cm thicknesses of distilled water; two air-glass surfaces, four water-glass surfaces.

Wave length in millimicrons	Density D_ω	Transmission [1] T_ω	Wave length in millimicrons	Density D_ω	Transmission [1] T_ω
350	0.088	0.817	600	0.045	0.902
360	.070	.851	610	.045	.902
370	.059	.873	620	.045	.902
380	.053	.885	630	.0455	.901
390	.051	.889	640	.046	.900
400	.050	.891	650	.046	.900
410	.0495	.892	660	.046	.900
420	.049	.893	670	.046	.900
430	.0485	.894	680	.046	.900
440	.048	.895	690	.0465	.898
450	.0475	.896	700	.047	.897
460	.047	.897	710	.049	.893
470	.0465	.898	720	.054	.883
480	.046	.900	730	.060	.871
490	.0455	.901	740	.064	.863
500	.045	.902	750	.065	.861
510	.0445	.903			
520	.044	.904	Hg 404.7 [2]	.0498	.892
530	.044	.904	Hg 435.8	.0482	.895
540	.044	.904	Hg 491.6	.0454	.901
			He 501.6	.0449	.902
550	.044	.904			
560	.044	.904	Hg 546.1	.0440	.904
570	.044	.904	Hg 578.0	.0440	.904
580	.044	.904	He 587.6	.0444	.903
590	.0445	.903	He 667.8	.0460	.900

[1] Values of transmission are derived from the adopted values of density. $D_\omega = -\log_{10} T_\omega$.
[2] These values for the Hg and He wave lengths are not fundamental experimental values as in the case of the solutions, tables 4 to 6, but are interpolated from the adopted density values above from 350mμ to 750mμ.

3. BEER'S LAW CORRECTION FOR SOLUTION A

Let **A** be the absorbancy, at any wave length, of a solution having a concentration of c grams per liter, and $\mathbf{A}_s$ the standard absorbancy at the same wave length, for a concentration c_s (=3.000 g) as given in Table 4, thicknesses in both cases being equal to 10 mm. If there is no Beer's law failure,

$$\mathbf{A} = \frac{c}{c_s}\mathbf{A}_s$$

If B represents the Beer's law correction factors given in Table 4, then B may be defined as

$$B = \frac{\mathbf{A}\left(\frac{c_s}{c}\right) - \mathbf{A}_s}{(c - c_s)\,\mathbf{A}_s}$$

Rearranging this equation gives

$$\mathbf{A} = \frac{\mathbf{A}_s}{c_s} c\,[1 + B\,(c - c_s)]$$

In the case of the **A** solutions, B is a negative quantity, as may be seen by reference to the experimental data of Table 17 (see Sec. VII, 6), which results in the correction being negative when $c > c_s$ and positive when $c < c_s$.

This formula was used in all computations of A, $\frac{A_s}{c_s}$ being computed and tabulated once for all as in Table 4. The fact that B is constant from 450mμ to 720mμ simplified the work of correction.

4. DESIGN OF FILTERS OF SERIES I—2,300°–4,000° K. TO MEAN NOON SUNLIGHT

(a) TO OBTAIN SPECTRAL ENERGY MATCH

Obviously the first step in the design of a filter to convert one spectral energy distribution, E, to another, E', is the computation of the relative spectral transmission of the ideal filter. This is obtained by computing, wave length by wave length, values of E'/E. These values of E'/E have a value of unity at 560mμ since the relative energies are all given with reference to 100 at this wave length. They may be reduced to a practical working basis by multiplying them by an appropriate factor, known by trial and experience to bring the relative transmissions, on the average, close to the spectral transmission of the final filter.

The next problem is to find what concentrations of the component solutions, A, B_1, and B_2 (considering also, of course, the data for the double cell filled with water) will best approximate the desired transmission or density.

The method devised for comparing the density of the ideal filter with the combined absorbancies of the three solutions at the trial concentrations plus the density of the cell filled with water is, briefly, as follows: A table of spectral absorbancies was prepared for each of the three solutions at a series of concentrations for the 17 spectral positions (every 20mμ) from 400mμ to 720mμ, inclusive. The concentrations of solution A were in 0.1 g steps, and of solutions B_1 and B_2 in 1.0 g steps. The absorbancies for each concentration were placed in horizontal rows, the values in the rows being arranged in order of wave lengths. A table of the spectral density values of the various ideal filters was also prepared.

These tables were mounted on rollers and placed in a box with a cover having slots so placed that one line of figures across each table is exposed to view. This is illustrated in Figure 11. The top roller contains the table of the spectral densities of the various ideal filters, the second line of figures, not mounted on a roller, are density values of the cell filled with water. The next three rollers contain the tables of the spectral absorbancies for solutions A, B_1, and B_2, respectively. The rollers may be turned by the handles outside of the box to expose to view any of the rows (horizontal) of figures. The cover of the box is ruled into 18 columns (vertical). In the first column, on the rollers, are the concentrations of the solutions in grams per liter and the name of the ideal filter. In the next 17 columns are the densities and absorbancies corresponding to the 17 wave lengths indicated at the tops of the columns. The operation of the device is as follows:

The three lower rollers are turned so as to expose the absorbancies corresponding to the concentrations of the solutions desired for the first trial. Next is found the difference between the density of the ideal filter and the sum of the absorbancies of the three solutions plus the density of the cell at each wave length. This requires but a

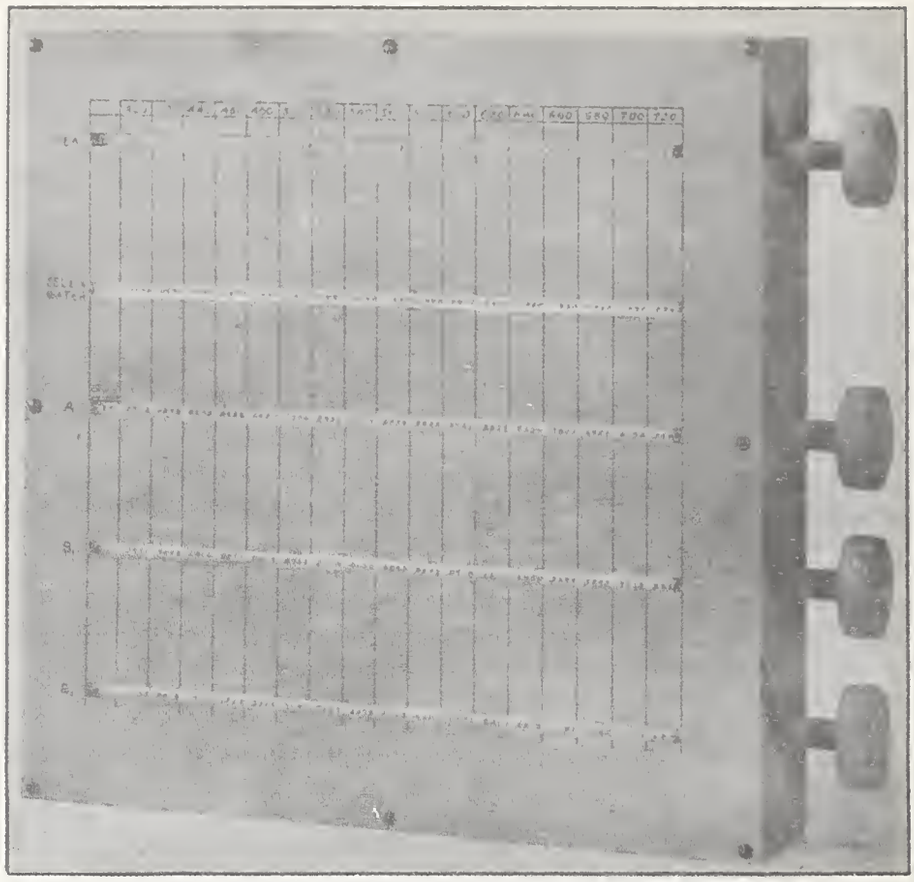


FIGURE 11.—*Computing device used in design of filters*

Detailed explanation is given in Section V, 4 (a).

few minutes with an adding or computing machine. The density differences are plotted on coordinate paper as a function of wave length. An ideal solution of the problem would be obtained if the density differences were equal in amount at all wave lengths.

Having thus plotted the density differences for the first trial, it is a relatively simple matter to determine what changes to make to improve the results, provided the general shape of the curve of each of the ingredients is kept in mind. These adjustments are made and the operation repeated until no further improvement can apparently be made. The length of time required to obtain satisfactory results by this method depends largely upon experience, from two to four hours work being usually necessary.

Other methods were considered before the one just described was adopted. For example, at each wave length there are three variables, and it would, therefore, be possible to choose three wave lengths fairly well distributed over the desired spectral range and make the actual and ideal values of density come into perfect agreement at these three wave lengths. However, the values at the remaining wave lengths are beyond control; that is, the actual density curve makes excursions from the ideal at various points and the magnitude of these excursions will depend upon the three wave lengths chosen for the computation.

The method of least squares is also applicable to the problem, but is quite complicated because the importance at a given wave length of obtaining a close approximation varies with the color temperature of the source. On this account a different set of weights would have to be employed for each source. It is not feasible to select one set of weights for each range of color temperature, because the continuity of the filters would be spoiled and interpolation for intermediate values would be difficult. The least square method appeared to be an unnecessary refinement.

The method finally adopted for solving the problem has advantages over either of these methods, chiefly because of its simplicity and flexibility. It allows some choice in the selection of concentrations so that a certain amount of control is possible over the light transmission of the filter. If the filter solutions had spectral properties such that a perfect match with the ideal filter could be obtained, no choice in concentrations would exist. However, this is not the case, since the actual densities depart from the ideal in several parts of the spectrum even at the most favorable concentrations. With low concentrations a good match in the central part of the visible spectrum is obtained with considerable departures at the ends; with higher concentrations the ends are improved considerably with increased departures in the central part of the spectrum.

Four filters, viz, 2,360°, 2,600°, 2,800°, and 3,000° K. to mean sun, were first designed by the adopted method. The concentrations of the three solutions as abscissas and the color temperature as ordinates were plotted and curves drawn. These curves gave trial data for intermediate filters, final adjustment being made as before. The range of the filters was extended to 3,500° K. and later to 4,000° K. by the same process. All of the data so obtained were again plotted (concentrations against color temperature) and smooth average curves drawn. From these curves the concentrations were read, at every 100° between the limits of 2,300° and 4,000° K., which were

used for deriving the data for the trial computations of the colorimetric characteristics, as outlined in the next section.

(b) TO OBTAIN COLOR MATCH

It is possible for two spectral energy distributions to be widely different and still give the same color. If a color match between the artificial and adopted sunlight were the only concern, solution B_1, for example, could be dispensed with. However, it is desirable for both photographic sensitometry and colorimetry that the spectral energy match be made as good as possible. This has been accomplished as outlined in the preceding section.

If it were possible to obtain, with the materials available, a perfect energy match with the ideal, a perfect color match would thereupon be automatically obtained. However, as previously noted, the energy distributions resulting from the process described in the preceding section depart from the ideal in several parts of the spectrum. From the magnitude and distribution of these deviations large color differences between the filtered energy and the ideal energy are not to be expected, and perfect color matches have been obtainable in all cases by making minor adjustments of the concentrations of one or more of the filter ingredients. Such adjustments do not materially change the energy matches previously obtained.

The method used for computing the color characteristics of any energy distribution follows the general description of such procedure as given in the Optical Society of America colorimetry report (4). The method makes use of the three so-called elementary excitations (red, green, and blue), based on the data of Abney and of Koenig and Dieterici as reduced by Weaver (4). These are given in Table 6 of the colorimetry report, but have been extrapolated and republished in the Optical Society of America spectrophotometry report (35), Table 5, page 230. These extrapolated excitations have been used in the present computations.

In addition to the choice of the red, green, and blue excitations, a spectral distribution defined as "neutral" (white) must be adopted. For this purpose the energy distribution of mean noon sunlight as given in Table 1 has been used.[13] The methods of computation were as follows:

---

[13] For the purposes of this paper, the computation of trilinear coordinates throughout the paper has been based on the arbitrary convention of making the center of the Maxwell triangle ($r=g=b=\frac{1}{3}$) represent the spectral distribution which we have chosen for mean sun; that is, mean sun has been taken as the "neutral stimulus." The decision as to what stimulus should be given this central position is conditioned upon the specific purposes which the system of colorimetric coordinates is to serve.

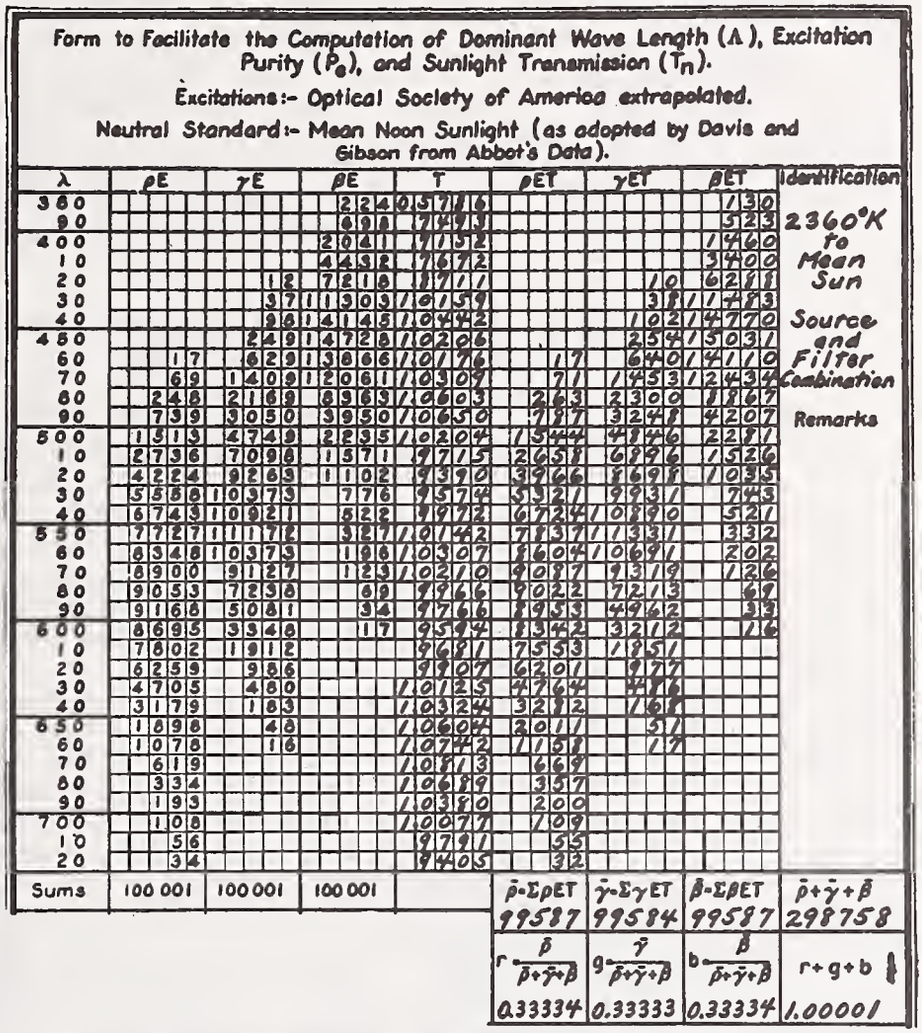
The question as to what stimulus should be assigned this position for colorimetric purposes, in general, is at present being discussed by a committee of the International Commission on Illumination, but no final recommendation has been made; and the question as to whether "noon sun" or "overcast sky" or "sun outside the earth's atmosphere" or a "black body" at some specified temperature should be formally chosen for this purpose is still in abeyance.

The choice of spectral distribution which we have made to represent "mean sun" has been based primarily on the requirements which a standard for photographic sensitometry must meet. For some years previously a slightly different spectral distribution of energy proposed by Priest (3) has served as a standard for "average noon sunlight at Washington" in the colorimetry section of the bureau and, to some extent, elsewhere. Priest sought to represent, for colorimetric purposes, the same data which we have sought to represent for the purposes of photographic sensitometry; the spectral distribution of energy he adopted is designated as "Abbot-Priest sun" throughout the present paper.

It may be noted that if "noon sun" should be formally chosen by the International Committee on Illumination to be placed at the center of the Maxwell triangle, Davis-Gibson mean sun would be equally as eligible to represent "noon sun" for this purpose as a black body at 5,000°K., "O. S. A. noon sun," or Abbot-Priest sun.

Naturally, for the purposes of this paper, it has been convenient to represent our ideal energy distribution (Davis-Gibson mean sun) by $r=g=b=\frac{1}{3}$; that is, it has been convenient to make Davis-Gibson mean sun the "neutral stimulus." Because of this difference in "neutral stimulus," the results of our computations of trilinear coordinates differ slightly from those which would be obtained on the basis of Abbot-Priest sun, which is placed at the center of the Maxwell triangle in most other colorimetric computations at the Bureau of Standards.

The Optical Society of America extrapolated excitations, as published, are weighted on the basis of an energy distribution of equal value at all wave lengths. On the form[14] shown in Figure 12 these excitation values have been multiplied, wave length by wave length, by the mean-sun relative energy data and the resulting values further



Form to Facilitate the Computation of Dominant Wave Length (λ), Excitation Purity (P_e), and Sunlight Transmission (T_n).

Excitations:- Optical Society of America extrapolated.

Neutral Standard:- Mean Noon Sunlight (as adopted by Davis and Gibson from Abbot's Data).

λ	ρE	γE	βE	T	ρET	γET	βET	Identification
360			2240	.05716			130	2360°K to Mean Sun Source and Filter Combination
90			698	.07493			523	
400			2041	.11552			1460	
10			4432	.17672			3700	
20		12	7218	.21711		10	6218	
30		37	11303	.10159		38	11483	
40		98	14145	.10442		102	14770	
450		249	14728	.10206		254	15031	
60	17	629	13866	.10176	17	640	14110	
70	69	1409	12061	.10309	71	1453	12434	
80	248	2169	8963	.06603	263	2300	8867	
90	739	3050	3950	.06550	787	3248	4207	Remarks
500	1313	4749	2235	.10204	1344	4846	2281	
10	2736	7098	1571	.07715	2658	6996	1526	
20	4224	9263	1102	.09390	3966	8698	1035	
30	5586	10373	776	.09574	5327	9931	743	
40	6743	10921	522	.09772	6724	10890	521	
550	7727	11172	327	.10442	7837	11331	332	
60	6348	10373	196	.10307	6607	10691	202	
70	8900	9127	129	.10210	9087	9319	126	
80	9053	7238	69	.09766	9022	7213	67	
90	9168	5081	34	.07766	8953	4962	33	
600	8695	3346	17	.09544	8342	3212	16	
10	7802	1912		.09681	7553	1851		
20	6259	986		.09107	6201	977		
30	4705	480		.10125	4764	486		
40	3179	183		.10324	3282	189		
650	1898	48		.10604	2011	51		
60	1078	16		.10742	1158	17		
70	619			.10813	669			
80	334			.10689	357			
90	193			.10380	200			
700	108			.10077	109			
10	56			.09791	55			
20	34			.09405	32			
Sums	100 001	100 001	100 001		$\bar{\rho}\cdot\Sigma\rho ET$ 99587	$\bar{\gamma}\cdot\Sigma\gamma ET$ 99584	$\bar{\beta}\cdot\Sigma\beta ET$ 99587	$\bar{\rho}+\bar{\gamma}+\bar{\beta}$ 298758
					$r=\frac{\bar{\rho}}{\bar{\rho}+\bar{\gamma}+\bar{\beta}}$ 0.33334	$g=\frac{\bar{\gamma}}{\bar{\rho}+\bar{\gamma}+\bar{\beta}}$ 0.33333	$b=\frac{\bar{\beta}}{\bar{\rho}+\bar{\gamma}+\bar{\beta}}$ 0.33334	r+g+b 1.00001

FIGURE 12.—*Computation of r, g, and b for the 2,360° K.-to-mean-sun filter*

In column headed T are given values of the ratio

$$\frac{T_{\text{filter}} E_{2,360}}{E'} = \frac{E''}{E'}$$

where, as a function of wave length, T_{filter} is the transmission of the filter, $E_{2,360}$ is the energy distribution for 2,360° K., and E' is the energy distribution of mean sun, the "neutral" stimulus. The values of $\frac{E''}{E'}$ are taken from the last column of Table 9, and are as given in chart 2 (with one less significant figure). (It should be noted that values of r, g, and b, computed on the basis of mean sun, E', as the "neutral" stimulus will, in general, be slightly different from those computed on the basis of Abbot-Priest sun (see fig. 3))

adjusted, respectively, by three factors so chosen that the three sums give the same total. These adjusted values are given in Figure 12 in the columns headed ρE, γE, and βE, totaling to 100,000 each.

[14] To facilitate the computations a form designed by Priest with Abbot-Priest sunlight as the "neutral" stimulus was revised by Davis using the adopted mean sunlight as the "neutral" stimulus. Part of this form thus revised is shown in Figure 12.

518-743 O - 73 - 45

If the specification of the color of a glass were to be computed on the basis of mean sun as the "neutral" stimulus, its spectral transmission would be entered in the column headed T. If the color corresponding to an energy distribution (such as a lamp-and-filter combination) is to be specified, the ratios of that energy distribution to mean sun must be obtained, wave length by wave length. As an example, this ratio in the case of the 2,360° K.-to-mean-sun filter is shown tabulated in Figure 12, under column T. These values are the same as the values of E''/E' given in chart 2, except that they are carried to an additional decimal place.

The products ρET, γET, and βET were then computed as illustrated and the respective sums $\overline{\rho}$, $\overline{\gamma}$, $\overline{\beta}$ were then taken, $\overline{\rho}+\overline{\gamma}+\overline{\beta}$ obtained, and the trilinear coordinates (r, g, b) computed, all as illustrated in Figure 12. In the case of mean sun, r, g, and b are each, by definition, exactly equal to one-third.

The first step in changing the energy distribution of any particular source-and-filter combination so that its color should agree with that of the adopted mean sun was to compute the r, g, and b values for this energy distribution as shown in Figure 12. In general, although the spectral energy distribution of the source-and-filter combination had been adjusted to as close an agreement with that of mean sun as was possible by the method outlined in the preceding section, the values of r, g, and b so obtained were not exactly 0.33333. The next step thereupon was to make still further adjustment, in all cases small, of one or more of the concentrations of the ingredients until the values of r, g, and b for the source-and-filter combination differed from 0.33333 by less than 0.00006.

These final adjustments were facilitated by plotting the r-b coordinates to a large scale near the "neutral" point. It was soon found that changes in the concentration of any one of the three components always shifted the point on the diagram along a certain direction and by an amount practically proportional to the change in concentration. After a little experience it was usually possible to obtain the final desired values with three or four trials.

In addition to choosing concentrations which would give correct values of r, g, and b, a further restriction was made, viz, that when these concentrations, for the respective A, B_1 and B_2 components, were plotted against color temperature they must fall upon smooth curves, so that values of concentrations for any intermediate temperature read from such curves would fit within the series.

As finally adopted, values of concentrations were obtained which gave (1) correct values of r, g, and b, viz, 0.3333, (2) smooth curves when plotted against color temperatures, and (3) energy distributions as illustrated in the charts. This was done at every 100° K. from 2,300° to 4,000° K., besides 2,360° and 2,848° K. These values of concentration are given in Table 8, the values at intermediate color temperatures being obtained from the concentration curves by interpolation.

A detailed example of the computation of E'' and E''/E' is given in Table 9, which is self-explanatory. The 2,360° K-to-mean-sun filter (chart 2) was chosen for this purpose. The data shown are for the respective concentrations as finally determined, giving values (fig. 12) of r, g, $b=0.3333$. The same laborious computation had to be carried through at each set of trial concentrations in order to get

values of E''/E' for the trial computations of r, g, and b. A procedure similar to that herein described was carried out for all of the 38 filters illustrated in the charts.

TABLE 3.—*Concentrations and light transmissions of filters of Series I, 2,300°–4,000° K. to mean noon sunlight* [1]

Color temperature (°K.) of source	Grams of copper sulphate and grams of mannite per liter in solution A	Grams of cobalt ammonium sulphate per liter in solution B	Grams of copper sulphate per liter in solution B	Light transmission	Color temperature (°K.) of source	Grams of copper sulphate and grams of mannite per liter in solution A	Grams of cobalt ammonium sulphate per liter in solution B	Grams of copper sulphate per liter in solution B	Light transmission
2,300	3.901	28.300	28.250	0.1224	3,200	1.784	10.890	14.570	.370
2,320	3.835	27.80	27.90	.1266	3,220	1.751	10.60	14.34	.377
2,340	3.770	27.31	27.54	.1309	3,240	1.719	10.32	14.11	.383
2,360	3.707	26.827	27.180	.1352	3,260	1.687	10.03	13.88	.390
2,380	3.644	26.35	26.83	.1397	3,280	1.655	9.75	13.66	.397
2,400	3.582	25.870	26.480	.1442	3,300	1.624	9.470	13.430	.403
2,420	3.522	25.40	26.13	.1487	3,320	1.594	9.19	13.21	.410
2,440	3.463	24.94	25.78	.153	3,340	1.564	8.92	12.99	.417
2,460	3.405	24.49	25.43	.158	3,360	1.534	8.64	12.77	.423
2,480	3.348	24.04	25.08	.163	3,380	1.504	8.37	12.55	.430
2,500	3.292	23.590	24.730	.168	3,400	1.475	8.100	12.330	.437
2,520	3.237	23.14	24.38	.173	3,420	1.446	7.83	12.11	.444
2,540	3.183	22.70	24.04	.177	3,440	1.417	7.57	11.90	.451
2,560	3.128	22.27	23.69	.183	3,460	1.388	7.31	11.68	.457
2,580	3.076	21.83	23.34	.188	3,480	1.360	7.05	11.47	.464
2,600	3.024	21.400	23.000	.193	3,500	1.332	6.790	11.260	.471
2,620	2.973	20.97	22.66	.198	3,520	1.305	6.53	11.05	.478
2,640	2.923	20.56	22.32	.203	3,540	1.278	6.29	10.84	.485
2,660	2.874	20.14	21.98	.209	3,560	1.251	6.05	10.64	.492
2,680	2.825	19.73	21.64	.214	3,580	1.225	5.81	10.43	.499
2,700	2.777	19.325	21.300	.220	3,600	1.199	5.570	10.220	.506
2,720	2.730	18.92	20.97	.225	3,620	1.173	5.33	10.02	.513
2,740	2.684	18.54	20.64	.231	3,640	1.148	5.09	9.82	.520
2,760	2.639	18.16	20.33	.236	3,660	1.123	4.85	9.62	.527
2,780	2.595	17.78	20.02	.242	3,680	1.098	4.61	9.42	.534
2,800	2.551	17.400	19.720	.248	3,700	1.074	4.380	9.210	.541
2,820	2.507	17.02	19.44	.253	3,720	1.050	4.15	9.02	.548
2,840	2.464	16.66	19.15	.259	3,740	1.026	3.92	8.81	.556
2,860	2.421	16.30	18.86	.265	3,760	1.002	3.70	8.61	.563
2,880	2.379	15.94	18.58	.271	3,780	.978	3.48	8.41	.570
2,900	2.337	15.580	18.300	.277	3,800	.955	3.260	8.210	.577
2,920	2.296	15.23	18.03	.283	3,820	.932	3.04	8.01	.584
2,940	2.256	14.89	17.77	.289	3,840	.910	2.82	7.82	.592
2,960	2.217	14.56	17.51	.295	3,860	.888	2.60	7.62	.599
2,980	2.178	14.23	17.25	.301	3,880	.866	2.39	7.42	.606
3,000	2.140	13.900	17.000	.307	3,900	.845	2.175	7.220	.613
3,020	2.102	13.58	16.74	.313	3,920	.824	1.96	7.03	.620
3,040	2.065	13.27	16.49	.320	3,940	.803	1.75	6.83	.628
3,060	2.028	12.97	16.24	.326	3,960	.782	1.54	6.64	.635
3,080	1.992	12.66	16.00	.332	3,980	.761	1.34	6.45	.642
3,100	1.956	12.360	15.760	.338	4,000	.741	1.130	6.250	.649
3,120	1.921	12.06	15.51	.335					
3,140	1.886	11.76	15.27	.351	2,848	2.445	16.520	19.020	.262
3,160	1.851	11.47	15.04	.357					
3,180	1.817	11.18	14.80	.364					

[1] Spectrophotometric illustrations are given in charts 1 to 20; methods of computation are described in Section V, 4; details of preparation are given in Section IX.

TABLE 9.—*Computation of energy distribution of source-and-filter combination, E'' of Chart 2, 2,360° K. to mean sun, starting with fundamental data of Tables 2, and 4 to 7*

Formulas:

(1) $E''=KT_FE_{2,360}$ (see chart 2), or $\log_{10}E''=K'+\log_{10}T_F+\log_{10}E_{2,360}$, or $\log_{10}\left(\frac{E''}{K''}\right)=K'''+\log_{10}T_F+\log_{10}E_{2,360}$. Values of $E_{2,360}$ given in Table 2.

(2) $-\log_{10}T_F=D_F$.

(3) $D_F=D_W+A_A+A_{B1}+A_{B2}$. Symbols defined in Section V, 2. Values of D_W given in Table 7.

(4) $A_A=\frac{A_s}{c_s}c[1+B(c-c_s)]=\frac{A_s}{c_s}c'$. Values of $\frac{A_s}{c_s}$ and B given in Table 4. $c=3.707$, $c_s=3.000$, c' tabulated below.

(5) $A_{B1}=\frac{A_s}{20}\times 27.180$. Values of $\frac{A_s}{20}$ given in Table 5.

(6) $A_{B2}=\frac{A_s}{14.481}\times 26.827$. Values of $\frac{A_s}{14.481}$ given in Table 6.

c'	λ	A_A	A_{B1}	A_{B2}	D_W	D_F	T_F[1]	$+\log_{10}T_F$	$\log_{10}E_{2,360}$	$\log_{10}\left(\frac{E''}{K''}\right)$ ($K'''=0.8435$)	$\frac{E''}{K''}$	E''_1 ($K''=0.988616$)	E' (from Table 1)	$E''-E'$ (see footnote, chart 2)	$\frac{E''_1}{E'}$
	$m\mu$														
3.392	350	0.5314	0.0122	0.0070	0.0880	0.6386	0.230	9.3614−10	0.1903	0.3952	2.48	2.45			
3.392	60	.4070	.0086	.0074	.0700	.4930	.321	.5070−10	.3404	.6909	4.91	4.85	15.97	−11.12	0.3037
3.398	70	.3024	.0063	.0093	.0590	.3770	.420	.6230−10	.4786	.9451	8.81	8.71	20.47	−11.76	.4255
3.408	80	.2158	.0048	.0120	.0530	.2856	.518	.7144−10	.6085	1.1664	14.67	14.50	25.06	−10.56	.5786
3.432	90	.1453	.0038	.0163	.0510	.2164	.608	.7836−10	.7308	1.3579	22.80	22.54	30.08	−7.54	.7493
3.497	400	.0973	.0031	.0232	.0500	.1736	.671	.8264−10	.8445	1.5144	32.69	32.32	45.19	−12.87	.7152
3.576	10	.0632	.0026	.0311	.0495	.1464	.714	.8536−10	.9513	1.6484	44.50	43.99	57.34	−13.35	.7672
3.623	20	.0399	.0022	.0415	.0490	.1326	.737	.8674−10	1.0527	1.7636	58.02	57.36	65.85	−8.49	.8711
3.655	30	.0262	.0019	.0630	.0485	.1396	.725	.8604−10	1.1480	1.8519	71.11	70.30	69.20	+1.10	1.0159
3.673	40	.0202	.0016	.0967	.0480	.1665	.682	.8335−10	1.2375	1.9145	82.13	81.20	77.76	+3.44	1.0442
3.681	450	.0209	.0015	.1432	.0475	.2131	.612	.7869−10	1.3222	1.9526	89.66	88.64	86.85	+1.79	1.0206
3.681	60	.0282	.0015	.1910	.0470	.2677	.540	.7323−10	1.4017	1.9775	94.95	93.87	92.25	+1.62	1.0176
3.681	70	.0436	.0016	.2247	.0465	.3164	.483	.6836−10	1.4776	2.0047	101.09	99.94	96.94	+3.00	1.0309
3.681	80	.0681	.0019	.2499	.0460	.3659	.431	.6341−10	1.5486	2.0262	106.22	105.01	99.04	+5.97	1.0603
3.681	90	.1043	.0024	.2727	.0455	.4249	.376	.5751−10	1.6162	2.0348	108.34	107.11	100.57	+6.54	1.0650
3.681	500	.1509	.0035	.3029	.0450	.5023	.315	.4977−10	1.6802	2.0214	105.05	103.85	101.77	+2.08	1.0204
3.681	10	.2147	.0052	.3227	.0445	.5871	.259	.4129−10	1.7410	1.9974	99.40	98.27	101.15	−2.88	.9715
3.681	20	.2945	.0075	.3129	.0440	.6589	.219	.3411−10	1.7982	1.9828	96.12	95.03	101.20	−6.17	.9390
3.681	30	.3816	.0107	.2690	.0440	.7053	.197	.2947−10	1.8524	1.9906	97.86	96.75	101.05	−4.30	.9574
3.681	40	.4748	.0151	.2062	.0440	.7401	.182	.2599−10	1.9043	2.0077	101.79	100.63	100.91	−.28	.9972

3.681	550	.5693	.0211	.1436	.0440	.7780	.167	.2220−10	1.9530	2.0185	104.35	103.16	101.72	+1.44	1.0142
3.681	60	.6601	.0294	.0919	.0440	.8254	.1495	.1746−10	2.0000	2.0181	104.26	103.07	100.00	+3.07	1.0307
3.681	70	.7399	.0397	.0571	.0440	.8807	.1316	.1193−10	2.0441	2.0069	101.60	100.44	98.37	+2.07	1.0210
3.681	80	.8037	.0530	.0383	.0440	.9390	.1151	.0610−10	2.0866	1.9911	97.97	96.85	97.18	−.33	.9966
3.681	90	.8503	.0704	.0293	.0445	.9945	.1013	.0055−10	2.1764	1.9754	94.49	93.41	95.65	−2.24	.9766
3.681	600	.8798	.0924	.0254	.0450	1.0426	.0907	8.9574−10	2.1646	1.9655	92.36	91.31	95.17	−3.86	.9594
3.681	10	.8908	.1203	.0230	.0450	1.0791	.0834	.9209−10	2.2008	1.9652	92.30	91.25	94.26	−3.01	.9681
3.681	20	.8896	.1529	.0213	.0450	1.1088	.0778	.8912−10	2.2354	1.9701	93.35	92.29	93.16	−.87	.9907
3.681	30	.8761	.1943	.0207	.0455	1.1366	.0730	.8634−10	2.2680	1.9749	94.38	93.31	92.16	+1.15	1.0125
3.681	40	.8540	.2446	.0204	.0460	1.1650	.0684	.8350−10	2.2994	1.9779	95.04	93.96	91.01	+2.95	1.0324
3.681	650	.8196	.3044	.0195	.0460	1.1895	.0646	.8105−10	2.3291	1.9831	96.18	95.09	89.67	+5.42	1.0604
3.681	60	.7816	.3724	.0180	.0460	1.2180	.0605	.7820−10	2.3576	1.9831	96.18	95.09	88.52	+6.57	1.0742
3.681	70	.7399	.4512	.0161	.0460	1.2532	.0558	.7468−10	2.3850	1.9753	94.47	93.39	86.37	+7.02	1.0813
3.681	80	.6994	.5327	.0141	.0460	1.2922	.0510	.7078−10	2.4105	1.9618	91.58	90.54	84.70	+5.84	1.0689
3.681	90	.6577	.6238	.0122	.0465	1.3402	.0457	.6598−10	2.4353	1.9386	86.82	85.83	82.69	+3.14	1.0380
3.681	700	.6147	.7162	.0100	.0470	1.3879	.0409	.6121−10	2.4587	1.9143	82.09	81.16	80.54	+.62	1.0077
3.681	10	.5742	.8045	.0085	.0490	1.4362	.0366	.5638−10	2.4809	1.8882	77.30	76.42	78.05	−1.63	.9791
3.681	20	.5337	.8915	.0070	.0540	1.4862	.0328	.5138−10	2.5023	1.8596	72.38	71.56	76.09	−4.53	.9405

$\sum_{400}^{720}$=2, 956. 03 2, 922. 40 2, 922. 38 +0. 02

2, 922. 38

2, 956. 03=0. 988616.

[1] These values are tabulated in chart 2. Values of E'' and E''/E' in chart 2, however, are given to one less figure than given above. Values of E''/E' are used in the computation of r, g, and b, as shown in Figure 12.

The complete computation of a filter is a long process and the chances for error are, of course, correspondingly large. To avoid any reasonable possibility of such errors, even though computing machines

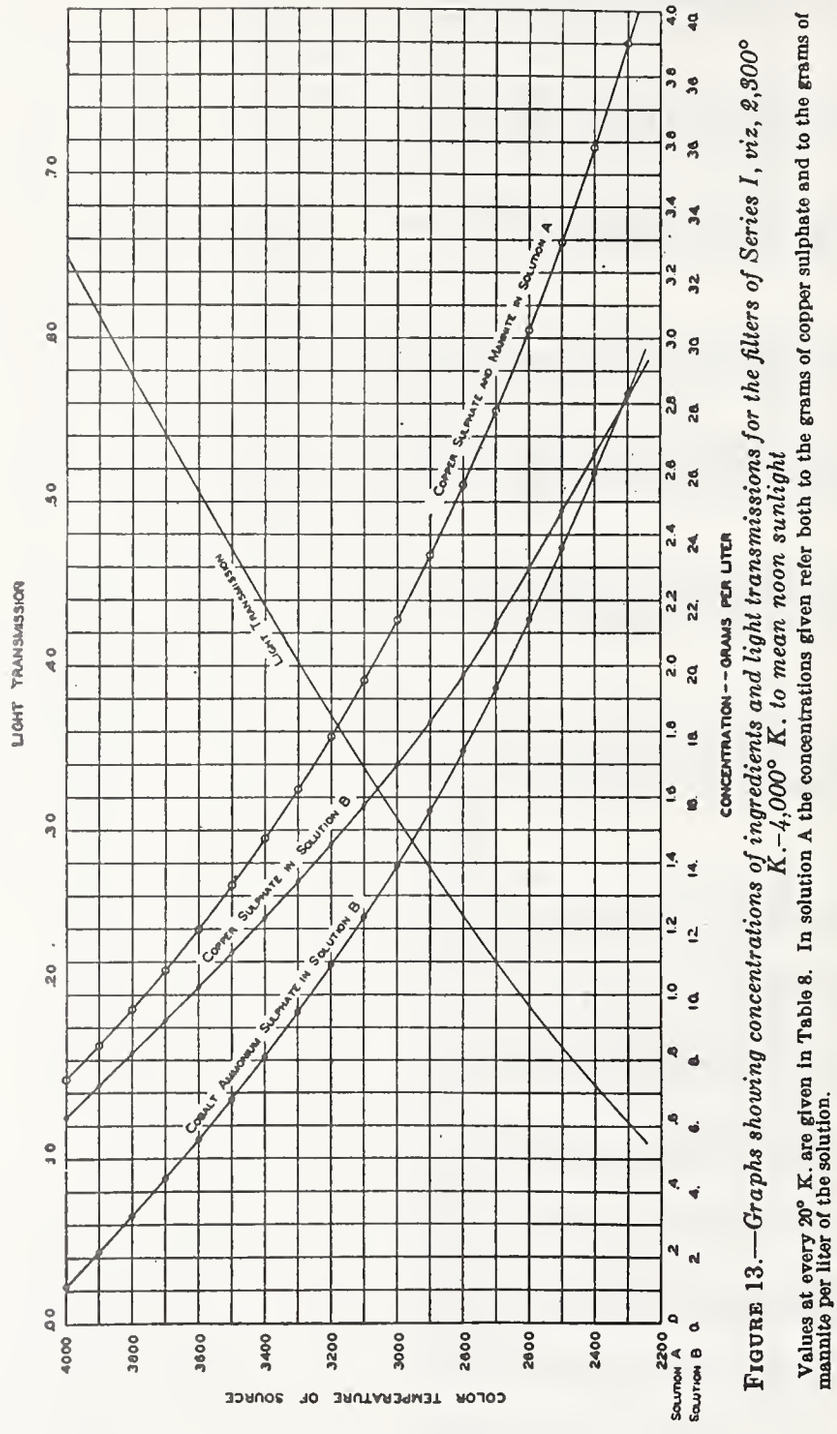


FIGURE 13.—*Graphs showing concentrations of ingredients and light transmissions for the filters of Series I, viz, 2,300° K.–4,000° K. to mean noon sunlight*

Values at every 20° K. are given in Table 8. In solution A the concentrations given refer both to the grams of copper sulphate and to the grams of mannite per liter of the solution.

were used in all computations, complete independent recomputations were made on each filter, starting with the fundamental spectrophotometric data.

The filters of Series I are described and illustrated in charts 1 to 20, inclusive, and in Table 8 and Figure 13.

(c) TO COMPUTE LIGHT TRANSMISSION OF FILTERS

It is customary in photographic sensitometry to specify the intensity of the light source in terms of visual candlepower. If the light used is obtained from a source-and-filter combination, it is essential (in accord with this custom) to know the effective candlepower of this source-and-filter combination. If the transmission of the filter for the light given by the source is known, the candlepower of the combination may be obtained by multiplying the candlepower of the source by the transmission of the filter. The transmissions of the filters are also of interest in their applications to colorimetry.

The light transmission of the filters may, of course, be obtained directly by photometric measurements, as, for example, by the flicker method. However, in the case of colored filters the results may be influenced by the photometric field size and brightness and may also vary from one observer to another because of the different values of relative visibility possessed even by observers of normal color sense.

This difficulty may be avoided by computing the light transmission from the spectral transmission, using standard visibility data. The values of light transmission so obtained should, for the present type of filters, be practically identical with the values that would be obtained if the light transmissions had been directly measured under standard photometric conditions by a large number of observers and averaged.

The light transmissions for the present filters have been computed by the following formula:

$$T_\theta = \frac{\sum_{400}^{720} (E_\theta V T)}{\sum_{400}^{720} (E_\theta V)}$$

The summations were carried out using values at every 10mμ from 400mμ to 720mμ. In the formula θ refers to the color temperature of the light source, this varying from 2,300° to 4,000° K. T_θ is the light transmission of the filter for a θ° K. source, T is the spectral transmission tabulated on the charts, E_θ is the relative energy for θ° K. as given in Table 2, and V is the relative visibility function as given in Table 10. These values of relative visibility are those recommended by Gibson and Tyndall (11, 12) and adopted by the International Commission on Illumination (10) and by the American Engineering Standards Committee and by the Illuminating Engineering Society (38).

The visibility data in Table 10 are given from 400mμ to 760mμ while the spectral range used to compute the light transmission of the filters does not go above 720mμ. It was found by trial that excluding values of T, E_θ, and V above 720mμ caused an error so small (not more than 0.00003 in the light transmission of any of the filters) that they could be safely omitted.

Values of light transmission for the various filters of Series I thus computed are given on the charts and in Table 8 and illustrated in Figure 13. Computations were carried out at the same color temperatures as used for the colorimetric computations. The intermediate values of Table 8 were obtained by plotting the computed values against color temperature and reading the desired values from this curve.

TABLE 10.—*Standard values of the relative visibility of radiant energy*

Wave length in millimicrons	Relative visibility [1]	Wave length in millimicrons	Relative visibility [1]
400	0.0004	600	0.631
410	.0012	610	.503
420	.0040	620	.381
430	.0116	630	.265
440	.023	640	.175
450	.038	650	.107
460	.060	660	.061
470	.091	670	.032
480	.139	680	.017
490	.208	690	.0082
500	.323	700	.0041
510	.503	710	.0021
520	.710	720	.00105
530	.862	730	.00052
540	.954	740	.00025
550	.995	750	.00012
560	.995	760	.00006
570	.952		
580	.870		
590	.757		

[1] These values were recommended by Gibson and Tyndall (B. S. Sci. Paper No. 475, p. 174; 1923; Trans. I. E. S., 19, pp. 192-193; 1924; Proc., International Commission on Illumination, sixth meeting, 1924, p. 235). They have been adopted as standard by the International Commission on Illumination (Proc., p. 67, sixth meeting) and by the American Engineering Standards Committee and the Illuminating Engineering Society (Trans. I. E. S., 20, pp. 629, 632; 1925).

5. DESIGN OF FILTERS OF SERIES II—2,848° TO 3,500°–10,000° K.

Because of the general interest in high color temperatures in the fields of colorimetry and astronomy and the limited means available for the reproduction of high color temperatures in the laboratory, it has seemed desirable to develop this second series of filters, in this case starting at a fixed lower color temperature of 2,848° K. By means of this series of filters, radiant energy evoking the color of a Planckian radiator at any temperature between 3,500° and 10,000° K. may be readily obtained in the laboratory. The particular starting point, 2,848° K. was chosen for the filters of Series II for the reason that it happens to be the color temperature of a certain primary standard lamp and thus came to be a common reference point. This particular color temperature is a very suitable one at which to operate gas-filled lamps.

Before proceeding with the computations for the filters of Series II, values of r, g, and b had to be computed for all the black-body temperatures between 3,500° and 10,000° K. for which filters were to be specifically designed. This was done and values of r, g, and b obtained as given in Table 11 and illustrated in Figure 14. (A similar illustration for color temperatures from 2,000° to 4,000° K. is given in the next section.)

After designing these filters for what seemed to be the best energy match, the colorimetric computations were carried out, striving now to make the values of r, g, and b not equal to 0.33333, as was the case

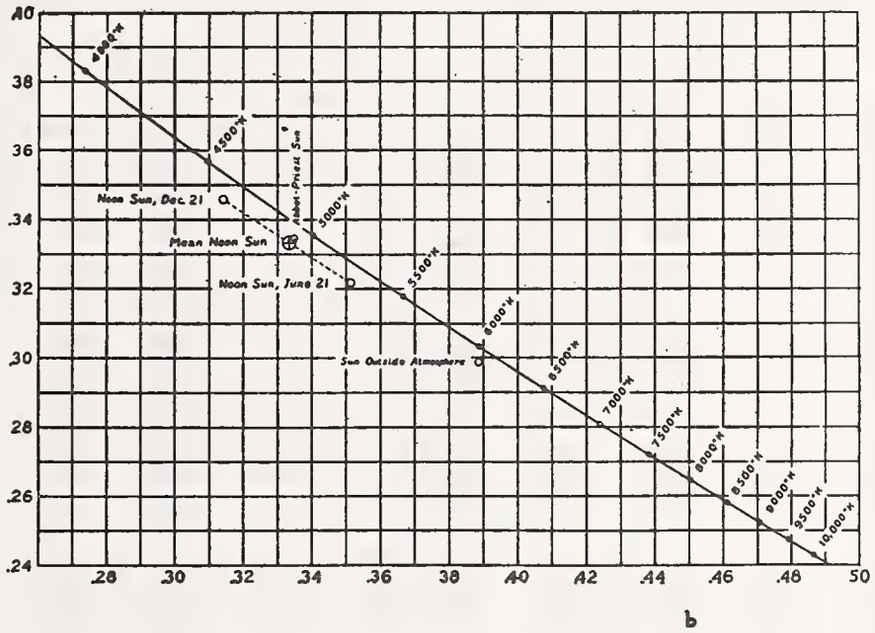


FIGURE 14.—*Section of the r–b diagram showing colorimetric positions of various phases of sunlight relative to the black-body locus*

(See fig. 16 for temperatures between 2,000° K. and 4,000° K.) Values of r, g, and b are given in Table 11, for every 20° K. up to 8,000° K. and for every 100° K. above that. All values of r, g, and b are computed on the basis of mean noon sunlight as the "neutral" stimulus.

with the filters of Series I, but equal to the respective values given in Table 11 for the particular color temperatures involved.

TABLE 11.—*Trilinear coordinates (r, g, b) of colors evoked by black-body radiant energy as a function of color temperature*

Color temperature [1]	Trilinear coordinates [2]			Color temperature	Trilinear coordinates [2]		
	r	g	b		r	g	b
°K.				°K.			
1,600	0.64249	0.32033	0.03718	1,750	0.6157	0.3343	0.0499
1,610	.6407	.3214	.0379	1,760	.6139	.3352	.0508
1,620	.6388	.3224	.0388	1,770	.6123	.3359	.0518
1,630	.6369	.3235	.0397	1,780	.6106	.3367	.0527
1,640	.6352	.3245	.0404	1,790	.6089	.3374	.0536
1,650	.6333	.3255	.0413	1,800	.60725	.33915	.05460
1,660	.6315	.3264	.0422	1,810	.6056	.3388	.0555
1,670	.6297	.3274	.0430	1,820	.6039	.3396	.0564
1,680	.6279	.3283	.0438	1,830	.6023	.3403	.0573
1,690	.6262	.3292	.0447	1,840	.6007	.3409	.0583
1,700	.62435	.33010	.04555	1,850	.5991	.3416	.0593
1,710	.6226	.3309	.0464	1,860	.5974	.3422	.0603
1,720	.6208	.3318	.0473	1,870	.5958	.3428	.0613
1,730	.6192	.3327	.0482	1,880	.5943	.3434	.0622
1,740	.6173	.3336	.0491	1,890	.5927	.3441	.0632

[1] C_2=14,350 micron-degrees. See Table 2.

[2] "Neutral" stimulus: Mean sun as adopted in this investigation. Excitations: O. S. A. extrapolated (J. O. S. A. and R. S. I., 10, p. 230; 1925). Computed values are given to five significant figures; other values, to four significant figures, were read from large scale graph plotted from computed values.

TABLE 11.—*Trilinear coordinates (r, g, b) of colors evoked by black-body radiant energy as a function of color temperature*—Continued

Color temperature [1]	Trilinear coordinates [2]			Color temperature [1]	Trilinear coordinates [2]		
	r	g	b		r	g	b
°K				°K			
1,900	0.59108	0.34470	0.06422	3,100	0.4480	9.3592	0.1927
1,910	.5895	.3452	.0652	3,120	.4463	.3589	.1947
1,920	.5879	.3458	.0662	3,140	.4446	.3587	.1967
1,930	.5864	.3464	.0672	3,160	.4429	.3584	.1987
1,940	.5848	.3469	.0682	3,180	.4411	.3581	.2007
1,950	.5833	.3475	.0692	3,200	.4394	.3578	.2027
1,960	.5818	.3480	.0702	3,220	.4377	.3575	.2047
1,970	.5802	.3486	.0712	3,240	.4360	.3572	.2067
1,980	.5787	.3491	.0722	3,260	.4344	.3569	.2086
1,990	.5772	.3495	.0733	3,280	.4328	.3566	.2106
2,000	.57571	.34998	.07431	3,300	.4312	.3563	.2125
2,020	.5727	.3509	.0764	3,320	.4297	.3560	.2144
2,040	.5697	.3518	.0785	3,340	.4281	.3557	.2163
2,060	.5668	.3526	.0806	3,360	.4265	.3553	.2182
2,080	.5639	.3534	.0827	3,380	.4249	.3550	.2201
2,100	.56101	.35416	.08483	3,400	.4234	.3546	.2220
2,120	.5582	.3549	.0870	3,420	.4219	.3543	.2239
2,140	.5554	.3555	.0891	3,440	.4203	.3540	.2258
2,160	.5526	.3562	.0912	3,460	.4188	3537	.2276
2,180	.5499	.3568	.0934	3,480	.4173	.3533	.2294
2,200	.54716	.35727	.09556	3,500	.41592	.35290	.23118
2,220	.5445	.3578	.0978	3,520	.4145	.3525	.2331
2,240	.5418	.3583	.1000	3,540	.4130	.3522	.2349
2,260	.5391	.3588	.1021	3,560	.4116	.3518	.2367
2,280	.5365	.3592	.1033	3,580	.4101	.3514	.2385
2,300	.5339	.3596	.1065	3,600	.4087	.3510	.2403
2,320	.5314	.3599	.1087	3,620	.4074	.3506	.2421
2,340	.5288	.3603	.1109	3,640	.4060	.3502	.2438
2,360	.52631	.36063	.11306	3,660	.4046	.3498	.2456
2,380	.5238	.3609	.1153	3,680	.4033	.3495	.2473
2,400	.52133	.36117	.11750	3,700	.4019	.3491	.2490
2,420	.5189	.3614	.1198	3,720	.4006	.3488	.2507
2,440	.5164	.3616	.1220	3,740	.3993	.3484	.2524
2,460	.5141	.3618	.1242	3,760	.3980	.3480	.2541
2,480	.5117	.3620	.1264	3,780	.3967	.3476	.2558
2,500	.50936	.36217	.12847	3,800	.3954	.3472	.2575
2,520	.5070	.3622	.1308	3,820	.3941	.3468	.2591
2,540	.5047	.3623	.1330	3,840	.3928	.3464	.2608
2,560	.5024	.3624	.1352	3,860	.3916	.3460	.2624
2,580	.5001	.3625	.1374	3,880	.3903	.3456	.2641
2,600	.49788	.36259	.13953	3,900	.3891	.3452	.2657
2,620	.4956	.3627	.1417	3,920	.3879	.3448	.2673
2,640	.4934	.3627	.1439	3,940	.3867	.3444	.2689
2,660	.4912	.3627	.1461	3,960	.3855	.3440	.2704
2,680	.4890	.3627	.1483	3,980	.3843	.3437	.2720
2,700	.48690	.36260	.15050	4,000	.38313	.34332	.27356
2,720	.4848	.3625	.1527	4,020	.3819	.3429	.2752
2,740	.4827	.3624	.1549	4,040	.3808	.3425	.2767
2,760	.4806	.3623	.1571	4,060	.3797	.3421	.2782
2,780	.4786	.3622	.1592	4,080	.3785	.3417	.2787
2,800	.47657	.36218	.16125	4,100	.3774	.3413	.2812
2,820	.4745	.3620	.1635	4,120	.3763	.3409	.2827
2,840	.4725	.3619	.1656	4,140	.3752	.3405	.2842
2,860	.4705	.3617	.1678	4,160	.3741	.3401	.2857
2,880	.4685	.3616	.1699	4,180	.3730	.3398	.2872
2,900	.46659	.36143	.17198	4,200	.3719	.3394	.2887
2,920	.4646	.3612	.1741	4,220	.3708	.3390	.2901
2,940	.4628	.3610	.1762	4,240	.3698	.3386	.2916
2,960	.4609	.3608	.1783	4,260	.3687	.3382	.2930
2,980	.4590	.3606	.1804	4,280	.3677	.3378	.2945
3,000	.45716	.36044	.18240	4,300	.3667	.3374	.2959
3,020	.4553	.3601	.1845	4,320	.3656	.3370	.2973
3,040	.4534	.3599	.1866	4,340	.3645	.3367	.2987
3,060	.4516	.3597	.1887	4,360	.3635	.3363	.3001
3,080	.4498	.3594	.1907	4,380	.3625	.3359	.3015

[1, 2] See footnotes 1 and 2, p. 55.

TABLE 11.—*Trilinear coordinates (r, g, b) of colors evoked by black. body radiant energy as a function of color temperature*—Continued

Color temperature [1]	Trilinear coordinates [2]			Color temperature [1]	Trilinear coordinates [2]		
	r	*g*	*b*		*r*	*g*	*b*
°*K.*				°*K.*			
4,400	0.3615	0.3355	0.3029	5,700	0.3118	0.3125	0.3757
4,420	.3605	.3351	.3043	5,720	.3112	.3122	.3766
4,440	.3596	.3347	.3056	5,740	.3106	.3119	.3775
4,460	.3586	.3343	.3070	5,760	.3100	.3116	.3784
4,480	.3576	.3339	.3084	5,780	.3094	.3112	.3793
4,500	.35675	.33352	.30973	5,800	.3088	.3109	.3802
4,520	.3557	.3332	.3110	5,820	.3083	.3107	.3811
4,540	.3548	.3329	.3123	5,840	.3078	.3103	.3819
4,560	.3539	.3325	.3136	5,860	.3072	.3100	.3828
4,580	.3530	.3321	.3149	5,880	.3066	.3097	.3837
4,600	.3521	.3317	.3162	5,900	.3061	.3094	.3845
4,620	.3512	.3313	.3175	5,920	.3056	.3092	.3853
4,640	.3503	.3309	.3188	5,940	.3050	.3089	.3862
4,660	.3494	.3305	.3201	5,960	.3044	.3086	.3870
4,680	.3485	.3302	.3214	5,980	.3039	.3083	.3878
4,700	.3476	.3298	.3226	6,000	.30332	.30801	.38867
4,720	.3468	.3294	.3238	6,020	.3028	.3077	.3895
4,740	.3459	.3290	.3251	6,040	.3023	.3074	.3903
4,760	.3451	.3286	.3263	6,060	.3018	.3072	.3911
4,780	.3443	.3282	.3275	6,080	.3013	.3069	.3919
4,800	.3434	.3278	.3287	6,100	.3008	.3066	.3927
4,820	.3426	.3274	.3299	6,120	.3002	.3063	.3935
4,840	.3418	.3271	.3311	6,140	.2997	.3060	.3942
4,860	.3410	.3268	.3323	6,160	.2992	.3057	.3950
4,880	.3401	.3264	.3335	6,180	.2987	.3054	.3958
4,900	.3393	.3260	.3347	6,200	.2982	.3052	.3966
4,920	.3386	.3257	.3358	6,220	.2977	.3049	.3973
4,940	.3378	.3253	.3370	6,240	.2972	.3046	.3981
4,960	.3370	.3249	.3382	6,260	.2968	.3043	.3988
4,980	.3362	.3246	.3393	6,280	.2963	.3041	.3996
5,000	.33543	.32419	.34038	6,300	.2958	.3038	.4003
5,020	.3346	.3238	.3416	6,320	.2953	.3035	.4011
5,040	.3339	.3235	.3427	6,340	.2949	.3032	.4018
5,060	.3331	.3231	.3438	6,360	.2944	.3030	.4026
5,080	.3324	.3228	.3449	6,380	.2939	.3027	.4033
5,100	.3316	.3224	.3460	6,400	.2935	.3024	.4040
5,120	.3309	.3220	.3471	6,420	.2930	.3022	.4047
5,140	.3301	.3217	.3482	6,440	.2926	.3019	.4054
5,160	.3294	.3213	.3493	6,460	.2921	.3017	.4062
5,180	.3287	.3210	.3503	6,480	.2917	.3014	.4069
5,200	.3280	.3207	.3514	6,500	.29117	.30124	.40759
5,220	.3273	.3203	.3525	6,520	.2908	.3009	.4083
5,240	.3265	.3200	.3535	6,540	.2903	.3007	.4089
5,260	.3258	.3197	.3546	6,560	.2899	.3004	.4096
5,280	.3252	.3193	.3556	6,580	.2895	.3002	.4103
5,300	.3245	.3189	.3567	6,600	.2890	.2999	.4110
5,320	.3238	.3186	.3577	6,620	.2886	.2997	.4117
5,340	.3231	.3183	.3587	6,640	.2882	.2994	.4123
5,360	.3224	.3179	.3597	6,660	.2877	.2992	.4130
5,380	.3218	.3176	.3607	6,680	.2873	.2990	.4137
5,400	.3211	.3173	.3617	6,700	.2869	.2988	.4143
5,420	.3204	.3169	.3627	6,720	.2865	.2985	.4150
5,440	.3198	.3166	.3636	6,740	.2861	.2982	.4157
5,460	.3192	.3163	.3645	6,760	.2857	.2980	.4163
5,480	.3185	.3160	.3655	6,780	.2853	.2978	.4169
5,500	.31788	.31566	.36647	6,800	.2849	.2975	.4176
5,520	.3173	.3153	.3674	6,820	.2844	.2972	.4183
5,540	.3167	.3150	.3684	6,840	.2840	.2970	.4189
5,560	.3160	.3147	.3693	6,860	.2836	.2968	.4195
5,580	.3154	.3143	.3702	6,880	.2832	.2966	.4202
5,600	.3148	.3140	.3712	6,900	.2828	.2964	.4208
5,620	.3142	.3137	.3721	6,920	.2824	.2961	.4215
5,640	.3136	.3134	.3730	6,940	.2820	.2959	.4221
5,660	.3130	.3131	.3739	6,960	.2817	.2957	.4227
5,680	.3124	.3128	.3748	6,980	.2813	.2955	.4233

[1, 2] See footnotes 1 and 2, p. 55.

TABLE 11.—*Trilinear coordinates (r, g, b) of colors evoked by black-body radiant energy as a function of color temperature*—Continued

Color temperature [1]	Trilinear coordinates [2]			Color temperature [1]	Trilinear coordinates [2]		
	r	*g*	*b*		*r*	*g*	*b*
°*K*.				°*K*.			
7,000	0.28086	0.29520	0.42394	9,500	0.2473	0.2737	0.4791
7,020	.2805	.2950	.4245	9,600	.2463	.2730	.4807
7,040	.2801	.2948	.4252	9,700	.2454	.2724	.4823
7,060	.2797	.2945	.4258	9,800	.2445	.2718	.4848
7,080	.2793	.2943	.4264	9,900	.2436	.2712	.4852
7,100	.2790	.2941	.4270	10,000	.24277	.27060	.48663
7,120	.2786	.2939	.4275	10,100	.2420	.2700	.4880
7,140	.2782	.2937	.4281	10,200	.2412	.2695	.4894
7,160	.2779	.2934	.4287	10,300	.2404	.2690	.4907
7,180	.2775	.2932	.4293	10,400	.2397	.2684	.4920
7,200	.2771	.2930	.4299	10,500	.2389	.2679	.4932
7,220	.2768	.2928	.4304	10,600	.2381	.2674	.4945
7,240	.2764	.2926	.4310	10,700	.2374	.2669	.4957
7,260	.2761	.2923	.4315	10,800	.2367	.2664	.4969
7,280	.2758	.2921	.4321	10,900	.2360	.2659	.4981
7,300	.2754	.2919	.4327	11,000	.2353	.2654	.4993
7,320	.2750	.2917	.4332	11,100	.2347	.2649	.5005
7,340	.2747	.2915	.4338	11,200	.2340	.2645	.5016
7,360	.2744	.2913	.4343	11,300	.2333	.2640	.5027
7,380	.2740	.2911	.4349	11,400	.2327	.2636	.5038
7,400	.2737	.2909	.4354	11,500	.2321	.2631	.5049
7,420	.2733	.2907	.4360	11,600	.2314	.2627	.5059
7,440	.2730	.2904	.4365	11,700	.2308	.2623	.5069
7,460	.2727	.2902	.4371	11,800	.2302	.2618	.5079
7,480	.2723	.2900	.4376	11,900	.2297	.2614	.5089
7,500	.2720	.2898	.4382	12,000	.22916	.26097	.50987
7,520	.2717	.2896	.4387	12,100	.2286	.2606	.5107
7,540	.2713	.2894	.4393	12,200	.2281	.2602	.5116
7,560	.2710	.2892	.4398	12,300	.2276	.2599	.5125
7,580	.2707	.2890	.4403	12,400	.2271	.2595	.5134
7,600	.2704	.2888	.4408	12,500	.2266	.2591	.5142
7,620	.2701	.2886	.4412	12,600	.2262	.2588	.5151
7,640	.2698	.2884	.4417	12,700	.2257	.2584	.5159
7,660	.2695	.2882	.4422	12,800	.2252	.2581	.5167
7,680	.2692	.2880	.4427	12,900	.2248	.2577	.5175
7,700	.2689	.2878	.4432	13,000	.2243	.2574	.5183
7,720	.2686	.2876	.4438	13,100	.2238	.2571	.5191
7,740	.2683	.2874	.4443	13,200	.2234	.2567	.5198
7,760	.2680	.2872	.4448	13,300	.2229	.2564	.5206
7,780	.2677	.2870	.4453	13,400	.2225	.2561	.5213
7,800	.2674	.2869	.4458	13,500	.2221	.2558	.5221
7,820	.2671	.2867	.4463	13,600	.2217	.2555	.5228
7,840	.2668	.2865	.4468	13,700	.2213	.2552	.5235
7,860	.2665	.2863	.4472	13,800	.2209	.2549	.5242
7,880	.2662	.2861	.4477	13,900	.2205	.2546	.5249
7,900	.2659	.2859	.4482	14,000	.22010	.25432	.52550
7,920	.2656	.2858	.4487	14,100	.2197	.2540	.5262
7,940	.2654	.2856	.4491	14,200	.2193	.2537	.5269
7,960	.2651	.2854	.4495	14,300	.2190	.2535	.5275
7,980	.2648	.2852	.4499	14,400	.2186	.2532	.5282
8,000	.26453	.28508	.45039	14,500	.2182	.2529	.5288
8,100	.2631	.2842	.4527	14,600	.2179	.2527	.5294
8,200	.2618	.2833	.4549	14,700	.2175	.2524	.5300
8,300	.2605	.2825	.4571	14,800	.2172	.2522	.5306
8,400	.2592	.2817	.4592	14,900	.2169	.2519	.5312
8,500	.2580	.2809	.4612	15,000	.2165	.2517	.5318
8,600	.2568	.2800	.4632	15,100	.2162	.2514	.5323
8,700	.2557	.2793	.4651	15,200	.2159	.2512	.5329
8,800	.2545	.2787	.4669	15,300	.2156	.2509	.5334
8,900	.2534	.2779	.4688	15,400	.2153	.2507	.5340
9,000	.25225	.27710	.47065	15,500	.2150	.2505	.5345
9,100	.2513	.2764	.4723	15,600	.2147	.2503	.5350
9,200	.2502	.2758	.4740	15,700	.2144	.2500	.5356
9,300	.2493	.2751	.4757	15,800	.2141	.2498	.5361
9,400	.2483	.2744	.4774	15,900	.2138	.2496	.5366

[1, 2] See footnotes 1 and 2, p. 55.

TABLE 11.—*Trilinear coordinates (r, g, b) of colors evoked by black-body radiant energy as a function of color temperature*—Continued

Color temperature [1]	Trilinear coordinates [2]			Color temperature [1]	Trilinear coordinates [2]		
	r	*g*	*b*		*r*	*g*	*b*
°*K.*				°*K.*			
16,000	0.21355	0.24938	0.53707	18,200	0.2083	0.2454	0.5463
16,100	.2133	.2492	.5376	18,300	.2081	.2452	.5466
16,200	.2130	.2490	.5381	18,400	.2079	.2451	.5470
16,300	.2127	.2488	.5385				
16,400	.2125	.2486	.5390	18,500	.2077	.2449	.5473
				18,600	.2075	.2448	.5477
16,500	.2122	.2484	.5395	18,700	.2073	.2446	.5480
16,600	.2120	.2482	.5399	18,800	.2072	.2445	.5483
16,700	.2117	.2480	.5404	18,900	.2070	.2443	.5487
16,800	.2115	.2478	.5408				
16,900	.2112	.2476	.5412	19,000	.2068	.2442	.5490
				19,100	.2066	.2441	.5493
17,000	.2110	.2474	.5417	19,200	.2064	.2439	.5496
17,100	.2107	.2472	.5121	19,300	.2062	.2438	.5499
17,200	.2105	.2471	.5425	19,400	.2061	.2436	.5503
17,300	.2103	.2469	.5429				
17,400	.2100	.2467	.5433	19,500	.2059	.2435	.5506
				19,600	.2057	.2434	.5509
17,500	.2098	.2465	.5437	19,700	.2056	.2432	.5512
17,600	.2096	.2464	.5441	19,800	.2054	.2431	.5515
17,700	.2094	.2462	.5445	19,900	.2052	.2430	.5518
17,800	.2092	.2460	.5448				
17,900	.2090	.2459	.5452	20,000	.20507	.24283	.55210
18,000	.20873	.24571	.54556	2,848	.47172	.36185	.16643
18,100	.2085	.2455	.5459				

[1,2] See footnotes 1 and 2, p. 55.

The methods of computation were the same as outlined in Section V, 4. The range of concentrations of the ingredients is about the same as was used in the mean-sun series, only the relative proportions are different. As just noted, each filter when combined with a black-body energy distribution at 2,848° K. gives by computation, an accurate color match with the respective high color temperature. A maximum deviation of but 0.00006 between the actual and desired values of *r*, *g*, and *b* was obtained. However, the energy distributions obtained with this series do not, in general, match the ideal as well as in Series I, due to the high relative energies in the violet and ultra-violet in black-body color temperatures above 4,000° K. (See chart 8 for 2,848° to sunlight, Series I, and chart 24, 2,848° to 5,000°, Series II.) The deviations in the energy distributions may be noted on the charts. The light transmissions, computed in all cases for 2,848° K., vary from 0.497 for the 3,500° K. filter to 0.0989 for the 10,000° K. filter.

Computations were thus carried out with filters designed (with 2,848° K. source) to give 3,500°, 4,000°, 4,500°, 5,000°, 5,500°, 6,000°, 6,500°, 7,000°, 8,000°, 9,000°, and 10,000° K. These computations gave values of concentration and transmission which, when plotted against color temperature, gave smooth, regular curves from which intermediate values were obtained as was done for Series I. Likewise, as before, complete independent recomputations were made on all the filters.

The filters of Series II are described and illustrated in charts 21 to 31, inclusive, and in Table 12 and Figure 15.

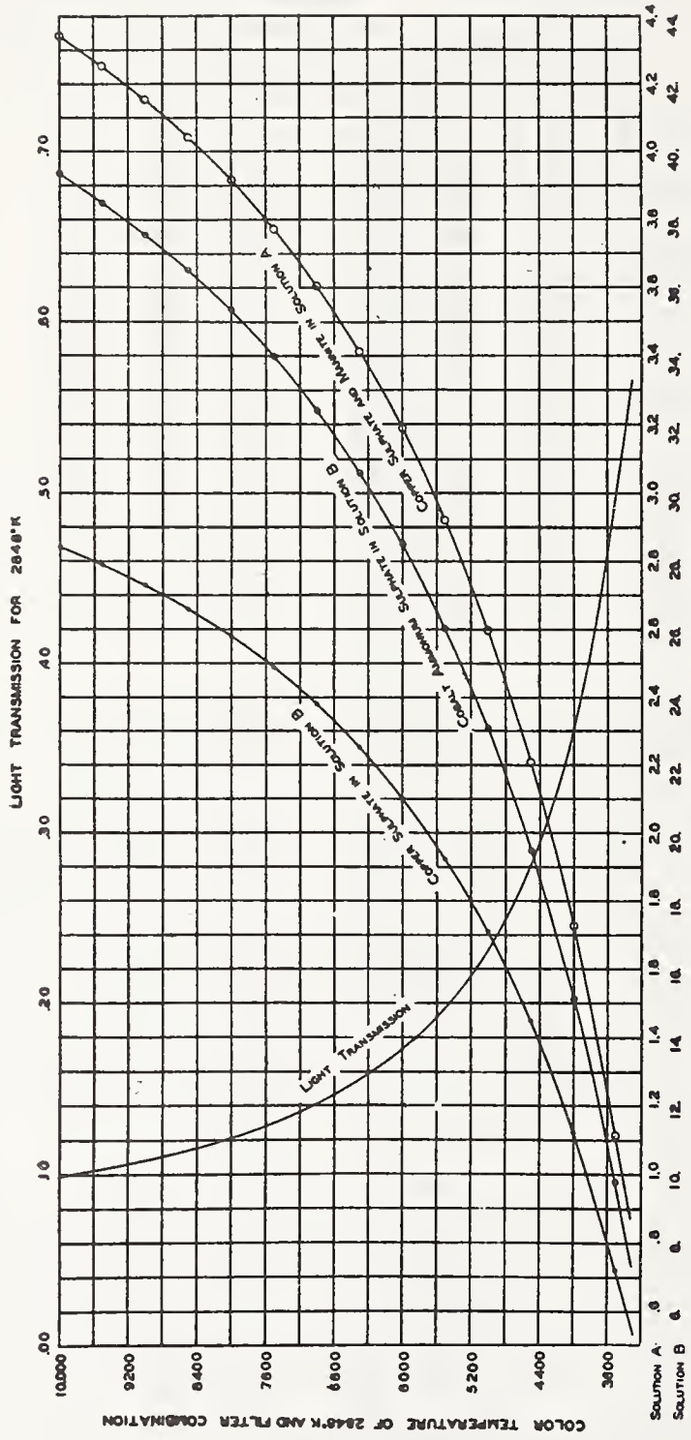


FIGURE 15.—*Graphs showing concentrations of ingredients and light transmissions for 2,848° K. in the case of the filters of Series II, viz, 2,848° K. to 3,500° K.-10,000° K.*

Values at every 100° K. are given in Table 12. In solution A the concentrations given refer both to the grams of copper sulphate and to the grams of mannite per liter of the solution.

TABLE 12.—*Concentrations and light transmissions of filters of Series II, 2,848° K. to 3,500°–10,000° K.*[1]

Color temperature (°K.) of filter with 2,848° K.	Grams of copper sulphate and grams of mannite per liter in solution A	Grams of cobalt ammonium sulphate per liter in solution B	Grams of copper sulphate per liter in solution B	Light transmission of filter for 2,848° K.	Color temperature (°K.) of filter with 2,848° K.	Grams of copper sulphate and grams of mannite per liter in solution A	Grams of cobalt ammonium sulphate per liter in solution B	Grams of copper sulphate per liter in solution B	Light transmission of filter for 2,848° K.
3,500	1.113	9.780	7.197	0.497	6,800	3.532	31.70	23.31	0.1459
3,600	1.249	10.96	8.07	.465	6,900	3.569	32.06	23.56	.1432
3,700	1.378	12.08	8.91	.435					
3,800	1.501	13.14	9.72	.408	7,000	3.605	32.400	23.800	.1407
3,900	1.616	14.16	10.49	.383	7,100	3.640	32.73	24.03	.1383
					7,200	3.675	33.06	24.25	1360
4,000	1.726	15.133	11.220	.361	7,300	3.708	33.38	24.47	.1338
4,100	1.832	16.07	11.93	.341	7,400	3.740	33.68	24.68	.1317
4,200	1.932	16.97	12.61	.324					
4,300	2.028	17.84	13.26	.309	7,500	3.772	33.98	24.88	.1297
4,400	2.120	18.67	13.88	.295	7,600	3.802	34.27	25.07	.1278
					7,700	3.832	34.56	25.26	.1260
4,500	2.208	19.470	14.490	.282	7,800	3.861	34.83	25.45	.1243
4,600	2.293	20.24	15.05	.270	7,900	3.889	35.10	25.63	.1227
4,700	2.374	20.99	15.59	.259					
4,800	2.452	21.71	16.11	.249	8,000	3.916	35.355	25.800	.1211
4,900	2.527	22.41	16.61	.240	8,100	3.943	35.60	25.97	.1196
					8,200	3.969	35.84	26.13	.1181
5,000	2.599	23.080	17.090	.231	8,300	3.994	36.08	26.29	.1167
5,100	2.668	23.71	17.55	.223	8,400	4.018	36.30	26.45	.1153
5,200	2.735	24.32	17.99	.216					
5,300	2.799	24.91	18.41	.209	8,500	4.042	36.52	26.60	.1140
5,400	2.860	25.47	18.82	.203	8,600	4.066	36.74	26.74	.1128
					8,700	4.089	36.95	26.89	.1116
5,500	2.920	26.020	19.215	.197	8,800	4.111	37.15	27.02	.1104
5,600	2.977	26.54	19.60	.192	8,900	4.133	37.35	27.16	.1093
5,700	3.033	27.05	19.97	.186					
5,800	3.086	27.55	20.32	.182	9,000	4.154	37.550	27.290	.1082
5,900	3.138	28.02	20.67	.177	9,100	4.175	37.74	27.42	.1071
					9,200	4.195	37.93	27.54	.1060
6,000	3.188	28.480	21.000	.173	9,300	4.215	38.12	27.66	.1050
6,100	3.236	28.93	21.32	.169	9,400	4.234	38.31	27.78	.1040
6,200	3.282	29.36	21.64	.165					
6,300	3.327	29.78	21.94	.161	9,500	4.253	38.49	27.89	.1031
6,400	3.370	30.19	22.24	.158	9,600	4.272	38.67	28.00	.1022
					9,700	4.290	38.84	28.11	.1013
6,500	3.412	30.580	22.520	.155	9,800	4.308	39.02	28.21	.1005
6,600	3.453	30.97	22.79	.152	9,900	4.325	39.19	28.32	.0997
6,700	3.493	31.34	23.06	1.487					
					10,000	4.342	39.353	28.420	.0989

[1] Spectrophotometric illustrations are given in charts 21 to 31; methods of computation are described in Section V, 5; details of preperation are given in Section IX.

6. DESIGN OF MISCELLANEOUS FILTERS

(a) ACETYLENE TO MEAN NOON SUNLIGHT

The acetylene flame has long been used as a light source for the sensitometry of photographic negative materials. The color of the light is whiter than that of other gas flames, or oil and spirit lamps, and approaches in quality the vacuum incandescent tungsten electric lamp. Since the flame is not homogeneous in color and intensity, it has been the custom to use a screen with an aperture close to the flame to limit the area used. While the color and intensity of the flame are slightly affected by atmospheric conditions (humidity and pressure) it is generally considered to be a reliable working standard of light quality. Its popularity is due chiefly to the simple and relatively inexpensive equipment necessary to operate it under standard conditions. The burner designed by Mees and Sheppard (39) and sold by the Eastman Kodak Co. is a standardized product and has practically displaced all other types of acetylene burners for photographic sensitometry.

The color temperature of the Eastman Kodak Co.'s burner is generally taken to be 2,360° K., although the energy distribution measured radiometrically by Coblentz gives a higher value as noted elsewhere. (Secs. III, 4 and VIII, 6.)

Figure 16 is a section of the *r-b* diagram showing the locus [15] of points representing colors of the black body from 2,000° to 4,000° K., together with the position of the acetylene burner computed from Coblentz's data ($r=0.51778$, $g=0.35719$, $b=0.12503$). (A similar illustration for color temperatures from 4,000° to 10,000° K. is given in the preceding section.)

A filter was designed to convert the color-quality of the light from the acetylene lamp to that of mean sun, with the idea that it could

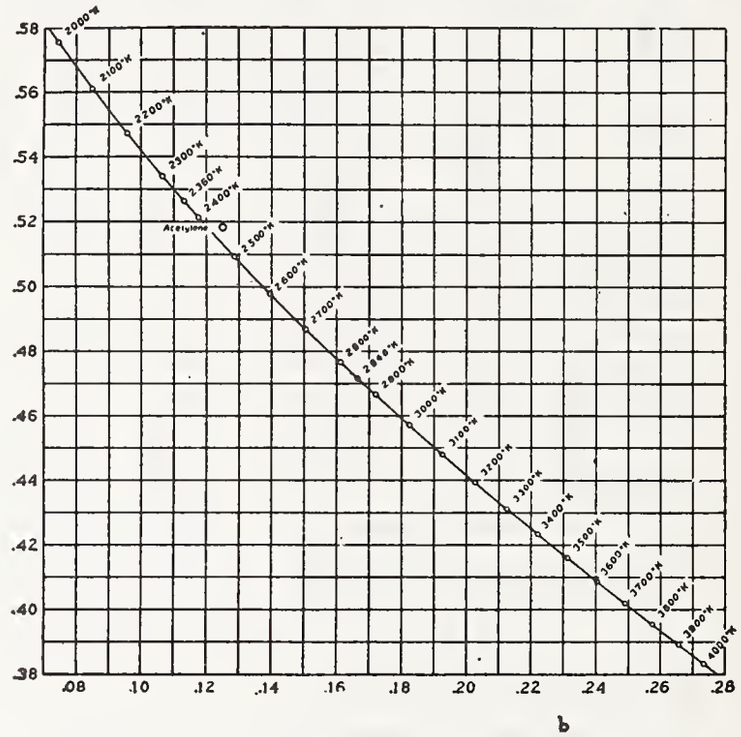


FIGURE 16.—*Section of the r–b diagram showing the black-body locus and the position of the acetylene flame (as computed from Coblentz's energy data, Table 3)*

(See fig. 14 for temperatures between 4,000° and 10,000° K.) Values of *r*, *g*, and *b* are given for every 20° K. in Table 11. All values of *r*, *g*, and *b* are computed on the basis of mean noon sunlight as the "neutral" stimulus.

be used for determining the color temperatures of incandescent lamps as outlined in Section VI, 2, (*d*), as well as for photographic sensitometry. The spectral energy distribution used in the computation of this filter was that of Coblentz as given in Table 3, reduced to 100.00 at 560mμ. The acetylene-to-mean-sun filter, by computation, gives a better energy match with the ideal than does that of the 2,360° K.-to-mean-sun filter, as may be seen by comparing charts 36 and 2.

The use of this filter in color temperature determinations is accompanied, however, by some uncertainty, because of the imperfect color

[15] The expression "colorimetric locus" will hereafter be used to designate the locus of points representing colors on the trilinear diagram. Likewise, individual points on the diagram will be referred to as "colorimetric positions."

match which may result when the acetylene flame and this filter is compared with, for example, a 2,360° K. source and its filter, and because of the uncertainty of the color temperature of the acetylene flame. This is further discussed in Section VIII, 6, where some tests with the filter are described. The use of the filter with the acetylene flame in photographic sensitometry should be entirely satisfactory.

(b) 2,848° K. TO NOON SUNLIGHT AT SUMMER AND WINTER SOLSTICES

By methods exactly analogous to those used in designing the filters of Series I and already described in detail, filters were designed for converting 2,848° K. to noon sunlight at Washington on June 21 and December 21, respectively, using Abbot's data as given in Table 1. As is the case with all the filters, an accurate color match was obtained between the ideal energy distribution and the source-and-filter combination. The filters are described in charts 34 and 35. The trilinear coordinates for these sunlight distributions are illustrated in Figure 14. They are as follows: June 21, $r=0.32184$, $g=0.32688$, $b=0.35128$; December 21, $r=0.34569$, $g=0.34027$, $b=0.31404$.

(c) 2,848° K. TO ABBOT-PRIEST SUNLIGHT

A description of the energy distribution known as Abbot-Priest sun, obtained with a 2,848° K. source and quartz-nicol rotatory-dispersion filter, is given in Section II and Figure 3. This representation of sunlight is quite satisfactory for colorimetric purposes and has been used in considerable work, both computation and direct observation, in the colorimetry section of the bureau. Because of the low light transmission of the rotatory-dispersion filter and relatively small field sizes available with it, and because the liquid filters may be used conveniently in many situations where the use of the rotatory-dispersion filter is not feasible, it was suggested that it would be desirable to duplicate the color of Abbot-Priest sunlight by means of a Davis-Gibson filter with its relatively high light transmission.

The filter so designed is described and illustrated in chart 33. Its light transmission for 2,848° K. is 0.251. A comparison of this filter and its resulting energy distribution with those for 2,848° K. to mean sun and 2,848° to 5,000° K. (charts 8 and 24) illustrates the effect of the type of energy distribution which it is sought to duplicate upon the concentrations of the ingredients, the energy distribution of the source-and-filter combination, and the light transmission of the filter.

Methods of computation leading to the best proportions of ingredients to give a precise color match and a reasonably good energy match were similar to those used for the filters of Series I and II, with the exception that, in the color computation, the form containing excitations for Abbot-Priest sunlight was used. The trilinear coordinates of this energy distribution (on the basis of mean sun as "neutral" stimulus) are illustrated in Figure 14. They are as follows: $r=0.33472$, $g=0.33052$, $b=0.33476$.

Another filter for securing the color of Abbot-Priest sunlight is also available. This is the one described in the next section and illustrated in chart 32 and Figure 17. The locus obtained with this filter by varying the color temperature of the light source passes through Abbot-Priest sunlight at 2,567° K. For this color temperature of source, the light transmission is 0.175.

518-743 O - 73 - 46

(d) 2,848° K. TO SUNLIGHT OUTSIDE EARTH'S ATMOSPHERE—VARIATION WITH COLOR TEMPERATURE OF SOURCE

As representative of daylight energy conditions—sun plus sky, or overcast sky—the energy distribution of sunlight outside the earth's atmosphere is in our opinion more nearly correct than is that of any black body between the temperatures of 6,000° and 7,000° K., although it might, perhaps, be expected that such true absorption as there may be by the water vapor of the atmosphere would tend to make the color of daylight slightly less red than that of the outside sun. While the color of daylight may, perhaps, be represented to a close approximation on the color temperature scale, its energy distribution when thus specified undoubtedly suffers in accuracy just as sunlight does when represented by 5,000° K.

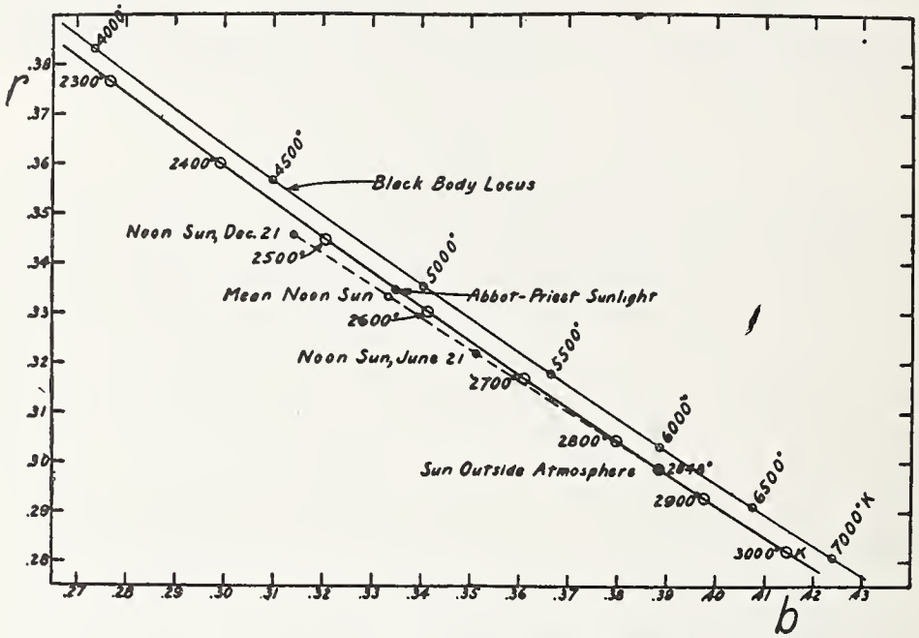


FIGURE 17.—*Colorimetric locus obtained by varying the color temperature of source, as indicated, with the filter for converting 2,848° K. to sunlight outside the earth's atmosphere*

This locus passes practically through the point representing the color of Abbot-Priest sunlight when the light source is at 2,567° K. The colors of daylight and noon sunlight may be closely matched with this filter and the variable source, as illustrated in Section VIII, 4 and Figure 32. The trilinear coordinates ("neutral" stimulus, mean sun), light transmissions, and color temperatures for this filter at every 20° K. of source are given in Table 13.

With a view to its possible use as a reproducible standard of artificial daylight, a filter was designed for converting, 2,848° K. to sunlight outside the earth's atmosphere, as represented by Abbot's data (Table 1) of 1917. The methods used were similar to those described above for the filters of Series II. A description of this filter is given in chart 32. Its light transmission for 2,848° K. is 0.180.

The trilinear coordinates for this energy distribution are: $r=0.29883$, $g=0.31261$, $b=0.38856$. Its location on the r-b diagram is illustrated in Figure 14.

Such tests as have been made with this and other filters (Sec. VIII, 4) indicate that a close color match may be obtained by means of this filter not only with daylight, but also with noon sunlight at any time

of the year, merely by varying the color temperature of the source. Accordingly, values of r, g, and b were computed for this filter with the source at every 100° of color temperature from 2,300° to 3,000° K. These computed values are given in Table 13. Intermediate values at every 20° K. were obtained by interpolation from a large scale plot. The locus of colors thus obtained with this filter by varying the color temperature of the source is shown in Figure 17.

The light transmissions of this filter for a black-body source varying over the range from 2,300° to 3,000° K. are also given in Table 13. They were computed at every 100° K. and values obtained at intermediate temperatures by reading from a large scale graph. These computations were similar to those described in Section V, 4, (*c*).

Methods of using this filter for the study of daylight and noon sunlight are outlined in Section VI, 2, (*b*).

TABLE 13.—*Data resulting from varying the color temperature of the source in the case of the filter designed to convert 2,848° K. to sun outside atmosphere—trilinear coordinates, light transmissions of filter, and color temperatures, as a function of color temperature of source.*

Colorimetric positions are illustrated in Figure 17. The filter is described in chart 32.

Color temperature (° K.) of source	Trilinear coordinates [1]			Light transmission	"Color temperature" by Davis's method [2]	Color temperature (° K.) of source	Trilinear coordinates [1]			Light transmission	"Color temperature" by Davis's method [2]
	r	g	b				r	g	b		
					°K.						°K.
2,300	0.37655	0.34684	0.27660	0.1700	4,217	2,700	0.31689	0.32199	0.36112	0.1776	5,812
2,320	.3731	.3457	.2812	.1704	4,285	2,720	.3143	.3207	.3650	.1779	5,905
2,340	.3698	.3445	.2857	.1708	4,353	2,740	.3118	.3194	.3678	.1783	6,001
2,360	.3665	.3433	.2902	.1712	4,423	2,760	.3093	.3182	.3726	.1786	6,099
2,380	.3632	.3421	.2947	.1716	4,493	2,780	.3069	.3169	.3762	.1789	6,199
2,400	.36006	.34088	.29907	.1721	4,561	2,800	.30452	.31565	.37983	.1792	6,304
2,420	.3569	.3397	.3034	.1725	4,638	2,820	.3021	.3144	.3835	.1795	6,409
2,440	.3538	.3384	.3078	.1729	4,713	2,840	.2998	.3131	.3872	.1799	6,517
2,460	.3507	.3372	.3120	.1733	4,789	2,860	.2974	.3119	.3908	.1802	6,631
2,480	.3477	.3359	.3163	.1736	4,867	2,880	.2951	.3106	.3943	.1805	6,746
2,500	.34470	.33471	.32059	.1740	4,945	2,900	.29290	.30937	.39773	.1808	6,864
2,520	.3417	.3334	.3248	.1744	5,025	2,920	.2907	.3081	.4012	.1811	6,987
2,540	.3387	.3322	.3290	.1748	5,106	2,940	.2885	.3069	.4046	.1814	7,112
2,560	.3358	.3309	.3332	.1751	5,188	2,960	.2863	.3057	.4079	.1817	7,240
2,580	.3320	.3296	.3374	.1755	5,271	2,980	.2842	.3045	.4113	.1820	7,369
2,600	.33029	.32835	.34136	.1759	5,357	3,000	.28218	.30327	.41455	.1823	7,497
2,620	.3275	.3271	.3454	.1762	5,445						
2,640	.3248	.3258	.3494	.1766	5,535	2,848	.29883	.31261	.38855	.1800	6,565
2,660	.3221	.3245	.3533	.1769	5,626						
2,680	.3195	.3232	.3573	.1773	5,718						

[1] "Neutral" stimulus: Mean sun as adopted in this investigation. Excitations: Optical Society of America extrapolated (J. Opt. Soc. Am. and Rev. Sci. Inst., **10**, p. 230; 1925). Computed values are given to five significant figures; other values, to four significant figures, were read from large-scale graph plotted from computed values.

[2] There is as yet no satisfactory, established method for computing the "color temperature"—that is, the nearest black-body color—for colors not lying precisely on the black-body locus. The values of color temperature given in Tables 13, 14, and 15 were obtained by an empiric method of computation developed by Davis. This method is, briefly, as follows: Values of rL_r, gL_g, and bL_b are computed for the energy distribution in question, where r, g, and b are the trilinear coordinates and L_r, L_g, and L_b are the luminosity coefficients (=0.45014, 0.54417, and 0.00569, respectively, as derived by Davis from least-squares adjustment of the O. S. A. excitations (extrapolated) to the visibility data of Table 10). These values are then multiplied by a factor, ω, of such magnitude that $\omega rL_r+\omega gL_g+\omega bL_b=0.33333$. These values are then compared, respectively, with similar values for the black-body energy distributions, which have been computed at closely adjacent intervals over the range from 1,600° to 20,000° K. Three "component temperatures" will thus be obtained and their arithmetical mean is taken as the color temperature. For the sake of relative values and maintaining regularity of differences in the tabulated values, these color temperatures have been computed with a greater precision than their accuracy can be guaranteed. It is impossible at this time to indicate definitely the accuracy of the values. The justification for the method is: (1) That it seems to give reasonable values, (2) that it is computationally precise, and (3) that no other wholly satisfactory method is available. The method just outlined has other features than here noted. These are reserved for discussion in a separate paper, along with a more detailed description and illustration of the above procedure, including the table of values of ωrL_r, ωgL_g, ωbL_b, summing to 0.33333, for the black-body series.

(e) 2,450° TO 3,500° K. AND TO 6,500° K.—VARIATION WITH COLOR TEMPERATURE OF SOURCE

The filters of Series II, with the source at 2,848° K. provide a means for reproducing any desired black-body color from 3,500° to 10,000° K. with high precision. It would be possible with any of these filters, by varying the color temperature of the source, to obtain a locus of colors slightly inclined to the black-body locus, and crossing it at the color temperature for which the filter was designed. Such a filter would thus enable various high color temperatures to be approximately obtained by merely varying the color temperature of the source.

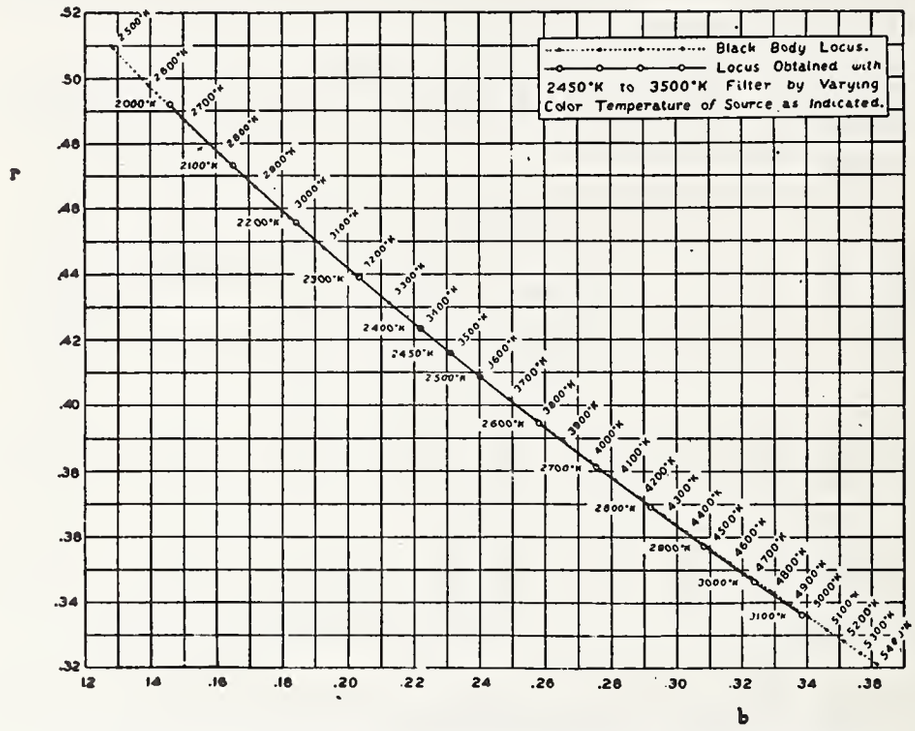


FIGURE 18.—*Colorimetric locus obtained by varying the color temperature of source, as indicated, with the filter for converting 2,450° to 3,500° K., compared to the black-body locus*

The trilinear coordinates ("neutral" stimulus, mean sun), light transmissions, and color temperatures for this filter at every 20° K. of source are given in Table 14.

This color temperature of source (2,848° K.) is eminently suitable for the filters of Series II and for the various miscellaneous filters where it is used. The higher the color temperature of source the better spectral energy match can be secured. However, the ordinary gas-filled lamps do not operate normally at much higher color temperatures than 2,848° K. so that to raise this value would ordinarily shorten the life of the lamps to an undesirable degree. It is believed the value of 2,848° K. is a very satisfactory compromise for the purpose.

When, however, it is desired to design a filter by which a range of color temperatures may be secured by varying the color temperature of the source, the point of exact color match with the black-body

locus should be chosen near the middle of the desired range, not near its upper limit as would be the case with the 2,848° K. filters, so that the deviations resulting from the crossing of the two loci may be kept as small as possible.

The practicable working range of a gas-filled lamp may be considered as from 2,000° to 3,100° K., although there are certain objections to operation near either extreme. The mid point of this range of colors is approximately at 2,450° K., as estimated on the assumption that the sensation scale along the black-body locus of colors is proportional to the spectral centroid (30, p. 1191). It was desired to have one filter by which the complete range of color temperatures conceivably called white could be covered. It was found that by means of a filter converting 2,450° to 6,500° K., a range of colors closely approxi-

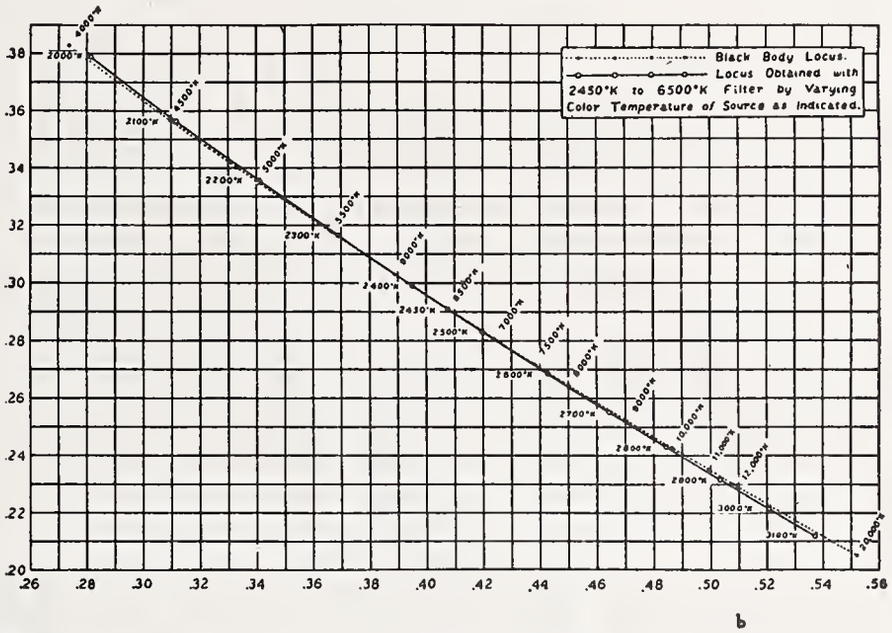


FIGURE 19.—*Colorimetric locus obtained by varying the color temperature of source, as indicated, with the filter for converting 2,450° to 6,500° K., compared to the black-body locus*

The trilinear coordinates ("neutral" stimulus, mean sun), light transmissions, and color temperatures for this filter at every 20° K. of source are given in Table 15.

mating those of a black body from about 4,000° to 19,000° K. could be obtained by varying the source from 2,000° to 3,100° K.

A filter was accordingly designed converting 2,450° to 6,500° K. with an accurate color match. Methods of design were similar to those previously described. A second filter was also designed, converting 2,450° to 3,500° K. so that by means of these two filters and the source alone a range of colors closely approximating black-body colors from 2,000° K. or below to about 19,000° K. could be obtained.

These filters are described and illustrated in charts 37 and 38. In Tables 14 and 15 are given the values of r, g, and b, the light transmissions, and the color temperatures (via Davis's method) attainable when the color temperature of the source is varied. Comparisons with the black-body locus are shown in Figures 18 and 19.

Deviations of the colors so obtained from the nearest black-body colors are in all cases considered to be less than one satron (Sec. VIII, 2), so that these filters may be considered as affording a means of practically duplicating all the black-body colors from 2,000° K. or below up to about 19,000° K.

TABLE 14.—*Data resulting from varying the color temperature of the source in the case of the filter designed to convert 2,450° to 3,500° K.—trilinear coordinates, light transmissions of filter, and color temperatures, as a function of color temperature of source*

Colorimetric positions are illustrated in Figure 18. The filter is described in chart 37

Color temperature (°K.) of source	Trilinear coordinates [1]			Light transmission	"Color temperature" by Davis's method [2]	Color temperature (°K.) of source	Trilinear coordinates [1]			Light transmission	"Color temperture" by Davis's method [2]
	r	g	b				r	g	b		
					°K.						°K.
2,000	0.49206	0.36195	0.14599	0.2850	2,649	2,600	0.39459	0.34733	0.25808	0.2982	3,816
2,020	.4883	.3619	.1498	.2855	2,685	2,620	.3919	.3465	.2616	.2986	3,859
2,040	.4845	.3619	.1536	.2861	2,721	2,640	.3892	.3457	.2651	.2989	3,903
2,060	.4807	.3618	.1575	.2866	2,757	2,660	.3865	.3449	.2686	.2993	3,947
2,080	.4770	.3617	.1614	.2871	2,793	2,680	.3839	.3441	.2720	.2996	3,992
2,100	.47354	.36148	.16519	.2876	2,829	2,700	.38138	.34325	.27538	.3000	4,038
2,120	.4697	.3612	.1690	.2881	2,865	2,720	.3790	.3424	.2787	.3003	4,082
2,140	.4662	.3609	.1729	.2886	2,901	2,740	.3764	.3416	.2821	.3007	4,126
2,160	.4626	.3606	.1767	.2891	2,939	2,760	.3739	.3407	.2854	.3010	4,171
2,180	.4592	.3603	.1805	.2895	2,976	2,780	.3715	.3398	.2887	.3013	4,216
2,200	.45576	.35991	.18433	.2900	3,012	2,800	.36907	.33895	.29198	.3016	4,262
2,220	.4523	.3595	.1882	.2905	3,051	2,820	.3666	.3381	.2953	.3019	4,309
2,240	.4489	.3590	.1920	.2909	3,088	2,840	.3643	.3372	.2985	.3023	4,356
2,260	.4456	.3586	.1958	.2914	3,127	2,860	.3619	.3363	.3018	.3026	4,404
2,280	.4423	.3581	.1996	.2919	3,165	2,880	.3596	.3354	.3050	.3029	4,453
2,300	.43904	.35757	.20339	.2923	3,202	2,900	.35735	.33451	.30814	.3032	4,502
2,320	.4358	.3570	.2071	.2927	3,242	2,920	.3551	.3336	.3113	.3035	4,551
2,340	.4327	.3564	.2109	.2932	3,281	2,940	.3529	.3327	.3144	.3038	4,600
2,360	.4295	3558	.2146	.2936	3,320	2,960	.3507	.3318	.3175	.3041	4,650
2,380	.4264	.3552	.2183	.2940	3,359	2,980	.3485	.3309	.3205	.3044	4,700
2,400	.42337	.35460	.22203	.2944	3,398	3,000	.34642	.33005	.32352	.3047	4,751
2,420	.4204	.3539	.2257	.2948	3,439	3,020	.3443	.3291	.3265	.3050	4,803
2,440	.4174	.3533	.2294	.2952	3,480	3,040	.3422	.3282	.3295	.3053	4,855
2,460	.4144	.3526	.2330	.2955	3,521	3,060	.3402	.3273	.3325	.3056	4,908
2,480	.4115	.3519	.2366	.2959	3,563	3,080	.3381	.3264	.3355	.3058	4,961
2,500	.40861	.35117	.24022	.2963	3,605	3,100	.33607	.32555	.33839	.3061	5,015
2,520	.4057	.3504	.2438	.2967	3,647						
2,540	.4029	.3497	.2474	.2971	3,688						
2,560	.4001	.3489	.2510	.2975	3,730						
2,580	.3973	.3481	.2545	.2978	3,773						

[1] "Neutral" stimulus: Mean sun as adopted in this investigation. Excitations: Optical Society of America extrapolated (J. Opt. Soc. Am. and Rev. Sci. Inst., **10**, p. 230; 1925). Computed values are given to five significant figures; other values, to four significant figures, were read from large-scale graph plotted from computed values.

[2] There is as yet no satisfactory, established method for computing the "color temperature"—that is, the nearest black-body color—for colors not lying precisely on the black-body locus. The values of color temperature given in Tables 13, 14, and 15 were obtained by an empiric method of computation developed by Davis. This method is, briefly, as follows: Values of rL_r, gL_g, and bL_b are computed for the energy distribution in question, where r, g, and b are the trilinear coordinates and L_r, L_g, and L_b are the luminosity coefficients (=0.45014, 0.54417, and 0.00569, respectively, as derived by Davis from least-squares adjustment of the O. S. A. excitations (extrapolated) to the visibility data of Table 10). These values are then multiplied by a factor, ω, of such magnitude that $\omega rL_r+\omega gL_g+\omega bL_b=0.33333$. These values are then compared, respectively, with similar values for the black-body energy distributions, which have been computed at closely adjacent intervals over the range from 1,600° to 20,000° K. Three "component temperatures" will thus be obtained and their arithmetical mean is taken as the color temperature. For the sake of relative values, and maintaining regularity of differences in the tabulated values, these color temperatures have been computed with a greater precision than their accuracy can be guaranteed. It is impossible at this time to indicate definitely the accuracy of the values. The justification for the method is: (1) That it seems to give reasonable values, (2) that it is computationally precise, and (3) that no other wholly satisfactory method is available. The method just outlined has other features than here noted. These are reserved for discussion in a separate paper, along with a more detailed description and illustration of the above procedure, including the table of values of ωrL_r, ωgL_g, ωbL_b, summing to 0.33333, for the black-body series.

TABLE 15.—*Data resulting from varying the color temperature of the source in the case of the filter designed to convert 2,450° to 6,500° K.—trilinear coordinates, light transmissions of filter, and color temperatures as a function of color temperature of source*

Colorimetric positions are illustrated in Figure 19. The filter is described in chart 38

Color temperature (°K.) of source	Trilinear coordinates [1]			Light transmission	"Color temperature" by Davis's method [2]	Color temperature (°K.) of source	Trilinear coordinates [1]			Light transmission	"Color temperature" by Davis's method [2]
	r	g	b				r	g	b		
					°K.						°K.
2,000	0.37893	0.33964	0.28138	0.0849	4,031	2,600	0.26875	0.28851	0.44274	0.0937	7,773
2,020	.3744	.3381	.2875	.0852	4,116	2,620	.2660	.2868	.4472	.0939	7,976
2,040	.3698	.3365	.2937	.0856	4,201	2,640	.2632	.2852	.4516	.0942	8,185
2,060	.3653	.3349	.2998	.0859	4,289	2,660	.2605	.2836	.4559	.0944	8,404
2,080	.3608	.3333	.3059	.0863	4,378	2,680	.2579	.2819	.4602	.0947	8,632
2,100	.35644	.33165	.31191	.0866	4,468	2,700	.25533	.28033	.46435	.0949	8,869
2,120	.3521	.3300	.3179	.0869	4,560	2,720	.2528	.2787	.4685	.0951	9,123
2,140	.3479	.3283	.3237	.0873	4,656	2,740	.2503	.2771	.4725	.0953	9,387
2,160	.3438	.3266	.3295	.0876	4,752	2,760	.2479	.2756	.4766	.0956	9,665
2,180	.3398	.3248	.3354	.0879	4,853	2,780	.2455	.2740	.4805	.0958	9,957
2,200	.33578	.32311	.34112	.0882	4,956	2,800	.24310	.27246	.48444	.0960	10,252
2,220	.3318	.3214	.3468	.0885	5,060	2,820	.2408	.2709	.4884	.0962	10,600
2,240	.3279	.3196	.3525	.0888	5,168	2,840	.2385	.2693	.4922	.0965	10,956
2,260	.3241	.3179	.3580	.0891	5,279	2,860	.2362	.2678	.4959	.0967	11,320
2,280	.3203	.3161	.3635	.0894	5,393	2,880	.2340	.2663	.4997	.0969	11,696
2,300	.31670	.31436	.36894	.0897	5,511	2,900	.23178	.26487	.50335	.0971	12,090
2,320	.3131	.3126	.3743	.0900	5,632	2,920	.2297	.2634	.5070	.0973	12,517
2,340	.3096	.3108	.3796	.0903	5,757	2,940	.2276	.2619	.5105	.0975	12,979
2,360	.3062	.3091	.3848	.0906	5,884	2,960	.2255	.2605	.5140	.0977	13,482
2,380	.3027	.3073	.3899	.0908	6,014	2,980	.2235	.2591	.5174	.0979	14,030
2,400	.29933	.30560	.39507	.0911	6,143	3,000	.22149	.25769	.52082	.0981	14,631
2,420	.2961	.3039	.4001	.0914	6,286	3,020	.2195	.2563	.5242	.0983	15,299
2,440	.2929	.3022	.4050	.0916	6,430	3,040	.2176	.2549	.5275	.0985	16,032
2,460	.2897	.3004	.4099	.0919	6,578	3,060	.2157	.2535	.5308	.0987	16,844
2,480	.2865	.2987	.4148	.0922	6,728	3,080	.2138	.2522	.5340	.0989	17,737
2,500	.28343	.29700	.41957	.0924	6,884	3,100	.21194	.25082	.53724	.0991	18,667
2,520	.2804	.2952	.4244	.0927	7,047						
2,540	.2774	.2935	.4290	.0929	7,217						
2,560	.2744	.2918	.4337	.0932	7,394						
2,580	.2716	.2902	.4383	.0934	7,580						

[1] "Neutral" stimulus: Mean sun as adopted in this investigation. Excitations: Optical Society of America extrapolated (J. Opt. Soc. Am. and Rev. Sci. Inst., **10**, p. 230; 1925). Computed values are given to five significant figures; other values, to four significant figures, were read from large scale graph plotted from computed values.

[2] There is as yet no satisfactory, established method for computing the "color temperature"—that is, the nearest black-body color—for colors not lying precisely on the black-body locus. The values of color temperature given in Tables 13, 14, and 15 were obtained by an empiric method of computation developed by Davis. This method is, briefly, as follows: Values of rL_r, gL_g, and bL_b are computed for the energy distribution in question, where r, g, and b are the trilinear coordinates and L_r, L_g, and L_b are the luminosity coefficients (=0.45014, 0.54417, and 0.00569, respectively, as derived by Davis from least-squares adjustment of the O. S. A. excitations (extrapolated) to the visibility data of Table 10). These values are then multiplied by a factor, ω, of such magnitude that $\omega rL_r+\omega gL_g+\omega bL_b=0.33333$. These values are then compared, respectively, with similar values for the black-body energy distributions, which have been computed at closely adjacent intervals over the range from 1,600° to 20,000° K. Three "component temperatures" will thus be obtained and their arithmetical mean is taken as the color temperature. For the sake of relative values and maintaining regularity of differences in the tabulated values, these color temperatures have been computed with a greater precision than their accuracy can be guaranteed. It is impossible at this time to indicate definitely the accuracy of the values. The justification for the method is: (1) That it seems to give reasonable values, (2) that it is computationally precise, and (3) that no other wholly satisfactory method is available. The method just outlined has other features than here noted. These are reserved for discussion in a separate paper, along with a more detailed description and illustration of the above procedure, including the table of values of ωrL_r, ωgL_g, ωbL_b, summing to 0.33333, for the black-body series.

VI. USE OF THE FILTERS

Certain of the uses to which these filters may be put have already been noted in connection with the discussions of the reasons for designing the various filters. Such uses are more specifically outlined in the present section and, in addition, other classes of work for

which the filters are readily adaptable are given. The use of the standard component solutions (Tables 4, 5, and 6), in checking or calibrating spectrophotometric apparatus is discussed later in Section VII, 10

1. IN PHOTOGRAPHIC SENSITOMETRY

(a) INTERNATIONAL UNIT OF INTENSITY FOR SENSITOMETRY

The need for an accurately reproducible filter for obtaining a spectral distribution of energy suitable for photographic sensitometry was, as already stated, one of the principal objectives of the present investigation. Likewise, in the fields of colorimetry and photometry, filters have been needed for converting low color temperatures to high color temperatures and for calibrating incandescent lamps for color temperature and intensity. In designing filters for use in these fields, a relatively large number have been developed giving colors and energy distributions approximating those of sunlight and daylight, and, therefore, suitable for use in photographic sensitometry.

The action of the Sixth International Congress of Photography in 1925 and the report of the American Committee on the Photographic Standard of Intensity were noted in the introduction. The subject was considered again at the Seventh International Congress of Photography meeting in London in July, 1928. At that time the American committee made another report (40) to the congress repeating its recommendation for a single high color temperature standard of light quality for use in sensitometry (as contrasted to the dual standard adopted in 1925), and urging the use of the Davis-Gibson filters, particularly the 2,360° K.-to-mean-sun filter, as a means of securing the desired sunlight distribution in the laboratory. A paper by the authors, entitled, "Artificial Sunlight for Photographic Sensitometry," was also read at the congress. (Proceedings, pp. 611–170.)

The action of the Seventh International Congress has been described elsewhere (41). It is sufficient here to quote the resolution adopted, this quotation being taken from the published Proceedings, page 173:

> This meeting of the Seventh International Congress recommends to the national committees that the photographic unit of intensity for the sensitometry of negative materials shall be defined as the intensity of a filtered source of radiation having a luminous intensity of 1 international candle, and produced by a gray body at a colour temperature of 2,360° K., together with a selectively absorbing filter made up as follows: Two solutions compounded according to the following formula, the complete filter to consist of 1 cm* layer of each solution contained in a double cell made by using three pieces of borosilicate crown glass (refractive index, $D=1.51$), each 2.5 mm thick.

Solution A

Copper sulphate ($CuSO_4.5H_2O$)	g	†3. 707
Mannite ($C_6H_8(OH)_6$)	g	†3. 707
Pyridine (C_5H_5N)	cc	**30. 0
Water (distilled) to make	cc	1, 000

Solution B

Cobalt ammonium sulphate ($CoSO_4.(NH_4)_2SO_4.6H_2O$)	g	†26. 827
Copper sulphate ($CuSO_4.5H_2O$)	g	†27. 180
Sulphuric acid (specific gravity 1.835)	cc	10. 0
Water (distilled) to make	cc	1, 000

* Tolerance in thickness shall be ±0. 05 mm.

† For practical purposes an accuracy to the second place of decimals is probably sufficient.

** EDITORS' NOTE.—It is obvious that this 30 cc refers to pure pyridine, and that the appropriate volume correction should be applied if pyridine of lesser purity is used.

It is recommended that the foregoing resolution shall come into force as a decision of this Seventh International Congress if and when ratified by the national committees represented at this congress.

(The editors' note given just above, as published in the Proceedings, p. 173, was not a part of the resolution adopted and is in error, since the pyridine used in designing the filters was of 98.4 per cent strength, as explained on p. 165 of the authors' paper published in the Proceedings.)

The filter herein specified is the Davis-Gibson filter described in chart 2 of the present paper. In accordance with the last sentence of the resolution, the matter was considered [16] by the Optical Society of America, through its council, meeting in February, 1929, and the recommendation of the Seventh International Congress adopted:

In ratifying this proposal the Optical Society of America understands that the intent of this recommendation is as follows:

1. The intention is to specify two things (*a*) the *unit* in which the intensities of light sources are to be expressed, and (*b*) the *quality* of light to be used.
2. The *unit* is to be the international candle, implying further that the intensities measured and stated will be luminous intensities as in visual photometry.
3. The *quality* of light to be used for sensitometry of negative materials is to be that which results from passing the radiation from a grey body at 2,360° K. normally through the filter described.
4. The grey body and the selectively absorbing filter, together shall be considered as the effective source in specifying the intensity (candlepower).
5. The procedure recommended for determining the intensity of the combined effective source is to multiply the intensity of the primary source (grey body) by the appropriate transmission factor of the filter which is 0.135. This factor has been computed from the spectral transmission of the filter via the relative energy distribution of 2,360° K. and the relative visibility function adopted by the sixth session of the International Commission on Illumination at Geneva, 1924.
6. This resolution does not state or imply the value of illumination to be used at the test plane during the sensitometric exposure, nor does it place any limitations on the intensity of the light source to be used.

(b) COMMENTS ON USE OF THE FILTERS

The essential details regarding the preparation of the filters are given in Section IX.

It is the custom to use low-intensity light sources (1 to 20 c. p.) for photographic sensitometry. Several lamps should be available, calibrated by one of the standardizing laboratories at 2,360° K. Lamps of small candlepower should be calibrated in terms of amperes (with potentiometer control) rather than volts. The filaments should be of such shape and dimensions as to approximate as nearly as possible a point source. The cylindrical or T type of bulb is to be preferred in gas-filled lamps, because the blackening which takes place with use in such lamps occurs mostly at the top where it does not interfere with the light taken in a horizontal direction.

These filters were designed to be used at a temperature of 25° C. where the highest accuracy is desired. However, a departure of 5° or 6° either way is of little or no importance in sensitometry. The effect of temperature is discussed in Section VII, 4. To avoid excess absorption of heat from the lamp by the filter, particularly when large lamps are used, the filter may be placed in the sensitometer with the light shutter between it and the light source. Thus the

[16] The official proceedings had not been published at the time the matter was considered by the optical society and the clarifying word "filtered" in the third line of the resolution was not present in the copy of the resolution at hand.

filter would be exposed to the radiant energy from the lamp only during the sensitometric exposure. It is also recommended that the side of the filter holding the B solution be placed facing the light source, thereby protecting the A solution, the most important one, from undue heating by the lamp.

2. IN COLORIMETRY

(a) NEUTRAL STIMULI

If the Planckian radiant energy at some temperature between 3,500° and 10,000° K. is eventually adopted as a standard neutral stimulus or white light, a filter converting 2,848° K. to this color is available from the filters of Series II, Table 12 and charts 21 to 31. If mean noon sunlight or sunlight outside the atmosphere, the latter representing average daylight, be adopted as standard white, the filters of Series I or the sun-outside-earth's-atmosphere filter (chart 32) are available. A filter (chart 33) duplicating the color of Abbot-Priest sunlight is also available.

As an aid in determining the true subjective white, the filter described in chart 38 and Table 15, has been devised, by which any color temperature from approximately 4,000° to 19,000° K. may be obtained by varying the color temperature of the source from 2,000° to 3,100° K. The use of this filter for this purpose is planned by Priest.

(b) METHODS FOR MEASURING THE COLOR OF NATURAL SUNLIGHT AND DAYLIGHT

A knowledge of the color of sunlight and daylight is of great importance both in photographic sensitometry and in colorimetry, as has already been pointed out. Such knowledge should include information regarding: (1) The variations that may take place both in daylight and sunlight with altitude of sun and with atmospheric conditions, and (2) statistical averages of such data, extensive and reliable enough to serve in the establishment of colorimetric standards of noon sunlight and daylight.

Such information as we have at present comes mostly from the extensive radiometric measurements by Abbot and his coworkers (42) on the spectral energy distribution of sunlight and the transmission of the atmosphere and on various isolated measurements of the color of sunlight and daylight made by Priest (28, 30, 43) with the rotatory-dispersion colorimetric photometer and the leucoscope. The data thus obtained give us an idea of the magnitude of the variations that take place in the color of sunlight and daylight, and serve to locate approximately the statistical averages of these variable colors.

Perhaps the chief reason for the lack of further data on these important subjects is the extreme difficulty of making reliable radiometric measurements and the lack of reliable color standards with which the colors of sunlight and daylight could be compared and measured. The rotatory dispersion colorimetric photometer has so far been the only one devised which can be made up from specification with known spectral characteristics and which is flexible enough to cover approximately the range of colors presented by natural sunlight, daylight, and skylight. This instrument has, however, not been readily available for such purposes.

Some of the filters already described afford excellent means for the precise measurement of the colors (r, g, and b coordinates) of all phases of sunlight and daylight. Inasmuch as the light transmissions of all these filters have been computed, it is also possible to measure illumination at the same time that the color measurements are made.

The following apparatus, in part or in full, in addition to the filters, is required:

1. A lamp calibrated in terms of color temperature (and intensity, if measurements of illumination are to be made). Such calibration need be only at 2,848° K. for a certain limited group of measurements as explained below. For the complete study of all phases of sunlight and daylight the lamp must be calibrated in terms of voltage or current over the color temperature range from 2,000° to 3,100° K. Such calibration may be made by means of the filters, as explained in Sections VI, 2 (d) and VIII, 5. The calibration of this working standard lamp, which should be a 500 or 1,000 watt Mazda-C lamp, should be checked from time to time to insure the desired accuracy. For this purpose, a calibrated primary standard Mazda-B lamp should be used, the checking being carried out via the filters in the same manner as the original calibration.

2. A good comparator or photometer by which the light from the sun or sky may be compared with that from the source-and-filter combination. A Martens photometer is especially good for this purpose. Both the natural and artificial sunlight or daylight should be diffused from identical pieces of magnesium carbonate or oxide or other suitable white diffusing material and the diffused light taken into the photometer. The usual precautions regarding stray light should be taken. If, in calibrating the working standard, the primary standard lamp is on that side of the photometer later taken by the natural sunlight or daylight, the method becomes a substitution method.

3. Auxiliary filters for compensating small color differences. While perfect color matches may occasionally be observed between the natural and artificial sunlight or daylight, there will, in general, be slight color differences remaining when the two fields are matched as closely as possible either by change of filter or by adjustment of the color temperature of the source. Such residual color differences may be easily eliminated by the use of Lovibond [17] glasses of low denomination. Where the color temperature of the source is varied, red (magenta) glasses only are necessary; otherwise, yellow or blue

[17] The Lovibond glasses consist, as indicated above, in a set of red, yellow, and blue glasses. The glasses of any one of the three series give colors of approximately constant hue, but of varying saturation and brilliance. Each of the red, yellow, and blue series of glasses is graded by the makers from 0.01 to 20.0. The 0.01 glass has very little absorption and its color is barely perceptible, while the 20.0 glass absorbs strongly in certain regions of the spectrum and gives a color of high saturation. The range from 0.01 to 20.0 is subdivided into a large number of steps, there being 155 glasses in each series (465 in all in a complete set of the three series) arranged and numbered in order of increasing saturation. By combining glasses of the same or different series, nearly all colors can be matched, especially if independent means is available for varying the brilliance. The glasses are prepared and sold by The Tintometer (Ltd.), Salisbury, England, and may be also obtained through importers. A descriptive advertising circular is issued by the makers.

The glasses of a given series are supposed to be additive; that is, 10.0+1.2=11.0+0.2=4.4+6.8, etc. There are certain inaccuracies in the scales, however, which have caused annoyance and error in their use and which have led to insistent demands that the bureau standardize the Lovibond glasses. The work which the bureau has done in this respect is partially described in the following publications: Gibson, Harris, and Priest, A Spectrophotometric Analysis of the Lovibond Glasses, B. S. Sci. Paper No 547; 1927. Judd and Walker, A Study of 129 Lovibond Red Glasses with Respect to the Reliability of their Nominal Grades, Oil and Fat Industries, 5, pp. 16–26; January, 1928. Priest, Judd, Gibson, and Walker, Calibration of Sixty-five 35-Yellow Glasses, B. S. Jour. Research, 1 (RP58), pp. 793–819; April, 1929.

glasses may also be required. This is indicated more specifically below. The colorimetric loci of the fractional Lovibond glasses are illustrated in Figure 20, wherein are shown also the black-body locus and the locus of the 2,848° K.-to-sun-outside-atmosphere filter obtained by varying the color temperature of the source. The coordinates r and g are used in this figure, as least perceptible color differences in any direction are more nearly proportional to distances

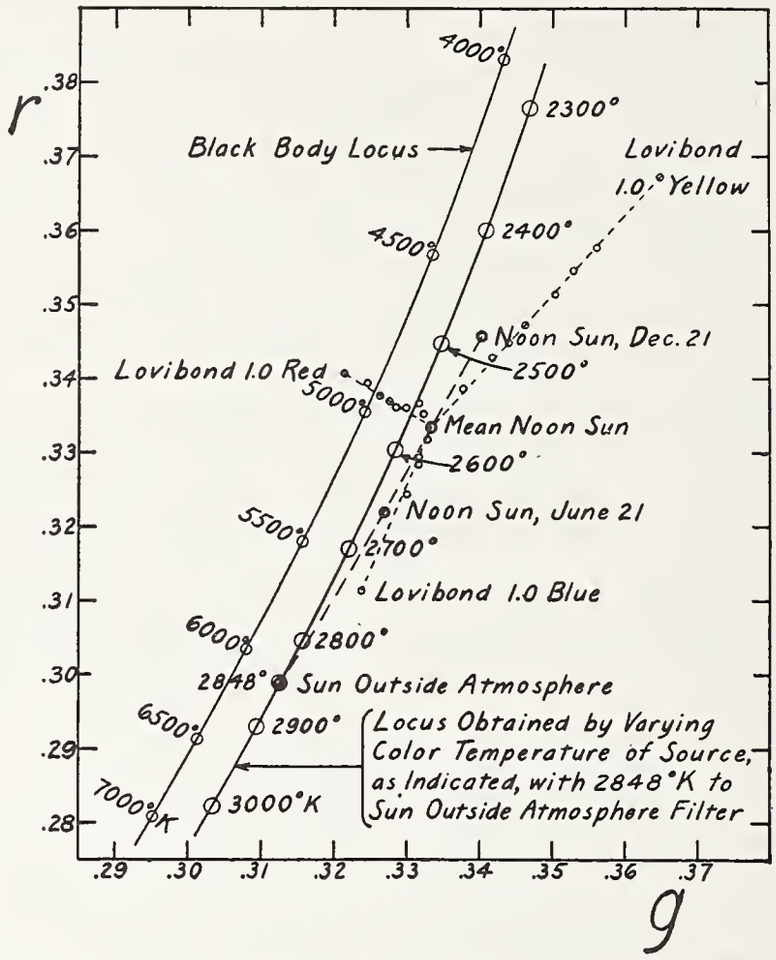


FIGURE 20.—*Section of the r–g diagram showing colorimetric positions of Lovibond glasses of low denomination which may be used with certain of the filters in determining the color of sunlight and daylight* (*See Sec. VI, 2, (b) for details*)

The Lovibond glasses shown (B. S. 9940) have the following maker's numerals: Red 1.0, 0.70, 0.54, 0.40, 0.32, 0.26, 0.19, 0.09; Yellow 1.0, 0.70, 0.60, 0.50, 0.40, 0.30, 0.20, 0.10; Blue 1.0, 0.50, 0.30, 0.20, 0.10.

on this scale than on the r-b scale. A comparison of this figure with Figure 17 will illustrate the differences in plotting in the two cases.

The effect of temperature change on the filters should not be ignored. This is discussed in Sections VII, 4 and VIII, 2, (*d*). If a 500 or 1,000 watt lamp is operated in close proximity to the filter, the temperature of the filter will rapidly increase. This may, in general, be prevented by turning off the lamp or interposing a shutter except during the actual observations and by placing a water cell between

the lamp and filter. The color of the latter may be ignored if made of borosilicate crown glass as specified for use in the filter, as illustrated in Section VIII, 2, (*d*).

The following combinations of source and filter may be used in the study of the color of sunlight and daylight.

(1) *Source Calibrated at 2,848° K. only.*—The following filters are available:

(*a*) 2,848° K. to sun outside earth's atmosphere, chart 32; (*b*) 2,848° K. to mean noon sun at Washington, chart 8; (*c*) 2,848° K. to noon sun at Washington on June 21, chart 34; (*d*) 2,848° K. to noon sun at Washington on December 21, chart 35; and (*e*) 2,848° K. to Abbot-Priest sun, chart 33—all as based on Abbot's data of 1917. On the basis of direct examination of daylight and noon sunlight made with filters (*a*) and (*b*) in September, 1928, and with filters (*a*) and (*d*) in December, 1928, the locus of colors represented by the first four filters is slightly greenish relative to the natural sunlight and daylight. Exact color matches may be obtained, however, by inserting Lovibond glasses of low denomination in combination with the source and filter—red alone, or red with blue, or red with yellow, depending on atmospheric conditions. This method was used in securing the data illustrated in Figure 32. The Lovibond glasses required varied from about 0.10 to 1.0 in each of the three colors. They were of known spectral transmission and thus of known *r*, *g*, and *b*, so that the colorimetric values may be derived from the measurements.

The same sort of measurements could be made, using filters of Series II; for example, 2,848° to 5,000°, 5,500°, 6,000°, or 6,500° K. On the basis of the measurements referred to above, the black-body locus is slightly purplish relative to the color of natural daylight and noon sunlight so that combinations of yellow and blue glasses should give the desired color match.

(2) *Source calibrated over range from 2,300° to 3,000° K.*—The most objectionable feature of the previous method is the necessity for the use of so many Lovibond glasses. If, however, the color temperature of the source may be varied, exact matches may be made with the red glasses only. If the color temperature of the source is varied, in the case of the filter for converting 2,848° K. to sun outside atmosphere (chart 32) the resulting locus of colors falls very close to that of the natural daylight and noon sunlight, crossing apparently at a slight angle, so that when perfect matches are not obtainable by varying only the source they may be obtained by inserting a red glass, usually of very low denomination, on one side or the other of the photometer. The values of *r*, *g*, *b*, the approximate color temperatures, and the light transmissions of this filter with varying source are all given in Table 13.

On the basis of the observations made in September and December, 1928 (fig. 32), this method is better than that of (1) above, or (3) below for the study of daylight and noon sunlight, although method (1) was used in securing the data illustrated. The only Lovibond glasses required are red glasses having maker's numerals of about 0.1, 0.2, 0.4, unless glasses less than 0.1 are desired for distinguishing extremely minute color differences. It is an interesting coincidence that the locus of colors obtainable with this filter passes almost exactly through the point representing the color of Abbot-Priest sun, so that, as previously noted, the latter color may be secured

not only with a 2,848° K. source and the filter described in chart 33, but also with a 2,567° K. source and the filter described in chart 32.

(3) *Source calibrated over range from 2,000° (or below) to 3,100° K.*—By means of the two filters designed as outlined in Section V, 6 (*e*)—viz, 2,450° to 3,500° K. and 2,450° to 6,500° K.—and by practically the same method as outlined under (2) just above, the colors of sunlight at practically any altitude, of daylight, and of blue sky may be determined. In addition to the Lovibond red glasses, the use of one or two blue glasses would eliminate the necessity of operating the lamp at the highest temperatures in measuring the color of blue sky. Data on these filters are given in charts 37 and 38 and in Tables 14 and 15.

(c) STANDARDS FOR GRADING DAYLIGHT GLASSES AND ARTIFICIAL WHITE-LIGHT ILLUMINANTS

That there is need for a ready means of checking the color of so-called "daylight" illuminants has been pointed out in a paper already referred to (17). The manufacturer of such lighting units or of daylight glass should also have available a means of calibrating or grading the material which he is selling. Such a means is available in the filters of the present paper.

He can take his choice of mean noon sunlight (the filters of Series I, charts 1 to 20), average daylight or overcast sky (chart 32), Abbot-Priest sunlight (chart 33), or any one or more of the color temperatures from 3,500° K. (yellowish) to 10,000° K. (bluish) (the filters of Series II, charts 21 to 31) including 6,500° K., which is close to the color of average daylight. Furthermore, filter No. 38, which by varying the color temperature of the source will give him approximately any color temperature from 4,000° to 19,000° K., will enable him to grade any piece of daylight glass or any artificial daylight illuminant at will.

For precise work, a good comparator, such as the Martens photometer, should be used. Simple visual comparison without any instrument will, however, immediately detect any serious deviations from the standard which it is desired to maintain and which is given by any one of the filters that may be selected for that purpose.

(d) CALIBRATION OF INCANDESCENT ELECTRIC LIGHTS IN TERMS OF COLOR TEMPERATURE

By designing the filters of Series I to give an accurate color match with mean sun, and, therefore, with each other, they may be used to calibrate an incandescent electric light in terms of color temperature over the range of the filters, and also, as indicated in the next section, to compare the intensities of two lights of different color temperatures under conditions of perfect color match, thus eliminating some of the uncertainties and difficulties of heterochromatic photometry.

The calibration of a lamp in terms of color temperature is relatively simple provided one has available a good photometer and a lamp already standardized for color temperature over the range to be used. The standardization of a lamp for color temperature has, however, been difficult or impossible for any but a few specially equipped standardizing laboratories. The only instrument for such purposes which is self-contained for its calibration—that is, not dependent for its calibration upon some other calibrated apparatus, other than

upon a standardized lamp—is Priest's rotatory dispersion colorimetric photometer (30). Blue glass or gelatine wedge photometers, even if available, are of no fundamental assistance, as they must be calibrated. "Black bodies" are difficult to operate and can not readily be used at the ordinary operating temperatures of the gas-filled tungsten lamp. The usual procedure, if one has wished to secure a lamp standardized for color temperature, has been to obtain such a lamp from the National Bureau of Standards or from some other standardizing laboratory.

The present series of filters will enable one to standardize his own lamp from 2,300° K. up to as high a temperature as is advisable and furthermore to check such calibration at any time desired. Aside from the filters the sole requisites for such standardization are a source whose color temperature is known at any single temperature within the range of the filters, together with a good photometer. It had been hoped that the acetylene flame might be used as a reproducible source of accurately known color temperature and thus free the worker from any dependence upon the standardizing laboratory. Certain questions have arisen, however, in connection with the acetylene flame (discussed in Section VIII, 6) and, unless approximate values only are desired, it is doubtless preferable to secure a lamp from a standardizing laboratory calibrated at some one temperature and use this as the basic fundamental standard. Color temperatures of 2,360° and 2,848° K. are commonly used reference points, 2,360° K. being especially favorable as the vacuum tungsten lamp has great constancy both in color and intensity at this color temperature. The following procedure is recommended for the standardization of a lamp over a range of temperatures by means of the filters of Series I:

1. All lamps to be used should be seasoned; that is, operated at normal voltage until they give a constant current at this constant voltage.

2. Upon receipt of the fundamental (primary) standard lamp (standardized at some one color temperature by a standardizing laboratory) secondary working standards should be calibrated in terms of it (without the use of any filters) and thereafter used for the basic color temperature. The primary lamp should never be used except to check the secondary standards.

3. Mean-sun filters should be prepared (*a*) for the basic temperature of the primary and secondary standard lamps and (*b*) for as many other color temperatures as are necessary in order to cover the desired range at sufficiently close intervals.

4. It is best to use a substitution method. The working standard and its filter should illuminate one side of the photometric field and a comparison lamp and its filter should illuminate the other. The voltage or current on the comparison lamp should then be varied until an exact color match is obtained in the two parts of the photometric field. The brightnesses in the two halves of the field should be continually matched during that procedure.

5. The particular color temperature chosen at which to operate the comparison lamp is perhaps unimportant. If it is a vacuum tungsten lamp, 2,360° K. is a suitable temperature. If it is a gas-filled lamp, it should probably be operated at some color temperature near the middle of the range being covered.

6. The working standard and filter should then be replaced by the lamp being standardized and its voltage or current determined with the various filters prepared for the desired color temperatures.

7. The number of readings taken for the single determinations and the amount of repetition and checking necessary will depend on the observer's skill, his sensibility to color differences, the consistency of the results, and the accuracy desired. A graph of the relation between color temperature and voltage or current of the lamp should have a perfectly regular (smooth) curvature.

8. This lamp being once standardized may be treated as a working standard and other lamps calibrated in terms of it, or several lamps may be standardized at the same time with the same filters.

The question of precision is touched upon in the following section, and a test of the reliability of the filters for color temperature measurements is described in Section VIII, 5.

3. IN PHOTOMETRY—COMPARISON OF INTENSITIES OF INCANDESCENT ELECTRIC LIGHTS OF DIFFERENT COLOR TEMPERATURES

The difficulties and uncertainties of heterochromatic photometry are well known to those engaged in the testing and standardizing of lamps. Even when the uncertainties of attempting to make direct photometric matches of two differently colored lights are eliminated by flicker photometry under standard conditions there still remains the important difference that may result when two observers, both of normal color sense but having different visibility functions, make respective determinations. A difference of several per cent may be thus obtained when the candlepower of a gas-filled tungsten lamp is measured relative to that of a vacuum tungsten lamp.

The use of the filters would not be convenient where the intensities of lamps of unknown color temperature are desired. But a calibration of intensity for any given lamp over a range of color temperatures may be readily obtained, at the same time as the color temperature calibration described in the previous section if desired, and always with perfect color match in the two parts of the photometric field. Also, of course, any number of lamps, at one or several known color temperatures, may be measured relative to a standard lamp operating at some other temperature, all color temperatures being 2,300° K. or above. The procedure would be very similar to that outlined for the color temperature standardization. The substitution method is even more important in this case. The light transmissions of the filters for their respective sources are given in Table 8 and in the charts.

It may be noted that in the methods of use of the filters as outlined above, both for the determination of color temperature and intensity, various uncertainties entering into the determination of the color and transmission of a filter, and discussed in Sections VII and VIII—viz, possible errors in the spectrophotometric data, possible lack of reproducibility of the chemicals and glass plates, temperature effects, etc.—have a tendency to cancel out, their importance decreasing as the differences in the color temperatures of the sources decrease. These filters should, therefore, have, along with their general applicability for all color temperatures above 2,300° K., a reliability equal to that obtained with the various special bluish or yellowish filters which have been used to bridge certain steps in the color

temperature scale of incandescent sources, even when these latter are carefully calibrated.

The precision of measurement is probably about the same whether filters are used or not, for it has been shown that (30, p. 1191) the precision of determination of the spectral centroid is independent of the color temperature, and also that (44) when the color temperature of a light is changed (for example, by altering the voltage on a lamp), the variation in the spectral centroid of the light transmitted through a white-light filter is slightly greater than the variation in the spectral centroid of the light alone. A real cause for decreased precision might, however, be found in the lowered brightness resulting from the use of filters on both sides of the photometer. With most incandescent lights, however, intensities are sufficiently great to make this objection of little importance.

An extension of the present series of filters to make them applicable to sources at color temperatures less than 2,300° K. is impracticable. The lower the color temperature of the source, the less must be the transmission of the filter required to produce artificial sunlight, which has been our objective in designing these filters. This low transmission, combined with the generally low intensities of sources of low color temperatures, would decrease the brightness undesirably. If there were a demand for filters applicable for use below 2,300° K., a new series could be designed, choosing, not sunlight, but light from the Planckian radiator at some relatively low temperature (say 2,848° K.) as the light to be delivered by the source-and-filter combination. With this objective, practicable filters could be computed for use in combination with sources at temperatures from below 2,000° to above 2,300° K.

(Approximate values of color temperature from 2,000° to 2,300° K. may be obtained by use of the 2,450° to 3,500° K. filter. For example, from Table 14 it may be noted that color temperatures of 2,649°, 2,829°, and 3,012° K. result when this filter is used with source at 2,000°, 2,100°, and 2,200° K., respectively. A lamp may be calibrated for 2,649°, 2,829°, and 3,012° K. by means of the filters of Series I as illustrated above. To calibrate a lamp for color temperatures of 2,000°, 2,100°, 2,200° K., and other values between 2,000° and 2,300° K., it is merely necessary to vary the voltage on this lamp until its color viewed through the 2,450°-to-3,500° K. filter is matched with that of a bare gas-filled tungsten lamp calibrated at the appropriate color temperatures by means of the filters of Series I. The values thus obtained may be somewhat uncertain because of the lack of exact match between the black-body color and that of the source-and-filter combination at these low color temperatures.) (Fig. 18.)

4. IN ASTRONOMY—DETERMINING THE COLOR TEMPERATURES OF THE STARS

The two filters described in Sections V, 6 (*e*), together with a calibrated light source, whereby color temperatures from 2,000° K. or below up to 19,000° K. may be obtained, should furnish a simple means of accurately determining the color temperatures of the stars. Aside from the usual astronomical equipment, the essential apparatus would consist of a suitable comparator, the two filters, and a lamp calibrated over a color temperature range from 2,000° to 3,100° K. with its accompanying voltmeter or ammeter.

518-743 O - 73 - 47

VII. SPECTROPHOTOMETRIC DATA

1. APPARATUS AND METHODS

(a) VISUAL METHOD

Except for some of the preliminary measurements not reported quantitatively in the paper, all spectrophotometric measurements were made in the colorimetry section of the bureau. The larger part of the measurements were made visually with the Koenig-Martens spectrophotomoter and auxiliary equipment. A description of this apparatus and method of use is given in a recent publication (45). In the present measurements the instrument has been used in two ways:

1. For the measurements leading to the standardization of the absorbancy and transmittancy values for the standard component solutions (medium concentrations) and the density and transmission values for the double cell filled with distilled water, both the homogeneous (Hg and He) light sources and the heterogeneous (incandescent Mazda C) light source were used, the former, in general, at the eight wave lengths noted below, with 0.5 mm slits, and the latter over the range from 420mμ or 450mμ to 650mμ or 710mμ, usually at every 10mμ and with 0.2mm slits (except at the ends of this range where wider slits are usually necessary).

2. For the studies of permanence, reproducibility, temperature effect, Beer's law, variation of pyridine content, and various miscellaneous measurements these eight wave lengths—viz, Hg 404.7, Hg 435.8, Hg 491.6, He 501.6, Hg 546.1, Hg 578.0, He 587.6, and He 667.8—in whole or in part, plus various intermediate wave lengths with the incandescent light if necessary, have furnished sufficient data for the information desired. The obvious advantage in their use is that differences in values obtained in a study of the various effects can not be ascribed to wave-length or slit-width errors. Furthermore, they cover the most important part of the wave-length range used in the present investigation. Various details concerning the measurements are noted later in the sections devoted to the respective subjects.

(b) AUXILIARY METHODS

Supplementing the visual method in the determination of the spectrophotometric data on the standard component solutions, thermoelectric, photoelectric, and photographic methods were used. The thermoelectric and photoelectric methods have been described elsewhere (46). The former was used over the range from 610mμ or 650mμ to 750mμ or 780mμ, at every 10mμ or 20mμ, depending upon the type of transmittancy, thus overlapping the visual method at several wave lengths. The photoelectric method was used over the range from 390mμ to 510mμ or 550mμ, likewise at every 10mμ or 20mμ, thus overlapping the visual method for an extensive range in the blue and green.

The photographic (Hilger sector photometer) method was used essentially as previously described (47, 48), except that for part of the measurements a ribbon-filament Mazda C lamp was used as source instead of the underwater spark. The primary purpose in the use of this method was to extend the data from the visible to 350mμ. It was used not only for the standard component solutions, but also in the study of Beer's law.

(c) MISCELLANEOUS

The thicknesses of cell used in the spectrophotometric determinations were 10, 20, 40, and 100 mm, these being accurate to one-tenth of 1 per cent or better. The particular thickness used in any case depended on the solution and the purpose of the measurements. Sometimes but one thickness, sometimes three or four thicknesses were used during a set of measurements on any given solution. The particular ones used will not, in general, be further stated. Regardless of the thickness used, values were always computed to 10 mm, inasmuch as this is the standard thickness used in the filters described herein.

The temperature of the solutions was in all cases kept at $25° \pm 1°$ C., unless otherwise noted. The apparatus for all four methods of measurement includes temperature control for solutions. This has been previously described (45, 47).

The concentrations given throughout this section are accurate to 0.001 g, and the volumes to 0.1 cc (except in the case of 1,000 cc or 1 liter, where 0.1 cc is not significant), whether indicated by the proper number of significant figures or not.

2. STANDARD FILTER COMPONENTS

In Tables 4, 5, 6, and 7 (Sec. V, 2) are given values of the density of the double cell with water, and of the absorbancy of the standard solutions A′, B_1', and B_2' as adopted for the present investigation. A precise description of the cell and of the solutions is given in Section IV, 1.

An outline of the experimental procedure, by means of which these data were obtained, follows. The final use of the data was, of course, kept in mind throughout the measurements. While it was convenient and desirable in view of the various computations, to consider the fundamental spectrophotometric data in terms of absorbancy (the slide rule used in the visual measurements on the Koenig-Martens spectrophotometer contained a $-\log_{10}$ scale, so that values of absorbancy were obtained directly from the readings of the spectrophotometer and not via a previous transmittancy computation), it is ultimately the transmittancy values, entering into the final transmission of the cell and its contents, comprising any given filter, which is of importance; for it is the product of transmission of the filter by the energy of the light source that gives the energy distribution of the source-and-filter combination. A difference or error in absorbancy of 0.00434 means a difference or error in transmittancy amounting to 1 per cent of the transmittancy, regardless of the values of absorbancy and transmittancy. Consequently it was never attempted to reach a constant percentage accuracy in the absorbancy data; that is, if ΔA represents the uncertainty in the values of A, it was never attempted to make $\frac{\Delta A}{A}$ constant regardless of A. It was attempted, rather, to maintain ΔA constant, less than 0.00434 so far as possible, so that $\frac{\Delta T}{T}$ should be constant and not greater than 1 per cent. A discussion of the reliability of the various data is given later.

In view of the uncertainty in the absorbancy caused by filtering the A solution through filter paper, it was, as elsewhere noted, decided

to base all standard adopted data on unfiltered solutions. This point is of little importance in the case of the B solution, but is not negligible with the A solution. All the spectrophotometric data given in this paper, unless otherwise specifically noted, were obtained with unfiltered solutions. In using the solutions during the measurements care was taken not to shake or tip the bottles unduly, and the solutions were transferred from bottle to cell usually by means of a pipette. The reason for this is that all chemicals contain a minute quantity of trash or insoluble material which settles to the bottom upon standing.

(a) TWO-COMPARTMENT GLASS CELL WITH DISTILLED WATER

Measurements were made by the visual method from 460mμ to 720mμ in addition to the Hg wave lengths 404.7mμ, 435.8mμ, 491.6mμ, 546.1mμ, and 578.0mμ. Photoelectric measurements were made from 400mμ to 580mμ and thermoelectric measurements from 640mμ to 780mμ. In addition, measurements were made by R. Stair, of the radiometry section of the bureau, by the thermoelectric method, at the Hg wave lengths 436mμ, 405mμ, 365mμ, 334mμ, 313mμ, 305mμ, and 280mμ.

The adopted values of density and transmission are given in Table 7 (see also fig. 10), those below 400mμ being based primarily on Mr. Stair's data. The uncertainty in the absolute values of transmission might be ± 1 per cent, considering the spectrophotometric data alone, although the relative accuracy is better than this. However, the value of transmission computed from the refractive indices of the glass and water, assuming no absorption, is equal to 0.904, and a value of 0.906 was reported by the optical instrument section for the "white light" transmission. The values of transmission in Table 7 are, therefore, considered accurate to ± 0.5 per cent or better, as the spectral transmission over the range of highest luminosity varies from 0.900 to 0.904.

The water absorption band at 750mμ is plainly shown. The slight absorption shown by the remaining data in the visible, and rapidly increasing in the ultra-violet, is ascribed mostly to the glass.

(b) SOLUTION A'

An outline of the experimental work leading to the adoption of this combination of materials for the principal (blue) solution has been given in Section IV, 1. Having decided on the most suitable amounts of copper sulphate and pyridine, viz, 3.000 g and 30.0 cc, respectively, but without mannite as yet, spectrophotometric measurements were made by the visual, thermoelectric, and photo-electric methods, covering the range from 390mμ to 750mμ. Values of absorbancy for 10 mm thickness at every 10mμ from 390mμ to 750mμ were adopted from these measurements (May, 1926). They refer to a single solution, No. 61–11, the measurements being made within a few days after preparation. After the hot summer weather had shown the copper-sulphate-pyridine solutions to be rather unstable at temperatures between 35° and 40° C. and mannite had been added as an essential constituent of these blue solutions, measurements of absorbancy were repeated, similar to but more extensive than those of May, 1926. New values of absorbancy were thereupon adopted

(October, 1926). They refer to solution No. 62–2, the data being taken from 3 to 32 days after preparation of solution. Certain small but fairly definite discrepancies which were more or less a function of time became apparent during these measurements. These were at first considered as a turbidity effect, but later work has proved them mostly a result of the very slow change in absorbancy which takes place with time. (This characteristic of the A solution is discussed in Section VII, 3.)

Another solution was thereupon prepared, No. 62–15. The eight reference wave lengths previously noted—viz, Hg 404.7mμ, Hg 435.8mμ, Hg 491.6mμ, He 501.6mμ, Hg 546.1mμ, Hg 578.0mμ, He 587.6mμ, and He 667.8mμ—were the only ones used in the study of this new solution. Measurements were made on the day of preparation and on the first, second, sixth, eighth, twenty-third, and twenty-fourth days after preparation. The changes that took place from the sixth to the twenty-fourth day were all less than 1 per cent of the transmittancy.

Based principally on the average of these measurements from the sixth to the twenty-fourth on solution No. 62–15, and guided also by miscellaneous measurements at these wave lengths on other A′ solutions, values of absorbancy were adopted (Nov. 19, 1926) which are the values listed at the bottom of Table 4. A summary of all the experimental data obtained with the eight Hg and He wave lengths on the various A′ solutions is given in Table 18.

The data of October, 1926, were then adjusted so that the data of November 19, 1926, would lie precisely upon the curve drawn through the adjusted values. These adjusted values of absorbancy (adopted Nov. 30, 1926) are the ones given in Table 4 from 400mμ to 750mμ (see also fig. 10). The finally adopted spectral absorbancy values of November 30, 1926 (with mannite) differed from the original data of May, 1926 (without mannite), in no case by more than 0.005. There is no reason to think that mannite influences the transmittancy of the copper-sulphate-pyridine solutions except to make them much more stable.

During February, 1927, measurements were made photographically (solution No. 65–5, seven days after preparation), overlapping the adopted data of November, 1926, from 400mμ to 510mμ and extending into the ultra-violet beyond 350mμ. The agreement with the adopted values above 400mμ was satisfactory. The values given in Table 4 from 350mμ to 390mμ were obtained from these measurements on solution No. 65–5, taking a graphic average of the data for 10, 20, and 40 mm thickness, all reduced to 10 mm as usual. These data were closely checked by later measurements on solution No. 65–8–C, 20 days after preparation, made during the study of Beer's law from 350mμ to 400mμ.

(c) SOLUTION B_1'

As previously stated, 20.000 g of copper sulphate per liter of acidified aqueous solution was chosen as standard. A filtered solution was prepared, No. 62–1, and spectrophotometric measurements made by visual, thermoelectric, and photoelectric methods during the first month after preparation. Values of absorbancy from 390mμ to 750mμ were thereupon adopted (September, 1926). These values were checked closely by the measurements made in the test of Beer's

law (solutions Nos. 62–7–A to E, filtered, October, 1926), but were revised somewhat in the red on the basis of later measurements.

After the decision had been made not to use filtered solutions, an unfiltered solution, No. 62–5, was prepared. A single set of measurements at the Hg wave lengths (see Table 19 for summary of measurements made with Hg and He wave lengths on the B'_1 solutions), made on the second day after preparation and using 100 mm cells, showed a very distinct decrease in transmittancy compared with the values previously obtained with filtered solutions, although this amounted to less than 1 per cent when reduced to a 10-mm basis. This discrepancy was considered a turbidity effect and it was decided to study the change with time in a manner similar to that described for the A′ solution, No. 62–15.

Accordingly, a new solution was prepared, No. 62–18, unfiltered, and measurements made on the day of preparation and on the first, fourth, twelfth, and fourteenth days after preparation, using the eight Hg and He wave lengths. It was found (Table 19) that the measurements made during the first four days after preparation checked those made on No. 62–5, unfiltered, made the second day after preparation, and that the measurements made on the twelfth and fourteenth days checked closely with the measurements made on the filtered solutions Nos. 62–1 and 62–7–C. Apparently, then, the insoluble matter causing the turbidity settles to the bottom of the bottle after the first few days and is removed entirely by filtering. Measurements on solution B′ (below) show that no further change of any importance takes place. With B_1 solutions, therefore, there is no practical difference between filtered and unfiltered solutions, provided the latter are not used for a week or so after preparation and care is taken not to disturb the insoluble material at the bottom of the bottle.

Based partially on the measurements with solution No. 62–18, particularly the data for 12 and 14 days after preparation, and also on the other measurements noted, absorbancy data for the Hg and He wave lengths were adopted (Nov. 13, 1926) and are given at the bottom of Table 5. Particular attention was paid to the He 667.8 measurements, seven different determinations being finally averaged to give the value adopted.

The previous data of September, 1926, were now adjusted to fit these data of November 13, 1926, in a manner similar to that done with solution A′. The only important change was in the red. The change required to make the previous data consistent with the He 667.8 values varied from less than 1 per cent of the transmittancy below 630mμ to about 3 per cent at 750mμ. The finally adopted data (Nov. 15, 1926) from 400mμ to 750mμ are given in Table 5 (see also fig. 10).

Photographic measurements in the ultra-violet to approximately 320mμ were made on 40-mm solutions (No. 65–3, unfiltered) about two weeks after preparation. The absorbancy was too low to obtain any data above 370mμ by this method, but it was not difficult to connect the visible and photographic data by a reasonable curve. The values of absorbancy are so small in this region that they might be considerably in error without affecting the transmittancy by more than a fraction of a per cent. The values from 350mμ to 390mμ in Table 5 were thereupon adopted.

(d) SOLUTION B'$_2$

It has already been noted that cobalt sulphate was at first used instead of the cobalt ammonium sulphate later adopted because of its greater stability. After preliminary measurements had indicated that 10.000 g of cobalt sulphate per liter was a suitable concentration, an unfiltered solution was prepared, No. 61–10, containing, however, but 4 cc of sulphuric acid. (The exact amount of sulphuric acid is apparently of little, if any, importance.) Measurements were made visually, thermoelectrically and photoelectrically, covering the wavelength range from 390mμ to 750mμ. Values of absorbancy were thereupon adopted (May, 1926). Later, after a study of Beer's law had been made, with filtered solutions, Nos. 62–8–A to E containing 10 cc of sulphuric acid, and no deviation from Beer's law detected, the values were revised slightly (October, 1926).

In the latter part of October, 1926, a new solution was prepared, No. 62–17, unfiltered, containing 10 cc of acid, and measurements were made on the day of preparation and on the first, fourth, thirteenth, and fifteenth days after preparation, using the eight Hg and He reference wave lengths. No change of absorbancy was detected over this interval of time.

Based principally on these measurements on solution No. 62–17, but guided also by the other previous measurements, the absorbancy values given at the bottom of Table 6 were adopted (November 13, 1926). In a manner similar to that for solutions A' and B$_1$', the revised data of October, 1926, were adjusted to fit these adopted data of November 13, 1926, and the data thus finally adopted (November 17, 1926) are given in Table 6, 400mμ to 750 mμ (see also fig. 10). The absorbancy values finally adopted November 17, 1926, differ from those originally obtained in May, 1926, in no case by more than 0.0028. The data summarized in Table 20 indicate that a turbidity effect, such as was found with copper sulphate, was not present, and it is apparently of no practical importance whether a B$_2$ solution is clarified by filtering or by decanting.

The loss of water of crystallization of cobalt sulphate (Sec. IV, 1) led to the analysis of solution No. 62–17, which was found to contain an amount of cobalt equivalent to 10.3 g of cobalt sulphate instead of 10.0 g as supposed. The change to cobalt ammonium sulphate was thereupon made. A single set of absorbancy measurements on a cobalt ammonium sulphate solution (No. 62–20) equivalent in cobalt content to No. 62–17, at the eight Hg and He wave lengths, agreed with the adopted data of November 13, 1926, to 0.0022 or better. Further test as to the agreement between absorbancy data for cobalt sulphate and cobalt ammonium sulphate was afforded by Beer's law measurements on cobalt ammonium sulphate. The measurements made (at 435.8mμ, 491.6mμ, 501.6 mμ, 546.1mμ, and 578.0mμ) on the solution having the standard concentration (No. 65–1–C) agreed with the adopted values to 0.0015 or better. A more important check, however, was obtained on the mixture of cobalt ammonium sulphate and copper sulphate, as indicated in the next section.

Photographic measurements were made from 450mμ to below 350mμ on cobalt ammonium sulphate as with copper sulphate. The absorbancy is very small below 400mμ. Measurements from one to two weeks after preparation were made on solution No. 65–4, 14.481 g,

and on No. 65–9, approximately 144.81 g, using 40-mm cells in both cases. With the latter solution Beer's law was assumed for computing to 14.481 g. Perfect agreement with the adopted data of Table 6 from 400mμ to 450mμ was not obtained. The differences, however, were not outside the uncertainties of the photographic data. The data were adjusted to fit the adopted values from 400mμ to 450mμ and are given as adopted in Table 6. Inasmuch as the absorbancy is itself less than 0.01 below 400mμ, there is no chance for any important error to be present in the adopted data.

Comparative measurements showed a difference in absorbancy between cobalt sulphate and cobalt ammonium sulphate below 400mμ. Cobalt sulphate has a minimum absorbancy at approximately 370mμ. For cobalt ammonium sulphate this minimum is shifted to about 355mμ, and the values of absorbancy are, in general, lower. As shown elsewhere, however, in the visible spectrum, the differences in transmittancy between cobalt sulphate and the equivalent amount of cobalt ammonium sulphate, at the concentrations used and for 10 mm of solution, must be considerably less than 1 per cent, if any difference exists at all.

(c) SOLUTION B′

Solution No. 65–2, unfiltered, was prepared containing 20.000 g of copper sulphate and 14.481 g of cobalt ammonium sulphate per liter of acidified aqueous solution. Measurements were made on this solution eight and nine days after preparation, using 10, 20, 40, and 100 mm of solution at the eight Hg and He wave lengths. The values were reduced to 10 mm thickness as usual and averaged. The maximum difference between these average observed absorbancies and those computed, adding the absorbancies tabulated for solutions B_1' and B_2', Tables 5 and 6, was but 0.0012, or less than one-third of 1 per cent of the transmittancy. (Table 21.)

In greater concentrations this strict additivity of the components of the B solution might not hold accurately, but over the range of concentrations used in the various filters, considering also various other spectrophotometric checks (Sec. VIII), it is certain that any error involved in using this additivity is negligible.

As implied in the preceding section, this set of measurements on Solution B′ also served as a check on the equivalence of cobalt ammonium sulphate and cobalt sulphate in the visible spectrum, for the latter had been used in adopting the values of Table 6 and the former in the preparation of solution No. 65–2.

3. PERMANENCE OF SOLUTIONS

There was no indication of change taking place in the separate copper sulphate and cobalt sulphate solutions, except the very slight turbidity effect noticed with the former, as already described. A much more important question is whether or not the mixture of the copper sulphate and the cobalt ammonium sulphate is unchanging in its transmittancy. Accordingly, when solution B′ above described (No. 65–2) was 300 days old, a set of measurements was made, similar to those made 9 days after preparation. The maximum deviation between the two sets of absorbancy values was 0.0028 (Table 21), about two-thirds of 1 per cent of the transmittancy, and no change at all was indicated outside of possible experimental error.

However, small but definite changes took place in the A′ solution with time. The absorbancy measurements on solution No. 62–15, using the eight Hg and He wave lengths, have already been noted. These were made on the day of preparation and on the first, second, sixth, eighth, twenty-third, and twenty-fourth days after preparation. Measurements were again made on this solution 384 and 385 days

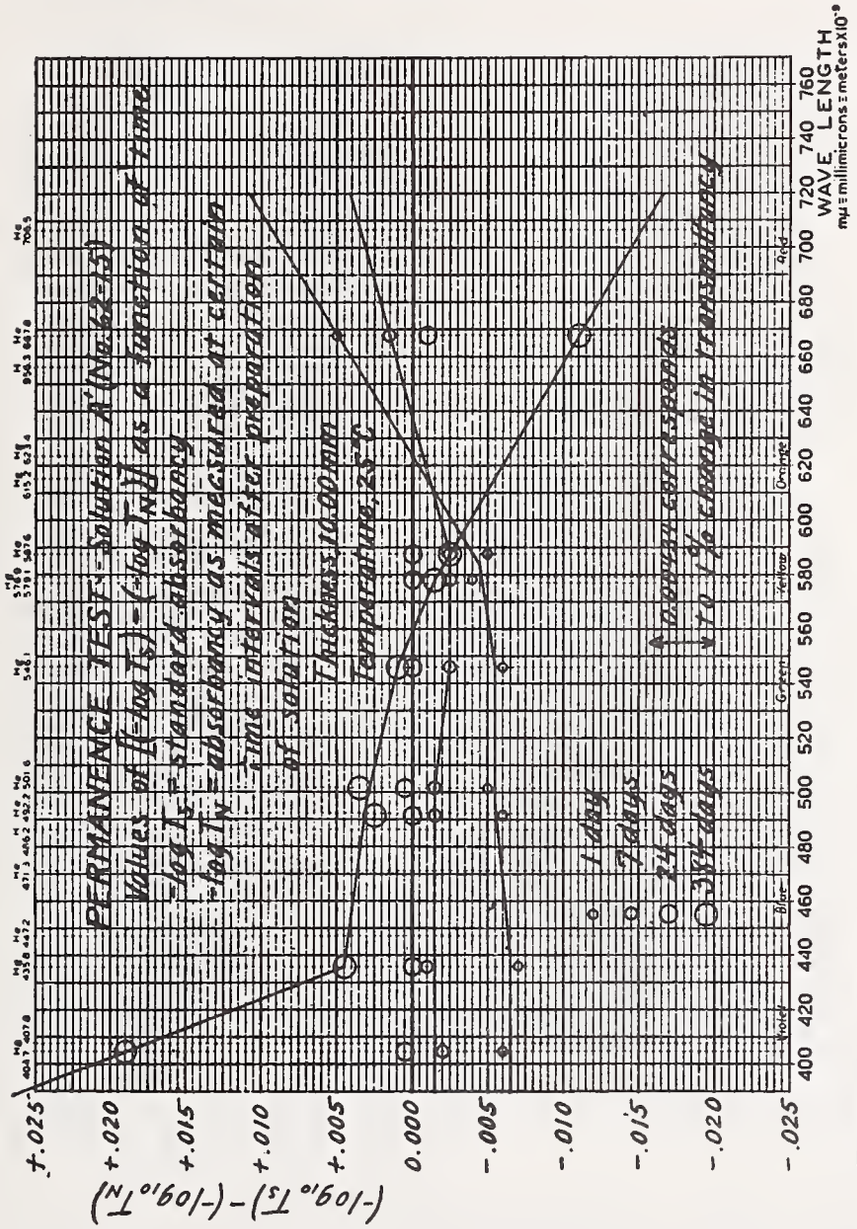


FIGURE 21.—*Effect of time on the absorbancy of the standard A′ solution—3 g each of copper sulphate and mannite, 30 cc of pyridine, per liter of aqueous solution*

Solution kept in glass-stoppered bottle exposed to ordinary room light and temperature conditions.

after preparation. The data for 0, 1, and 2 days were averaged together, likewise the data for 6 and 8 days, 23 and 24 days, and 384 and 385 days; so that, in effect, we have data on this solution representative of the first, seventh, twenty-fourth, and three hundred and eighty-fourth day after preparation. These data, given in Table 18 and plotted in Figure 21, show the differences in absorbancy between the data representative of the time intervals as noted and

the standard values of Table 4. The change throughout the brighter part of the spectrum is small, but there is a definite decrease in transmittancy in the red and an increase in the violet, resulting in the solution becoming slightly bluer after long standing. However, over a period from a week to a month, probably two months, the values do not deviate from the adopted values by more than 1 per cent of the transmittancy. The resulting effect of this and other factors on the color of completed filters is discussed further in Section VIII.

4. TEMPERATURE EFFECTS

As noted elsewhere, the standard temperature adopted for these solutions was 25° C. and all measurements of transmittancy or absorbancy reported in this paper refer to this temperature unless otherwise specifically noted. It is well known that aqueous inorganic salt solutions may show a relatively rapid change in transmittancy with temperature, whereas aqueous dye solutions are often practically unaffected.

It might, perhaps, have been assumed that any variation of absorbancy thus caused would be proportional to the concentration of the solution for a given temperature range. It seemed best, however, to test this.

(a) SOLUTION A

Three solutions were prepared (Nos. 65–17–A, 65–18–A, and 65–19–A), having concentrations of copper sulphate (and mannite) equal to 3.707, 2.445, and 0.741 g per liter, respectively. The pyridine content of each solution was 30 cc per liter, as usual. These concentrations are those used in the 2,360°, 2,848°, and 4,000° K.-to-sunlight filters, covering nearly the complete range of concentrations used.

For each concentration, measurements of absorbancy were made at 25° C., at 40° C., and again at 25° C. These measurements were made in December, 1927, from one to two weeks after preparation of the solutions. The eight Hg and He wave lengths were used in addition to 460mμ, 520mμ, 560mμ, 610mμ, 640mμ, 690mμ, and 720mμ. The 3.707 g solution was measured at 10 mm thickness, the 2.445 g at 20 mm, and the 0.741 g at 40 mm.

Values of absorbancy were then reduced to 10 mm thickness for all three solutions. The two sets of 25° C. values were averaged and taken to represent 25° C., as no permanent change was detected. The differences in absorbancies—that is, $A_{25°} - A_{40°}$—at the various wave lengths were then taken for each solution, and it was found that at practically every wave length this difference was proportional to the concentration of the solution within the experimental uncertainty.

These values of $A_{25°} - A_{40°}$ were thereupon divided by the concentration and are plotted in Figure 22. The general lack of any consistent differences in the values computed from the different concentrations shows that any attempt to correct these values on the basis of the Beer's law data already obtained would have been an unwarranted refinement. (The values at 404.7mμ, where the Beer's law correction is several times as large as at the other wave lengths, probably illustrate the effect of ignoring such a correction.)

There are given also in Figure 22 the mean values of $A_{25°} - A_{40°}$ per gram and a curve plotted through the adopted values. These adopted values are given in Table 16, and, with those obtained for solution B, were used in computations of the temperature effect on the completed filters. (Sec. VIII.)

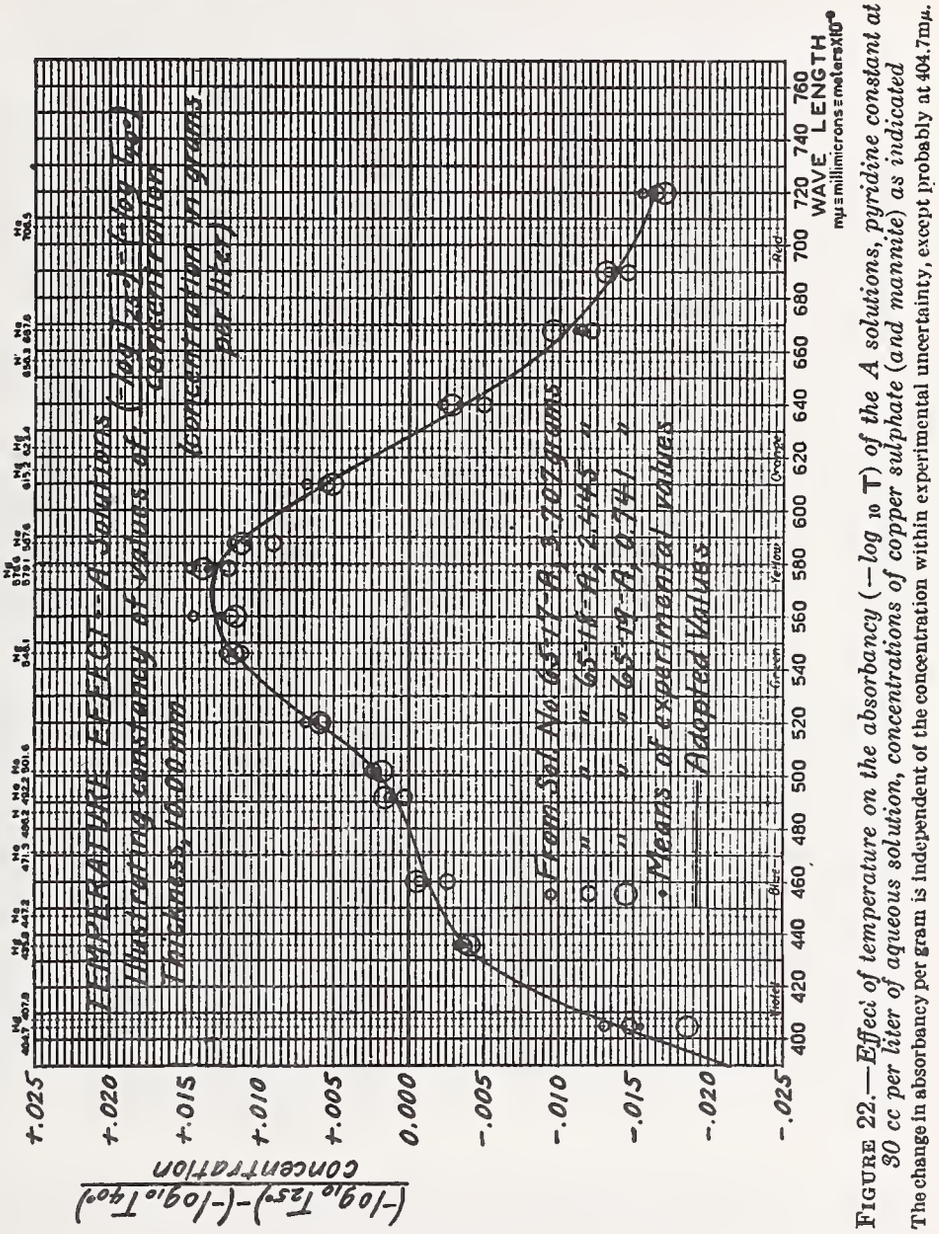


FIGURE 22.—*Effect of temperature on the absorbancy* ($-\log_{10} T$) *of the A solutions, pyridine constant at 30 cc per liter of aqueous solution, concentrations of copper sulphate (and mannite) as indicated*

The change in absorbancy per gram is independent of the concentration within experimental uncertainty, except probably at 404.7mμ.

(b) SOLUTION B

The transmittancies of the components of the B solution were known to decrease with increasing temperature. As in the study of permanence, however, the effect should be tested on the B mixture, and not computed from data on the two components, even if such were available in suitable form. Accordingly three B solutions were

prepared, having concentrations corresponding to the filters 2,360°, 2,848°, and 4,000° K. to mean sun, viz:

Number	Concentration (in grams per liter)	
	Copper sulphate	Cobalt ammonium sulphate
65-17-B..	27.180	26.827
65-18-B..	[1] 16.520	[1] 19.820
65-19-B..	6.250	1.130

[1] This solution was made up with the weights inadvertently reversed from the intended values. The mistake was not discovered until after the measurements were completed, hut no error or disadvantage results from using the data, the values of ahsorhancy being applied to the weights as actually used.

The measurements were made in December, 1927, from two to three weeks after preparation of the solutions. Thicknesses of solution of 10, 20, and 40 mm, respectively, were used, and the other experimental conditions were the same as for the A solutions above described.

Previous measurements (49) on a copper-sulphate solution at 57 g per liter have shown that the change with temperature was inappreciable at the shorter wave lengths, increasing as the absorbancy increased toward the longer wave lengths. It was also known (50, plate 30) that the principal temperature change with cobalt sulphate was in the green with little change in the red where the absorbancy is small. Furthermore, the weights of the two components are not greatly different for two of the three solutions.

Accordingly, the values of $A_{25°}-A_{40°}$ were reduced to a 10 g. basis (in a manner analogous to the A solutions, which were reduced to a 1 g basis) assuming the change in absorbancy below 600mμ to be caused by the cobalt ammonium sulphate alone and that above 600mμ by the copper sulphate alone. These values are plotted in Figure 23 for two of the three solutions. Values for solution No. 65–19–B are not given. The concentrations were so low that observational uncertainty masked the real effect over most of the range. These data were accordingly not used.

The curve drawn more or less through the average of the values for the two solutions was adopted as representative of the temperature effect on solution B per 10 g of each component. The separate values derived and estimated from this curve for each component are

given in Table 16 and plotted in Figure 23. Remembering that 0.00434 is equivalent to a 1 per cent change in transmittancy, such

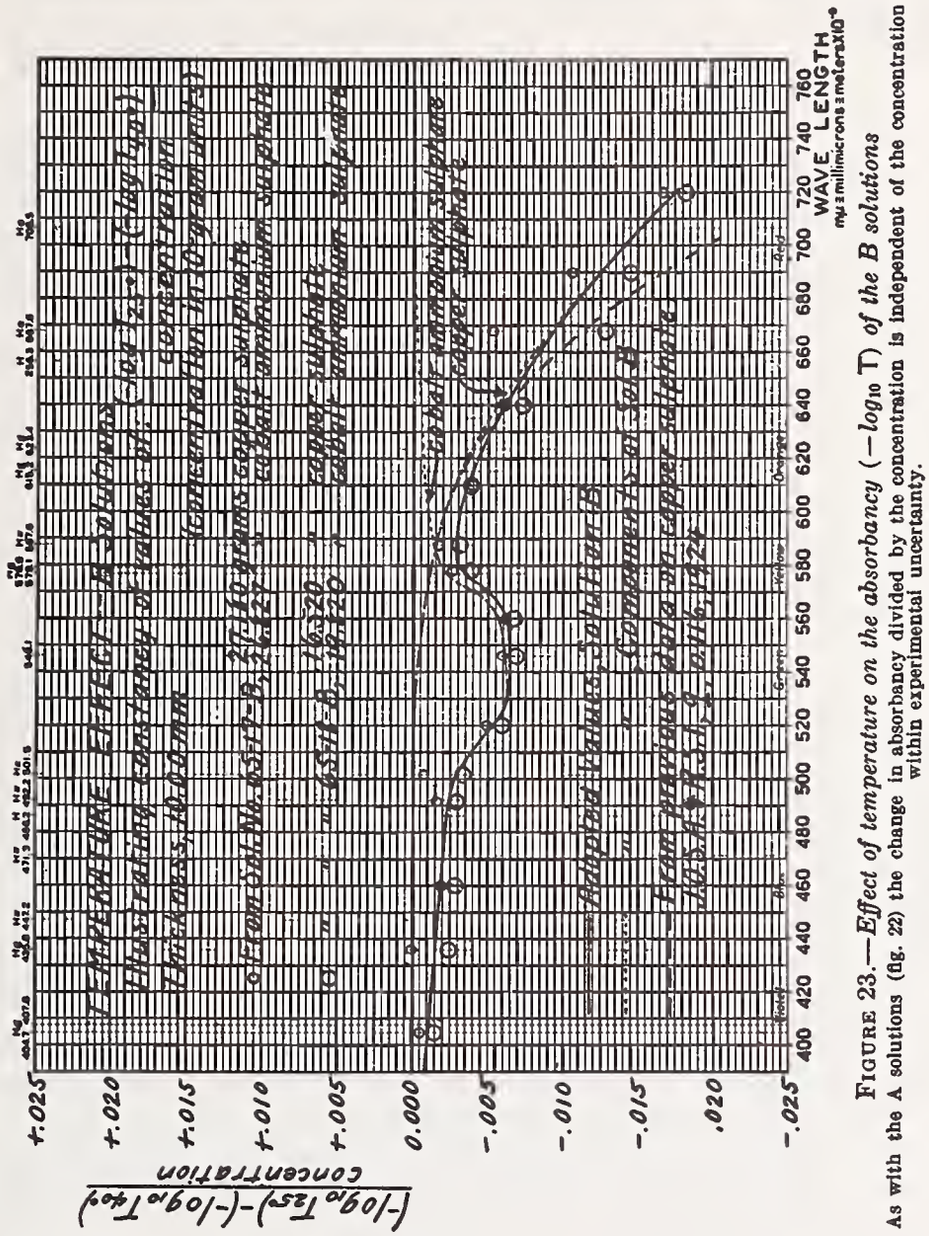


FIGURE 23.—*Effect of temperature on the absorbancy* ($-\log_{10} T$) *of the B solutions*
As with the A solutions (fig. 22) the change in absorbancy divided by the concentration is independent of the concentration within experimental uncertainty.

uncertainty as is present in the data is probably negligible.[18] All the values of Table 16 are considered uncertain in the fourth decimal.

[18] An apparent discrepancy may be noted (fig. 23) between the present data for copper sulphate and the data referred to in a previous publication (49), particularly at wave lengths greater than 650mμ. This difference is probably not outside of the possible combined experimental error in the two cases, but may also be caused by the different experimental conditions, viz, 57 g per liter, no acid, no mixture with cobalt ammonium sulphate in the previous determination, as contrasted with the respective differing conditions of the present determination.

TABLE 16—*Changes in absorbancy*[1] *with temperature for component filter solutions*

Wave length in millimicrons	Values of $\frac{A_{230}-A_{400}}{\text{concentration}}$ for			Wave length in millimicrons	Values of $\frac{A_{230}-A_{400}}{\text{concentration}}$ for		
	A solution per gram	B_1 solution per 10 g	B_2 solution per 10 g		A solution per gram	B_1 solution per 10 g	B_2 solution per 10 g
380	−0.0267	0.0000	−0.0008	560	+0.0129	−0.0006	−0.0057
390	−.0212	.0000	−.0009	570	+.0132	−.0009	−.0040
				580	+.0127	−.0012	−.0018
400	−.0162	.0000	−.0010	590	+.0110	−.0016	−.0013
410	−.0116	.0000	−.0011				
420	−.0078	.0000	−.0012	600	+.0084	−.0021	−.0010
430	−.0049	.0000	−.0014	610	+.0057	−.0028	−.0008
440	−.0032	.0000	−.0015	620	+.0026	−.0036	−.0007
				630	−.0007	−.0045	−.0006
450	−.0021	.0000	−.0017	640	−.0037	−.0056	−.0005
460	−.0013	.0000	−.0018				
470	−.0006	.0000	−.0019	650	−.0066	−.0068	−.0004
480	.0000	.0000	−.0021	660	−.0090	−.0081	−.0003
490	+.0008	.0000	−.0023	670	−.0109	−.0095	−.0002
				680	−.0124	−.0110	−.0001
500	+.0020	.0000	−.0027	690	−.0138	−.0125	.0000
510	+.0038	.0000	−.0039				
520	+.0062	.0000	−.0053	700	−.0148	−.0141	.0000
530	+.0087	−.0001	−.0062	710	−.0157	−.0158	.0000
540	+.0107	−.0002	−.0062	720	−.0164	−.0175	.0000
550	+.0123	−.0004	−.0060				

[1] All the values of this table are uncertain in the fourth decimal place. See Section VII, 4, and Figures 22 and 23 for details of derivation.

5. VARIATION OF PYRIDINE CONTENT IN SOLUTION A

To study and illustrate the effect of varying the amount of pyridine used, five solutions were prepared, Nos. 62–26–A to E, all containing 3 g of copper sulphate and mannite, and containing, respectively, 10, 20, 30, 50, and 100 cc of pyridine per liter of solution. Measurements were made in November, 1927, from 4 to 12 days after preparation of the solutions. Solution thicknesses of 10 and 100 mm were used. The eight Hg and He wave lengths were used, in addition to 460mμ, 620mμ, and 720mμ in some cases.

The values obtained are illustrated in Figure 24 compared with the standard data for solution A′ as given in Table 4. The effect of increasing the pyridine content is to increase the absorbancy over most of the spectral range and to shift the wave length of maximum absorbancy toward shorter wave lengths. The effect in the blue and violet is small, but may, of course, increase in the ultra-violet.

The results of this test have been further illustrated in Figure 8 (Sec. IV, 1) where the absorbancy is plotted as a function of the pyridine content at the various wave lengths. The smooth curves therein drawn pass through the standard values for 30 cc; and the effect of any variation of pyridine strength may thus be obtained at these wave lengths from this figure.

Similar results had been found by preliminary measurements made in 1926. Computations based on the later data illustrated in Figures 8 and 24 show that little if any improvement in the energy distribution of the completed filters could have been obtained with a pyridine content other than the standard concentration chosen. This is illustrated in Section VIII, 3.

6/ BEER'S LAW

(a) SOLUTION A

Measurable deviations from Beer's law were found for solution A. The principal measurements were made on solutions Nos. 62–13–A to E, containing 1, 2, 3, 4, and 5 g, respectively, of copper sulphate

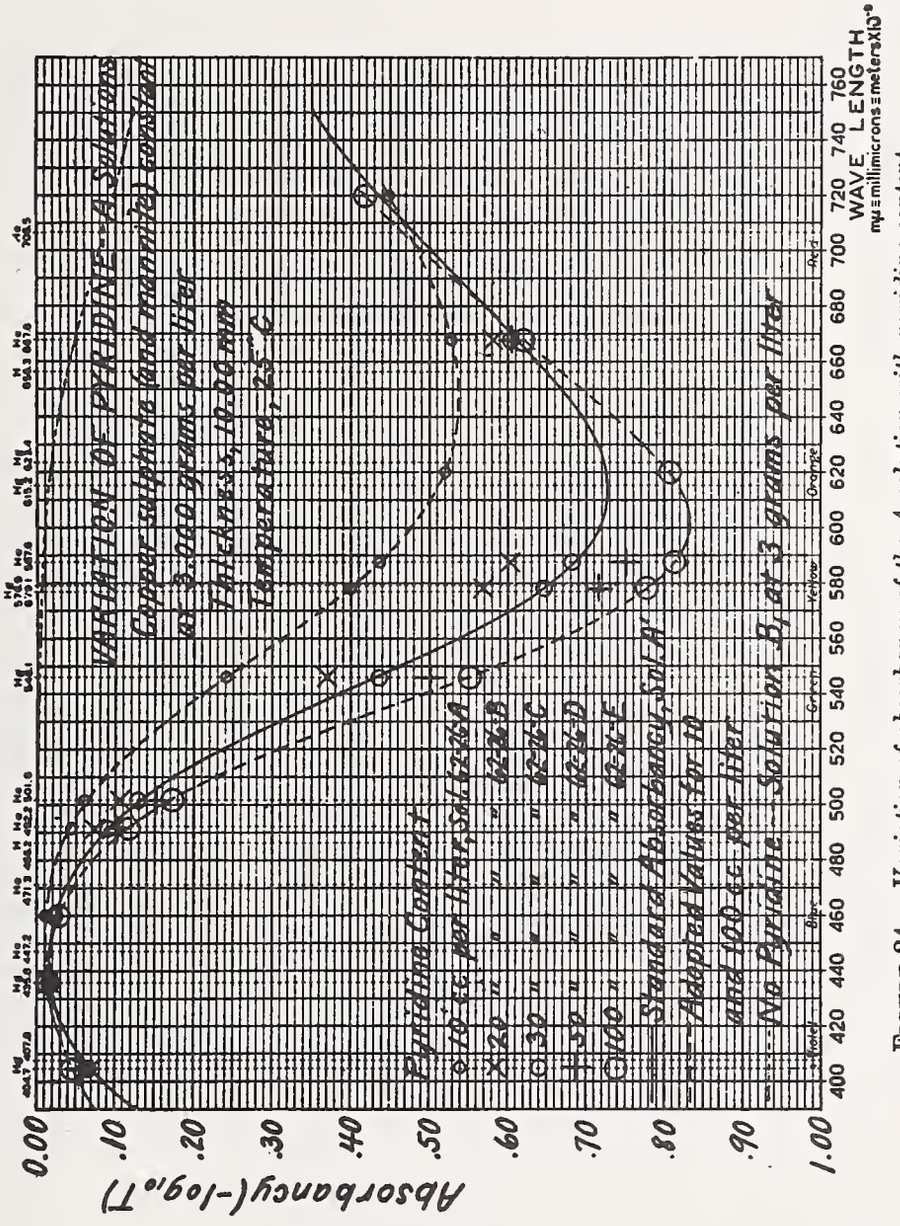


FIGURE 24.—*Variation of absorbancy of the A solution with pyridine content*
These data are plotted as a function of pyridine content in Figure 8.

and mannite per liter of solution, each solution containing also 30 cc of pyridine. Measurements were made nine days after preparation of the solutions, at 404.7mμ, 435.8mμ, 546.1mμ, 578.0mμ, and 680mμ for each solution. Thicknesses of 10, 20, and 40 mm were used, the thickness used depending on the concentration. Reduction of the data to 10 mm and 3 g resulted in the values given in Table 17,

columns 2 to 6. Each value of $A\frac{c_s}{c}$ there tabulated represents the absorbancy measured, divided by the thickness (in centimeters), and further divided by one-third of the concentration (in grams of copper sulphate and mannite per liter). They are thus proportional to the specific absorptive indices, and their variation with concentration illustrates the failure of Beer's law. Within experimental uncertainty and to an accuracy sufficient for the purpose, this change in the absorbancies of Table 17 may be taken as constant from one concentration to another at any wave length. An average change per gram is accordingly tabulated in cloumn 7 of Table 17, these values being computed by averaging the differences between the 3 g values of column 4 and the values of columns 3 and 5 and one-half the differences between column 4 and columns 2 and 6.

TABLE 17.—*Beer's law data for solution A*

Thickness 10.00 mm, temperature 25° C. Other experimental data are illustrated in Figure 25. The Beer's law correction factor is defined as (*see* Secs. V, 3 and VII, 6):

$$B=\frac{A\left(\frac{c_s}{c}\right)-A_s}{(c-c_s)\,A_s},$$

wherein A_s and c_s refer to the 3-g solution.

Wave length in millimicrons	Values of $A\left(\frac{c_s}{c}\right)$ for solutions Nos. 62-13-A to E having the following concentrations in grams of copper sulphate and mannite per liter (pyridine content 30.0 cc throughout)					Mean value of $\frac{A\left(\frac{c_s}{c}\right)-A_s}{c-c_s}$ (from preceding columns)	$\frac{A\left(\frac{c_s}{c}\right)-A_s}{(c-c_s)\,A_s}$ equals B (from columns 7 and 4)
	1.000	2.000	3.000	4.000	5.000		
404.7	0.088	0.080	0.076	0.068	0.073	−0.0049	−0.065
435.8	.023	.022	.022	.020	.025	−.0002	−.009
546.1	.448	.438	.435	.427	.423	−.0059	−.014
578.0	.658	.652	.647	.640	.634	−.0060	−.009
630	.577	.573	.576	.568	.570	−.0021	−.004

This average change per gram is seen to approximate an average change in transmittancy of about 1 per cent. For purposes of applying the Beer's law correction, it is more convenient, however, to express the change as a fractional change in the absorbancy. This is done in column 8 of Table 17, these values resulting from dividing the values of column 7 by those of column 4. They are plotted in Figure 25. This fractional change in absorbancy per gram difference from the standard has been defined as B (Sec. V, 3) in the equations

$$A=\frac{A_s}{c_s}c\left[1+B\,(c-c)_s\right]$$

and

$$B=\frac{A\left(\frac{c_s}{c}\right)-A_s}{(c-c_s)\,A_s}.$$

These measurements were made on unfiltered solutions, the usual condition unless otherwise noted. Similar measurements on a series of filtered solutions gave a similar set of values of B, these being

also shown in Figure 25. Measurements were also made with 2 and 4 g unfiltered solutions, with mannite constant at 3 g, and likewise with mannite constant at 10 g. These results are all plotted in Figure 25. Only at 404.7mμ was it felt that the value of B was certainly different from -0.010. Accordingly, this value was adopted for the wave-length range from 450mμ to 720mμ.

In order to confirm the results obtained at 404.7mμ and to extend the correction into the ultra-violet, photographic measurements were made. Six solutions were prepared, Nos. 65–8–(½)A and 65–8–A to E, containing, respectively, ½ g and 1 to 5 g of copper sulphate and mannite per liter of solution. Each solution contained 30 cc of pyridine per liter. Measurements were made three weeks after preparation of the solutions. The results were derived in a manner similar to the values of Table 17 and gave the values of B shown in Figure 25. In the same figure are shown also the adopted values of B—from 720mμ to 400mμ on the basis of the previous visual measurements described above, and from 390mμ to 350mμ on the basis of the photographic measurements. These adopted values are those given in Table 4. Measurements of the effect were again made visually at 404.7mμ, 435.8mμ, and 491.6mμ with the six solutions used in the photographic measurements, four weeks after preparation. These values are also given in Figure 25.

It might reasonably be questioned, after examination of Figure 25, whether the adopted values of the Beer's law correction factor represent the experimental data as satisfactorily as some alternative curve which might be plotted. As a matter of fact, if the true Beer's law correction factors were of primary importance, more data would have to be taken to determine them with high accuracy. Here, as elsewhere throughout the investigation, the influence of the data on the complete filters was effective in determining the extent of the data.

For example, could any other set of Beer's law correction factors be adopted on the basis of the experimental data of Figure 25 which would change the computed spectral transmission of any of the filters (and, therefore, the resulting energy distribution) by as much as 1 per cent? That there could be no such set of values may be shown as follows:

If $\Delta \mathbf{A}$ represent the error in $\mathbf{A}$ caused by an error ΔB in B, then from the first equation above, we have

$$\Delta \mathbf{A} = \Delta B \left[\frac{c\ (c - c_s)\ \mathbf{A}_s}{c_s} \right]$$

For any given values of ΔB, c_s, and $\mathbf{A}_s$, it is obvious that $\Delta \mathbf{A} = 0$ when $c = 0$ or when $c = c_s$. For values of c greater than c_s, $\Delta \mathbf{A}$ continuously increases as c increases. By differentiating $\Delta \mathbf{A}$ with respect to c and equating to zero, we obtain $c = c_s/2$, so that for values of c between 0 and c_s, $\Delta \mathbf{A}$ has a maximum absolute value when $c = c_s/2$, or in the present case when $c = 1.5$ g of copper sulphate.

We wish, however, to know what values of ΔB will cause $\Delta \mathbf{A}$ to equal 0.00434, thereby producing a 1 per cent error in the transmittancy. The equation connecting $\Delta \mathbf{A}$ and ΔB may be rewritten

$$\Delta B = \frac{\frac{c_s}{c} \Delta \mathbf{A}}{(c - c_s)\ \mathbf{A}}$$

518-743 O - 73 - 48

as is shown in Figure 25. It follows, for any given values of $\Delta\mathbf{A}$, c_s, and $\mathbf{A}_s$, that ΔB is infinite when $c=c_s$ or when $c=0$, that for values of c greater than c_s, ΔB continuously decreases as c increases, and that for c less than c_s, ΔB has a minimum absolute value at $c=c_s/2$.

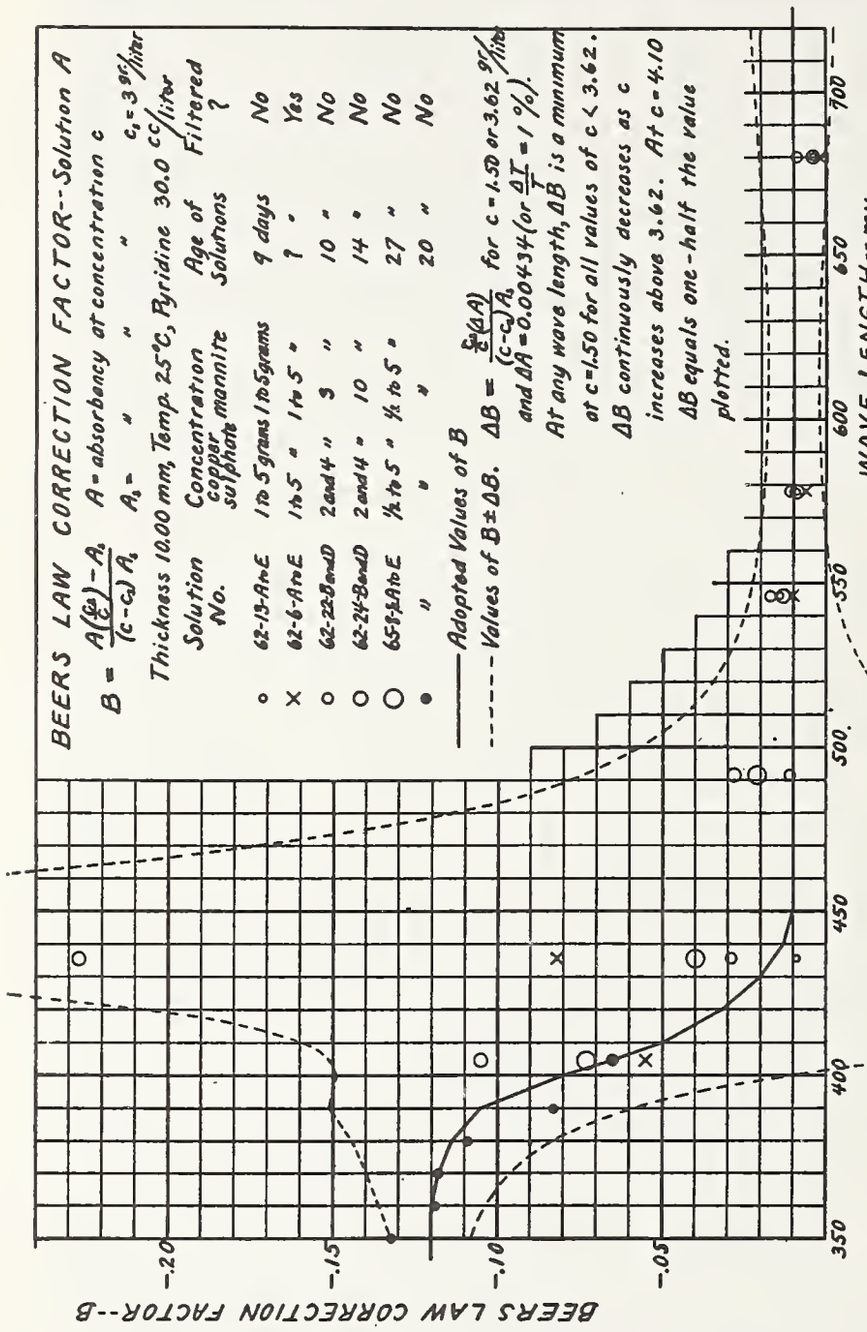


FIGURE 25.—*Beer's law data for the A solutions*

In the figure are shown the complete experimental data (the plotted points), the adopted values of the correction factor B (the continuous curve), and values of $B \pm \Delta B$ (the dashed curve). The values of ΔB illustrated would produce errors of 1 per cent in the transmittancy for values of $c=1.50$ or 3.62 g of copper sulphate. They would cause errors of less than 1 per cent for all other values of c less than 3.62. They would cause errors greater than 1 per cent for values of c greater than 3.62, these being 2 per cent at $c=4.10$ and 3 per cent at $c=4.5$ g. The figure illustrates that any other values of the Beer's law correction factor that could have been adopted from these experimental data would have produced negligible differences in the transmittancy of the solution, except possibly in the red for the highest concentrations.

Values of ΔB have been computed for values of $\Delta\mathbf{A}=0.00434$, for values of $\mathbf{A}_s$ as given in Table 4, with $c_s=3$, and with a value of $c=1.50$ (or 3.62, which gives practically the same absolute values of ΔB). These values of ΔB, which are required to produce a 1

per cent error in the transmittancy, are illustrated in Figure 25 where values of $B \pm \Delta B$ are plotted as a function of wave length.

It will be noted that these dashed curves in practically all cases lie much beyond the range of the experimental data, and it has just been shown that for any values of c less than 3.62, other than 1.50, the deviations would be even greater, becoming infinite at $c=0$ and 3. For values of c greater than 3.62 the values of ΔB which will cause a value of $\Delta A = 0.00434$ decrease, becoming one-half the values illustrated at $c=4.10$, and one-third the values at $c=4.50$, which is a greater concentration than used in any of the filters. Even for this extreme case, any curve which might reasonably be put through the plotted data of Figure 25 would still be within the 1 per cent limits, except possibly above 600mμ.

Values of ΔB computed as indicated above from the equation given in Figure 25, with c, c_s, and ΔA constant, are, of course, inversely proportional to the standard absorbancy A_s, which explains the general shape of the $B \pm \Delta B$ curves. The irregularity near 400mμ would be eliminated by adopting slightly different values of B in this region.

(b) SOLUTION B

No failure of Beer's law was detected for the components of the B solution. Measurements were made on copper sulphate, cobalt sulphate, and cobalt ammonium sulphate, as follows: Five copper sulphate solutions were prepared, Nos. 62–7–A to E, filtered, containing, respectively, 5, 10, 20, 30, and 40 g of copper sulphate per liter of solution. Each solution contained 10 cc of sulphuric acid. Thicknesses of 40 and 100 mm were used. Measurements were made at 435.8mμ, 546.1mμ, 578.0mμ, and 640mμ. The values of absorbancy for each solution were then computed to 10 mm thickness and 20 g of copper sulphate. When this was done the maximum total variation for any of the five solutions at any of the four wave lengths was 0.004, with no indication of any consistent variation with concentration.

In a similar manner five solutions of cobalt sulphate were prepared, Nos. 62–8–A to E, filtered, containing, respectively, 2.5, 5, 10, 15, and 20 g of cobalt sulphate per liter of solution. Each solution contained 10 cc of sulphuric acid. Thicknesses of 40 and 100 mm were used. Measurements were made at 435.8mμ, 491.6mμ, 546.1mμ, and 578mμ. Upon reduction of the absorbancies to 10 mm and 10 g, the maximum total variation for the five solutions at any of the four wave lengths was but 0.002.

Three solutions of cobalt ammonium sulphate were measured, Nos. 65–1–A, C, and F, unfiltered, containing, respectively, 4.827, 14.481, and 28.962 g of cobalt ammonium sulphate per liter of solution. Each solution contained 10 cc of sulphuric acid. Thicknesses of 40 and 100 mm were used. Measurements were made at 435.8mμ, 491.6mμ, 501.6mμ, 546.1mμ, and 578.0mμ. Upon reduction of the absorbancies to 10 mm and 14.481 g, the maximum total variation for any solution at any of the five wave lengths was 0.0014. They are undoubtedly merely a result of experimental uncertainty. Furthermore, both copper sulphate and cobalt sulphate in aqueous solution have previously been found to obey Beer's law over a larger range of concentrations (51, p. 31; 52, p. 775).

There still might remain the question of whether the mixture of copper sulphate and cobalt ammonium sulphate would obey Beer's law, as mixed in the present filters to form solution B. This was not tested directly, but two of the experimental checks elsewhere described show that any failure is quite negligible. These checks are:

1. The measurements on the completed 4,000° K.-to-mean-sun filter. (Sec. VIII, 1.) The concentrations of both components of solution B are in this case near their lower limits. The very good check obtained between the computed and directly measured values leaves no reasonable question of the practical obedience to Beer's law at the low concentrations.

2. The test of reproducibility (Sec. VII, 8) on solution B of the 2,360° K.-to-mean-sun filter. In this case the concentrations of both components are near their higher limits. The values thus obtained, compared with the standard computed values for the 2,360° K. B solutions, show that there can be no important deviation of Beer's law at the higher concentrations.

7. RELIABILITY OF THE SPECTROPHOTOMETRIC MEASUREMENTS

The data given in the present section are intended primarily to illustrate the reliability of the adopted absorbancy values given in Tables 4, 5, and 6 as pertaining to solutions prepared from the chemicals described in Section IV, 2. They do not, except where clearly indicated, illustrate the reproducibility obtainable from different lots of chemicals, which subject is considered in Section VII, 8, following.

(a) AT THE Hg AND He WAVE LENGTHS

The most fundamental and reliable spectrophotometric data on the solutions are the absorbancy values given for the standard solutions, Nos. A′, B_1', B_2', and B′, at the eight reference Hg and He wave lengths, viz, 404.7mμ, 435.8mμ, 491.6mμ, 501.6mμ, 546.1mμ, 578.0mμ, 587.6mμ, and 667.8mμ. The use of these light sources eliminates the possibility of errors arising from incorrect wave-length calibrations and the use of finite slit widths with a continuous spectrum source. The data with the homogeneous light sources at these eight wave lengths are, in general, quite extensive, including usually measurements on several different solutions and at various thicknesses and intervals after preparation.

A complete summary of the experimental values obtained on the standard solutions is given in Tables 18 to 21. As a function of wave length is tabulated first the adopted absorbancy values and then for the various solutions, as indicated, values of $A - A_s$ where A_s refers to the adopted absorbancy values and A the experimental values. Algebraic means are also tabulated as may be noted. In compiling Tables 18 to 21, no data which were taken at any time during the investigation have been omitted, whether obtained before or after the standard values were adopted, except that in Table 18 only those data are shown which were obtained on solutions prepared from commercial pyridine. Where values are tabulated representative of two or more determinations they are an arithmetical mean, unweighted, all values being reduced to 10 mm thickness before being averaged.

The data for solutions B_1' and B_2' show, in general, higher precision and better agreement with the adopted values than do the data for solution A′, particularly if the data for the first four days on solution B_1' be excluded, as they were when the standard values were adopted. This difference between the A′ solution and the components of the B′ solution results, at least principally, from two causes:

1. In the concentrations as used and for these eight wave lengths the absorbancies for solution A′ reach a higher value than for either B_1' or B_2'. Above 546.1mμ, thicknesses of 10 and 20 mm were the only ones used for solution A′, whereas for solution B_1' and B_2', thicknesses from 10 to 100 mm could be used, except only at 667.8mμ with B_1'. Increased thickness makes for more precise determination when the values are reduced to a 10 mm basis.

2. Solutions B_1' and B_2' show no important change with time as does A′. (Sec. VII, 3.)

The absorbancies for solution B′, the sum of those adopted for B_1' and B_2', should, therefore, have a reliability fully the equal of, probably better than, solution A′. It is rather difficult to summarize the data further than is done in Tables 18 to 21. It seems, however, correct to say—considering the respective algebraic means of the deviations of the complete data from the adopted data, for a time interval from a week to perhaps two months after preparation, for the chemicals as analyzed and used, and possibly excepting the data at 404.7mμ and 667.8mμ—that the adopted absorbancies at these eight wave lengths for solutions A′, B_1', B_2', and B′ in 10 mm thicknesses are accurate to 0.001 or better, or one-quarter per cent of the transmittancy. (The accuracy of the photometric scale of the Koenig-Martens spectrophotometer is, of course, involved in such a conclusion. As shown in a recent paper by McNicholas (45), by tests with accurately calibrated rotating sectors, part of which tests were carried out during the period of the present measurements, no error in the photometric scale of this particular instrument has ever been found, provided that measurements are made under such experimental conditions that the angle of photometric match is not close to the 0° or 90° points. The accuracy of the present data is further certified by the fact that the adopted data are in all cases based on measurements with two or more thicknesses of solution.)

TABLE 18.—*Experimental data on solution A', standard aqueous pyridine solution of copper sulphate and mannite, unfiltered, at the eight Hg and He wave lengths*

[Thickness 10.00 mm, temperature 25° C.]

Wave length in millimicrons	Standard absorbancy A_s	Absorbancy as measured, A, minus adopted standard absorbancy, A_s.[1]														
		Solution number													Algebraic mean [4]	
		62-15 [2]	62-15 [2]	62-15 [2]	62-15 [2]	62-2	62-2	None	62-13-C	62-13-C	62-26-C	62-26-C	65-8-C	65-24 [3]		
		Determinations														
		3	2	2, 3, 4	2, 3, 4	1, 2	2, 3	1, 2	1	1	2	1	1	2, 3		
		Days after preparation													All data	Excluding values for > 60 days
		0, 1, 2	6, 8	23, 24	384, 385	3	27, 32	101, 117	9	42	4, 5	12	27	19–22		
Hg 404.7	0.068	+0.006	+0.002	−0.0005	−0.019	+0.009	−0.001	−0.0165	+0.008	−0.006	−0.001	−0.0045	+0.0045	0.000	−0.0029	+0.0019
Hg 435.8	.018	+.007	+.001	.000	−.0045	+.0045	−.002	−.0025	+.004	−.003	−.002	−.001	+.001	−.001	+.0001	+.0013
Hg 491.6	.091	+.006	+.0015	+.0005	−.0025	----------	----------	−.003	----------	−.002	−.001	−.0015	−.0015	−.0025	+.0002	+.0011
He 501.6	.131	+.0055	+.0015	−.0005	−.0035	----------	----------	−.005	----------	.000	−.0035	−.001	----------	+.001	−.0005	+.0007
Hg 546.1	.434	+.0055	+.0025	.000	−.001	−.002	−.0015	−.0055	+.001	+.001	+.002	----------	----------	−.003	+.0003	+.0013
Hg 578.0	.646	+.0035	+.0025	.000	+.0015	−.004	−.001	−.001	+.001	.000	.000	----------	----------	+.0025	+.0006	+.0006
He 587.6	.685	+.005	+.0025	.000	+.0025	----------	----------	−.001	----------	.000	−.004	----------	----------	+.001	+.0012	+.0012
He 667.8	.610	−.0045	−.0015	+.001	+.011	----------	----------	+.008	----------	−.002	−.0075	----------	----------	−.004	−.0001	−.0032

[1] $A-A_s=0.00434$ is equivalent to a 1 per cent difference in transmittancy.
[2] These values, illustrating changes in solution A' with time, are plotted in Figure 21. See also data on solution (unnumbered) at 101 and 117 days after preparation
[3] These data were obtained in June, 1929. A new stock of pyridine, from different source, was used. Corresponding value at 620 $m\mu=+0.004$. See section VII, 8 (end) for details.
[4] Weighted according to the number of determinations; No. 65-24 not included.

TABLE 19.—*Experimental data on solution B_1', standard acidified aqueous solution of copper sulphate, at the eight Hg and He wave lengths*

[Thickness 10.00 mm., temperature, 25° C.]

Wave length in millimicrons	Standard absorbancy A_s	Absorbancy as measured, A, minus adopted standard absorbancy, A_s[1]								
		Solution number							Algebraic means[3]	
		62-18[2]	62-18	62-1	62-1	62-5[2]	62-7-C	62-7-C		
		Determinations								
		3	2, 3, 4	1	1	1	1	1		
		Days after preparation								
		0, 1, 4	12, 14	4	27	2	7	35	All data	Excluding values for <5 days
		Solution filtered?								
		No	No	Yes	Yes	No	Yes	Yes		
Hg 404.7	0.0021	+0.0031	−0.0004	+0.0012	+0.0005	+0.0031	----------	+0.0001	+0.0015	0.0000
Hg 435.8	.0013	+.0030	−.0001	+.0008	+.0002	+.0030	0.0000	.0000	+.0013	.0000
Hg 491.6	.0019	+.0030	+.0003	----------	----------	----------	----------	−.0005	+.0015	.0000
He 501.6	.0028	+.0023	+.0004	----------	----------	----------	----------	−.0006	+.0012	+.0001
Hg 546.1	.0135	+.0021	−.0009	+.0003	−.0001	+.0030	+.0001	−.0005	+.0005	−.0006
Hg 578.0	.0368	+.0019	−.0011	+.0004	.0000	+.0030	+.0002	−.0002	+.0004	−.0006
He 587.6	.0487	+.0018	+.0002	----------	----------	----------	----------	−.0013	+.0006	−.0001
He 667.8	.319	+.001	.000	----------	----------	----------	----------	−.001	+.0003	−.0002

[1] $A-A_s=0.00434$ is equivalent to a 1 per cent difference in transmittancy.
[2] These data illustrate turbidity effect of first few days with unfiltered solutions.
[3] Weighted according to the number of determinations.

TABLE 20.—*Experimental data on solution B_2' standard acidified aqueous solution of cobalt sulphate or cobalt ammonium sulphate, at the eight Hg and He wave lengths*

[Thickness 10.00 mm, temperature 25° C.]

Wave length in millimicrons	Standard absorbancy A_s	Absorbancy as measured, A, minus adopted standard absorbancy, A_s[1]							
		Material							
		Cobalt sulphate					Cobalt ammonium sulphate		
		Solution number							Algebraic mean,[2] all data
		62-17	62-17	61-10	62-8-C	62-8-C	62-20	65-1-C	
		Determinations							
		3	2	3	1	1	1	1	
		Days after preparation							
		0, 1, 4	13, 15	(?)	3	38	4	16	
		Filtered?							
		No	No	No	Yes	Yes	No	No	
Hg 404.7	0.0144	+0.0006	+0.0004	+0.0003	----------	+0.0034	+0.0022	----------	+0.0009
Hg 435.8	.0437	+.0007	+.0002	+.0010	−0.0004	+.0023	+.0017	0.0000	+.0008
Hg 491.6	.1497	−.0001	+.0006	----------	−.0007	+.0001	+.0008	+.0015	+.0003
He 501.6	.1661	+.0009	+.0002	----------	----------	+.0025	−.0008	+.0001	+.0006
Hg 546.1	.0901	+.0010	−.0003	+.0019	−.0010	+.0013	+.0002	+.0007	+.0008
Hg 578.0	.0219	+.0005	−.0001	+.0005	−.0004	+.0013	+.0007	−.0001	+.0004
He 587.6	.0167	+.0004	+.0001	----------	----------	+.0023	+.0007	----------	+.0006
He 667.8	.0089	+.0005	.0000	----------	----------	+.0021	+.0008	----------	+.0006

[1] $A-A_s=0.00434$ is equivalent to a 1 per cent difference in transmittancy.
[2] Weighted according to the number of determinations.

TABLE 21.—*Experimental data on solution B', standard acidified aqueous solution of copper sulphate and cobalt ammonium sulphate, unfiltered, at the eight Hg and He wave lengths*

[Thickness 10.00 mm, temperature 25° C.]

Wave length in millimicrons	Standard absorbancy A_s			Absorbancy as measured, A, minus adopted standard absorbancy, A_s[1]	
				Solution number	
				65-2	65-2
	For solution B'_1	For solution B'_2	For solution $B'=B'_1+B'_2$	Determinations	
				3, 4	3, 4
				Days after preparation	
				8, 9	300, 301
Hg 404.7	0.0021	0.0144	0.0165	−0.0010	−0.0038
Hg 435.8	.0013	.0437	.0450	−.0010	−.0028
Hg 491.6	.0019	.1497	.1516	+.0012	+.0011
He 501.6	.0028	.1661	.1689	−.0007	+.0005
Hg 546.1	.0135	.0901	.1036	+.0010	+.0001
Hg 578.0	.0368	.0219	.0587	−.0007	−.0009
He 587.6	.0487	.0167	.0654	−.0008	−.0022
He 667.8	.319	.0089	.328	+.001	.000

[1] $A-A_s=0.00434$ is equivalent to a 1 per cent difference in transmittancy.

(b) FROM 350mμ TO 720mμ

The discussion in Section VII, 1 and 2, regarding the methods of measurement and the extent of the data, should be recalled in this connection. It will be remembered that the data were in all cases adjusted so that the adopted values at the eight Hg and He wave lengths fall precisely upon a smooth curve drawn through the values adopted at every 10mμ.

In the neighborhood of these eight reference wave lengths, therefore, the accuracy of the data at every 10mμ should be practically equal to that for these eight wave lengths. It may well be true, however, at some of the intermediate wave lengths, particularly where the intervals are the largest, that the reliability of the data is slightly decreased. It is estimated that the accuracy of the standard absorbancies in these cases should hold to 0.002 or better, or to one-half of 1 per cent of the transmittancies. Outside of the range from 404.7mμ to 667.8mμ, the uncertainty may become greater.

8. SPECTROPHOTOMETRIC REPRODUCIBILITY OF THE CHEMICALS

The spectrophotometric accuracy of the data obtained with the chemicals as analyzed and used in this investigation is further discussed in Section VIII, 1, where spectrophotometric data are given for various completed filters. A question of equal importance is whether or not the data would be duplicated when materials are

obtained from various sources, with or without careful analysis. This has been investigated with the cooperation of Messrs. Jones and Richardson, of the research laboratory of the Eastman Kodak Co., in connection with special studies of the 2,360° K.-to-mean-sun filter recommended as a standard for photographic sensitometry in a resolution adopted in 1928 by the Seventh International Congress of Photography. (Sec. VI, 1.)

Three sets of solutions were prepared at the Eastman Kodak Co. and submitted to the bureau for spectrophotometric measurement. The materials were obtained from the sources listed in Table 25, and will be considered more in detail later in the section. The first two sets consisted of five A and five B solutions, numbered 1A to 5A and 1B to 5B, having compositions as specified in chart 2 for the 2,360° K.-to-mean-sun filter. So far as this first set of A solutions was concerned, the purposes of the test were partially defeated because the authors neglected to inform those preparing the solutions that they should not be filtered. Solutions Nos. 1A to 5A and 1B to 5B were filtered. As shown in Section VII, 7, it is a matter of practical indifference whether the B solutions are filtered or not; but this is not the case with the A solutions, so that the spectrophotometric measurements with the first set of A solutions may show among other things the effect of filtering the solutions.

Later on another set of A solutions, Nos. 1A–2 to 5A–2 was prepared at the Eastman Kodak Co. These A–2 solutions were not filtered; they were prepared from the same lots of copper sulphate as the previous A solutions; they were prepared from new lots of pyridine obtained from the same sources as before; the pyridine was carefully tested for strength, and the solutions prepared using the equivalent of 30.0 cc per liter of 98.4 per cent strength pyridine in each solution.

The spectrophotometric measurements on all three sets of solutions were made at the bureau. Values of absorbancy were measured at certain selected wave lengths, at 25° C., and at two or more thicknesses (except with the B solutions, where measurements were made at but one thickness in the regions of minimum absorbancy).

The results are shown in Tables 22, 23, and 24, wherein are given: (1) Standard values of absorbancy, that is, those obtained by computation using the standard spectrophotometric data of Tables 4 to 6 in the same manner as was done for the A and B components of all of the filters, (2) values of absorbancy obtained experimentally with the respective solutions, (3) differences between the standard and these observed absorbancies, (4) the percentage differences between the standard and observed transmittancies, and (5) the percentage ratios of the observed to the standard absorbancies, in those cases where the absorbancies are great enough to make these ratios of possible significance from an analytical standpoint.

TABLE 22.—*Test of spectrophotometric reproducibility of chemicals for solution A, filtered, 2,360° K. to mean sun; thickness, 10.00 mm, temperature, 25°C., age, 1 month. Pyridine not tested before preparation of solutions*

[Solutions prepared at Eastman Kodak Co.; spectrophotometric measurements at National Bureau of Standards]

Wave length in millimicrons	Spectrophotometric data [1]						
	Standard values	Experimental values for solutions Nos.					Mean 1A to 5A
		1A	2A	3A	4A	5A	
Hg 404.7	0.0802	0.0836	0.0786	0.0825	0.0760	0.0956	
		+.0034	−.0016	+.0023	−.0042	+.0154	
		+.8	−.4	+.5	−1.0	+3.5	+0.7
Hg 435.8	.0220	.0256	.0218	.0218	.0193	.0300	
		+.0036	−.0002	−.0002	−.0027	+.0080	
		+.8	.0	.0	−.6	+1.8	+.4
Hg 491.6	.1117	.1132	.0990	.1110	.1038	.1026	
		+.0015	−.0127	−.0007	−.0079	−.0091	
		+.3	−2.9	−.2	−1.8	−2.1	−1.3
He 501.6	.1607	.1576	.1504	.1598	.1541	.1561	
		−.0031	−.0103	−.0009	−.0066	−.0046	
		−.7	−2.4	−.2	−1.5	−1.1	−1.2
Hg 546.1	.5325	.5182	.5109	.5254	.5236	.5192	
		−.0143	−.0216	−.0071	−.0089	−.0132	
		−3.3	−5.0	−1.6	−2.0	−3.0	−3.
		97.3	95.9	98.7	98.3	97.5	
Hg 578.0	.7926	.7784	.7569	.7827	.7844	.7690	
		−.0142	−.0357	−.0099	−.0082	−.0236	
		−3.3	−8.2	−2.3	−1.9	−5.4	−4.2
		98.2	95.5	98.8	99.0	97.0	
He 587.6	.8405	.8292	.8072	.8288	.8380	.8232	
		−.0113	−.0333	−.0117	−.0025	−.0173	
		−2.6	−7.7	−2.7	−.6	−4.0	−3.5
		98.7	96.0	98.6	99.7	97.9	
Inc. 620	.8896	.8804	.8579	.8823	.8897	.8738	
		−.0092	−.0317	−.0073	+.0001	−.0158	
		−2.1	−7.3	−1.7	.0	−3.6	−2.9
		99.0	96.4	99.2	100.0	98.2	
Inc. 667.8	.7485	.7303	.7264	.7441	.7474	.7394	
		−.0182	−.0221	−.0044	−.0011	−.0091	
		−4.2	−5.1	−1.0	−.3	−2.1	−2.5
		97.6	97.0	99.4	99.9	98.8	
Average values of A/A_s (per cent) =		98.2	96.2	98.9	99.4	97.9	----------

[1] At each wave length the values given represent the following quantities:
First line, values of absorbancy, A_s for standard values, A for experimental values.
Second lines, values of $A-A_s$.
Third line, values of $(A-A_s)/0.00434$, representing the percentage difference in the standard and observed transmittancies.
Fourth line, values of A/A_s (per cent), at wave lengths indicated.

TABLE 23.—*Test of spectrophotometric reproducibility of chemicals for solution A, unfiltered, 2,360° K. to mean sun; thickness, 10.00 mm, temperature, 25° C., age, 1 to 2 weeks. Pyridine tested for strength and amount used equivalent to standard*

[Solutions prepared at Eastman Kodak Co.; spectrophotometric measurements at National Bureau of Standards]

Wave length in millimicrons	Spectrophotometric data [1]						
	Standard values	Experimental values for solutions Nos.					Mean 1A-2 to 5A-2
		1A-2	2A-2	3A-2	4A-2	5A-2	
Hg 404.7	0.0802	0.0899	0.0856	0.1154	0.0801	0.1055	
		+.0097	+.0054	+.0352	−.0001	+.0253	
		+2.2	+1.2	+8.1	.0	+5.8	+3.5
Hg 435.8	.0220	.0286	.0280	.0464	.0195	.0296	
		+.0066	+.0060	+.0244	−.0025	+.0076	
		+1.5	+1.4	+5.6	−.6	+1.8	+1.9
Hg 491.6	.1117	.1105	.1130	.1213	.1049	.1046	
		−.0012	+.0013	+.0096	−.0068	−.0071	
		−.3	+.3	+2.2	−1.6	−1.6	−.2
He 501.6	.1607	.1678	.1594	.1725	.1494	.1584	
		+.0071	−.0013	+.0118	−.0113	−.0023	
		+1.6	−.3	+2.7	−2.6	−.5	+.2
Hg 546.1	.5325	.5307	.5260	.5306	.5268	.5138	
		−.0018	−.0065	−.0019	−.0057	−.0187	
		−.4	−1.5	−.4	−1.3	−4.3	−1.6
		99.7	98.8	99.6	98.9	96.5	
Hg 578.0	.7926	.7876	.7830	.7833	.7839	.7752	
		−.0050	−.0096	−.0093	−.0087	−.0174	
		−1.2	−2.2	−2.1	−2.0	−4.0	−2.3
		99.4	98.8	98.8	98.9	97.8	
He 587.6	.8405	.8434	.8318	.8362	.8445	.8310	
		+.0029	−.0087	−.0043	+.0040	−.0095	
		+.7	−2.0	−1.0	+.9	−2.2	−.7
		100.3	99.0	99.5	100.5	98.9	
Inc. 620	.8896	.8888	.8808	.8854	.8905	.8772	
		−.0008	−.0088	−.0042	+.0009	−.0124	
		−.2	−2.0	−1.0	+.2	−2.9	−1.2
		99.9	99.0	99.5	100.1	98.6	
Inc. 667.8	.7485	.7495	.7433	.7522	.7454	.7454	
		+.0010	−.0052	+.0037	−.0031	−.0031	
		+.2	−1.2	+.9	−.7	−.7	−.3
		100.1	99.3	100.5	99.6	99.6	
Average values of A/A, (per cent) =		99.9	99.0	99.6	99.6	98.3	

[1] For explanation of values, see footnote to Table 22.

TABLE 24.—*Test of spectrophotometric reproducibility of chemicals for solution B, filtered, 2,360° K. to mean sun; thickness, 10.00 mm, temperature, 25°C., age, 1 month*

[Solutions prepared at Eastman Kodak Co.; spectrophotometric measurements at National Bureau of Standards]

Wave length in millimicrons	Spectrophotometric data [1]						
	Standard values	Experimental values for solutions Nos.					Mean 1B to 5B
		1B	2B	3B	4B	5B	
Hg 404.7	0.0296	0.0278	0.0288	0.0308	0.0269	0.0397	
		−.0018	−.0008	+.0012	−.0027	+.0101	
		−.4	−.2	+.3	−.6	+2.3	+0.3
Hg 435.8	.0828	.0771	.0788	.0814	.0838	.0841	
		−.0057	−.0040	−.0014	+.0010	+.0013	
		−1.3	−.9	−.3	+.2	+.3	−.4
Hg 491.6	.2799	.2646	.2703	.2811	.2770	.2706	
		−.0153	−.0096	+.0012	−.0029	−.0093	
		−3.5	−2.2	+.3	−.7	−2.1	−1.6
		94.5	96.6	100.4	99.0	96.7	
He 501.6	.3115	.2906	.3006	.3128	.3072	.3012	
		−.0209	−.0109	+.0013	−.0043	−.0103	
		−4.8	−2.5	+.3	−1.0	−2.4	−2.1
		93.3	96.5	100.4	98.6	96.7	
Hg 546.1	.1852	.1776	.1801	.1916	.1821	.1830	
		−.0076	−.0051	+.0064	−.0031	−.0022	
		−1.8	−1.2	+1.5	−.7	−.5	−.5
		95.9	97.2	103.5	98.3	98.8	
Hg 578.0	.0906	.0869	.0878	.0920	.0898	.0893	
		−.0037	−.0028	+.0014	−.0008	−.0013	
		−.9	−.6	+.3	−.2	−.3	−.3
He 587.6	.0971	.0952	.0948	.0964	.0995	.0984	
		−.0019	−.0023	−.0007	+.0024	+.0013	
		−.4	−.5	−.2	+.6	+.3	.0
Inc. 620	.1742	.1751	.1764	.1763	.1782	.1772	
		+.0009	+.0022	+.0021	+.0040	+.0030	
		+.2	+.5	+.5	+.9	+.7	+.6
		100.5	101.3	101.2	102.3	101.7	
Inc. 667.8	.4500	.4414	.4465	.4492	.4513	.4521	
		−.0086	−.0035	−.0008	+.0013	+.0021	
		−2.0	−.8	−.2	+.3	+.5	−.4
		98.1	99.2	99.8	100.3	100.5	
Average values of A/A, (per cent):							
Cobalt (mean of 491.6, 501.6, and 546.1 mμ) =		94.6	96.8	101.4	98.6	97.4	----------
Copper (weighting 667.8/620 = 2/1) =		98.9	99.9	100.3	101.0	100.9	----------

[1] For explanation of values, see footnote to Table 22.

Item (4) above illustrates the percentage effect of the deviations on the energy distribution of the source-and-filter combination; item (5) affords a spectrophotometric analysis of the relative copper or cobalt content of the solutions, as qualified below.

A summary of the analyses of these three sets of solutions is given in Table 25. Explanations for the larger discrepancies in Tables 22 to 24 between the standard spectrophotometric data and the values obtained with these solutions become immediately obvious. The following cases should be noted in particular:

1. *Solution 2A.*—The pyridine was yellow and of but 92 per cent strength in the solution. As noted elsewhere, it requires about 6 per

TABLE 25.—*Analysis of chemicals used in tests of spectrophotometric reproducibility*

[Materials obtained by Eastman Kodak Co., solutions prepared at Eastman Kodak Co.; tests of pyridine strength for solutions 1A-2 to 5A-2 made at Eastman Kodak Co.; other tests of pyridine strength made at National Bureau of Standards. Analyses of copper and cobalt salts and solutions made in chemistry division, National Bureau of Standards; in solutions 1A to 5A, 1A-2 to 5A-2, and standard, 3.707 g of copper sulphate and mannite per liter were used, with pyridine content as noted; in solutions 1B to 5B and standard, 27.180 g of copper sulphate and 26.827 g of cobalt ammonium sulphate per liter were used. Spectrophotometric measurements made at National Bureau of Standards]

Solution No.	Copper sulphate (theoretical copper content of salt=25.46 per cent)							Cobalt ammonium sulphate (theoretical cobalt content of salt=14.91 per cent)						Pyridine				Mannite
		Copper content, percentage of theoretical			Percentage of impurity in salt				Cobalt content, percentage of theoretical									
	Material obtained from—	From chemical analysis of salts	From chemical analysis of solutions	From spectrophotometric comparison with standard	Nickel (as Ni)	Iron (as Fe)	Appearance of salt [1]	Material obtained from—	From chemical analysis of salts (includes nickel)	From chemical analysis of B solutions (includes nickel)	From spectrophotometric comparison with standard	Percentage of nickel impurity in salt (as Ni)	Appearance of salt [1]	Material obtained from—	Number of cc per liter	Percentage strength	Appearance of undiluted stock [1]	Material obtained from—
Standard	Mallinckrodt	100.0?	99.7	--------	0.03	0.01	([2])	P. W. R. and B. S.	100.0	--------	--------	0.14	([2])	Baker	30.0	98.4	([2])	B. S.
1A	do	99.3	([3])	([4])	([5])	([5])	O. K.							Mallinckrodt.	30.0	98.1	O. K.	E. K.
1A-2	do	99.3	([3])	99.6	([5])	([5])	O. K.							do	30.2	97.7	(?)	E. K.
1B	do	99.3	97.3	[6] 98.6	([5])	([5])	O. K.	Will	98.6	95.6	94.6	([5])	Wet.					
2A	Merck	99.1	([3])	([4])	([5])	([5])	O. K.							Merck	30.0	[7] 98.9	Yellow.	Merck.
2A-2	do	99.1	([3])	98.7	([5])	([5])	O. K.							do	29.8	99.2	(?)	Merck.
2B	do	99.1	98.9	[5] 99.6	([5])	([5])	O. K.	Will	98.6	97.5	96.8	([5])	Wet.					
3A	Baker	99.1	([3])	([4])	([5])	([5])	O. K.							Baker	30.0	97.0	O. K.	E. K.
3A-2	do	99.1	([3])	99.3	([5])	([5])	O. K.							do	30.0	98.4	Yellow.	E. K.
3B	do	99.1	([3])	[6] 100.0	([5])	([5])	O. K.	Baker	100.0	([3])	101.4	([5])	O. K.					
4A	General chemistry.	99.9	([3])	([4])	([5])	([5])	O. K.							E. K	30.0	97.8	O. K.	E. K.
4A-2	do	99.9	([3])	99.3	([5])	([5])	O. K							do	29.9	98.7	O. K	E K
4B	do	99.9	([3])	[6] 100.7	([5])	([5])	O. K	General chemistry.	100.6	([3])	98.6	([5])	O. K.					
5A	Will	98.7	([3])	([4])	0.14	0.07	Poor.							E. K	30.0	97.8	O. K.	E. K.
5A-2	do	98.7	([3])	98.0	0.14	0.07	Poor.							do	29.9	98.7	O. K.	E. K.
5B	do	98.7	98.9	[6] 100.6	0.14	0.07	Poor.	Will	98.6	99.3	97.4	([5])	Wet.					

[1] See Section IX for details regarding proper appearance of chemicals.
[2] See Section IV, 2 for discussion of materials used for standard solutions.
[3] Not tested.
[4] Values for the A solutions derived from Table 22, in a manner analogous to those given here for the A-2 solutions and derived from Table 23, would not be reliable because of the filtering of the A solutions and the deviation of the pyridine strength from standard. The A-2 solutions were not filtered and the pyridine content was adjusted to the true value.
[5] Not tested. The spectrophotometric data indicate that such possible impurities were of no practical importance in these cases.
[6] These B values do not have the reliability of the A-2 values because of the paucity of the data.
[7] All the values of pyridine strength given in the table were obtained from analysis of undiluted stock. An analysis of the pyridine strength of solution 2A gave a value of but 92.0 per cent strength.

cent error in pyridine strength to equal 1 per cent error in copper strength. Since this solution is about 4 per cent weak (in copper content or its equivalent) the pyridine weakness is not the sole explanation of the discrepancy. However, as noted above, this and the other A solutions (1A to 5A) were filtered. In view of the fact that the filter paper is liable to remove some of the copper (Secs. VII, 2, (*b*) and IX), no further explanation of the copper weakness in these A solutions should be necessary. It is to be emphasized that these data of Table 22 can not be used to show lack of reproducibility in the Davis-Gibson filters.

2. *Solution 3A–2.*—The pyridine used was yellow. Spectrophotometric analysis of this pyridine (Sec. VII, 9 (*d*)) showed that this cause would explain about half of the relatively large deviations in the transmittancies at 404.7mμ and 435.8mμ. The reason for the lack of such deviation in solution 2A is not clear.

3. *Solutions 1A–2, 2A–2, and 3A–2.*—The remaining discrepancy at 404.7mμ with solution 3A–2 and other positive deviations at this wave length would be adequately explained by assuming small amounts of picoline impurity in the pyridine. (Sec. VII, 9 (*b*).) This, however, would not explain the positive deviation at 404.7mμ in the case of the 5A–2 solution as the same pyridine was used here as in the 4A–2 solution.

4. *Solutions 5A, 5A–2, and 5B.*—The deviations at 404.7mμ and to a less extent at 435.8mμ indicate quite certainly a common source of trouble. An analysis of the copper sulphate, common to all three solutions, showed a relatively high proportion of nickel and iron impurities. No study of nickel sulphate in pyridine solution has been made, nor of iron solutions with or without pyridine. The analyzed impurity of nickel would cause barely +1 per cent error at 404.7mμ in the B solution, compared with the observed +2.3 per cent error. (The cobalt component of solution 5B is common to 1B and 2B, which shows that nickel impurity in the cobalt can not contribute appreciably to this 2.3 per cent.) While no certain conclusion can thus be drawn, it is believed that the discrepancies at 404.7mμ, allowing for the usual spectrophotometric uncertainty, could be adequately accounted for by the impurities of this sample of copper sulphate. As noted in Table 25 its general appearance was quite inferior, and the chemical and spectrophotometric analysis (solution 5A–2 and possibly 5A) indicate copper weakness of 1 or 2 per cent. It is reasonably certain that solutions 5A and 5A–2 were not weak in pyridine because the same supplies were used as in solutions 4A and 4A–2, respectively, which showed much smaller deviations.

5. *Solutions 1B, 2B, and 5B.*—The same lot of cobalt ammonium sulphate was used in preparing these solutions. The salt was distinctly wet, adhering to the sides of the bottle. Analysis of the salt showed 98.6 per cent of the theoretical cobalt content. However, chemical analysis of the three solutions (Table 25) gave cobalt contents which account for most of the discrepancies, the chemical and spectrophotometric analyses agreeing to about 1 per cent in two of the three cases.

If the data be rejected for those solutions in which yellow pyridine, wet cobalt ammonium sulphate, or inferior looking copper sulphate were used (in addition to the filtered A solutions), the remaining solutions show a very satisfactory degree of reproducibility. It is

true that the pyridine used in solutions 1A–2 to 5A–2 was tested and adjusted for strength, but it is also true that the total variation in strength in these and the pyridines used in solutions 1A to 5A, as tested in the undiluted stock, varied only from 97.0 to 99.0 per cent. The total variation in all the samples of pyridine examined during this investigation, 17 in number, varied only from 97.0 to 100.0 per cent or ±1.5 per cent approximately from the 98.4 per cent taken as standard. As stated above, 6 per cent variation in pyridine is practically equivalent to a 1 per cent variation in copper content; and 1.5 per cent variation in pyridine causes less than 1 per cent variation in transmittancy at the wave length (587.6mμ, approximately) where this variation is the greatest.

Excluding the solutions just noted leaves Nos. 1A–2, 2A–2, 4A–2, 3B, and 4B as representative of what may be expected when materials are used without chemical or spectrophotometric analysis, but when certain precautions are used in the specifications and choice of the materials and when the A solution is not filtered. The maximum deviations in transmittancy of the observed from the standard computed values are from 1 to 1½ per cent for the B solutions and from 2 to 2½ per cent for the A solutions. Furthermore, little increase in these maximum deviations are noticeable if data from the rejected solutions be included at those wave lengths where the objections stated do not apply—for example, (1) with solutions 1A to 4A from 404.7mμ to 501.6mμ, where loss of copper strength by filtering would be barely noticeable; (2) with solution 3A–2 at the longer wave lengths, where the yellow pyridine has practically no absorption when diluted to 30 cc per liter; (3) with solutions 1B and 2B at 404.7mμ and 435.8mμ and above 546.1mμ, where variations of a few per cent in cobalt content would be of no importance.

It will be noticed that the region where the B solution is most apt to be in error is in the green, including particularly the 491.6mμ and 501.6mμ lines, in which region the A–2 solution is on the average the least in error. Therefore, the combination of the A and B solutions (prepared from properly chosen materials as noted just above) will give for the complete filter a maximum deviation little if any greater than is given by the A solution alone; for example, the combination of either 3B or 4B with either 1A–2, 2A–2, or 4A–2 gives a maximum deviation of 2½ per cent, excepting only in one case (4A–2+4B at 501.6mμ) where a 3½ per cent deviation is obtained.

These maximum deviations compare favorably with those obtained when a completed 2,360° K. filter was examined spectrophotometrically (Sec. VIII, 1), the same materials having been used in preparing the completed filter as were used in deriving the adopted data, and are not certainly outside of the combined spectrophotometric uncertainties. Filters using lower concentrations would be subject to still less uncertainty. With due consideration for the source and appearance of the materials, therefore, as summarized in Section IX, it would seem that filters can readily be prepared whose resulting energy distributions will not deviate from those herein given by more than 2 or 3 per cent.

Two other items of interest regarding the reproducibility of pyridine may be noted.

1. A new supply of pyridine was obtained at the bureau in 1929 from Eastman Kodak Co., titrating 97.7 per cent strength. (It

will be recalled that the standard absorbancy data for solution A′, as illustrated in Tables 4 and 18, were obtained with pyridine from Baker Chemical Co. which was of 98.4 per cent strength.) An A′ solution was prepared, No. 65–24, using the same copper sulphate as used throughout the investigation and with the new pyridine—3 g of copper sulphate, 3 g of mannite, and 30 cc of pyridine per liter of solution. Spectrophotometric measurements on this solution made three weeks after preparation gave, at every wave length tested (Table 18, column 15), values agreeing with the adopted standard values to better than 1 per cent of the transmittancy.

2. The following statement, to which reference has been made in Section IV, 2, appeared in a memorandum (15 Great Britain 1) submitted under date of June 13, 1929, by the subcommittee on colorimetry of the National Illumination Committee of Great Britain:

> Several of the Davis-Gibson filters have been made up at the National Physical Laboratory using c. p. pyridine obtained from commercial dealers without taking any precautions about purity. The properties of these filters agreed so closely with those published by Davis and Gibson that we may conclude, so far as Great Britain is concerned, that pyridine can be obtained commercially of sufficient purity to obviate the need for any elaborate precautions.

9. AUXILIARY DATA

(a) AQUEOUS PYRIDINE SOLUTION OF COBALT SULPHATE

A solution containing 6.000 g (supposedly) of $CoSO_4.7H_2O$, 30.0 cc of pyridine, and distilled water to a total of 1 liter, was prepared and filtered, No. 61–12, and spectrophotometric measurements made from one to three weeks after preparation. Visual, photoelectric, and thermoelectric methods were used, with thicknesses of 20, 40, and 100 mm. The adopted values are illustrated in Figure 26 in comparison with the standard B_2' solution. They are not tabulated because of the uncertainty in the cobalt content. (Sec. IV.)

(b) AQUEOUS α-PICOLINE SOLUTION OF COPPER SULPHATE WITH MANNITE

It was noted in Section IV, 2 that the commonest impurity of importance apt to be present in pyridine is α-picoline, the pyridine used in the standard solutions having a picoline impurity of about 2½ per cent. The effect of the picoline was studied in two ways: (1) As an impurity, 3 g copper-sulphate (with mannite) solutions being prepared, containing, respectively, by volume, 0, 5, and 10 per cent of picoline in the pyridine; (2) without pyridine, a 3 g copper sulphate (with mannite) solution being prepared, containing 35 cc of picoline (the molecular equivalent of 30 cc of pyridine) per liter of aqueous solution.

The results obtained are shown in Figure 27. The most striking differences between the pyridine and the picoline solutions are the general shift of the latter curve toward longer wave lengths and the notable increase in its absorbancy at 404.7mμ.

This marked change in absorbancy at 404.7mμ affords a sensitive method of analysis of the picoline impurity in pyridine. It will be noticed that in the green, yellow, and red the effect of the picoline is not very marked and, such as it is, is very similar to the change produced by a simple weakening of the pyridine strength (fig. 24);

but the change at 404.7mμ produced by the picoline is in striking contrast to the very small change occurring at this wave length when the pyridine content is varied.

On the basis of the measurements at 404.7mμ with the solutions containing 0, 5, and 10 per cent of picoline impurity (fig. 27), the standard solution had 2½ per cent of such impurity, and within the

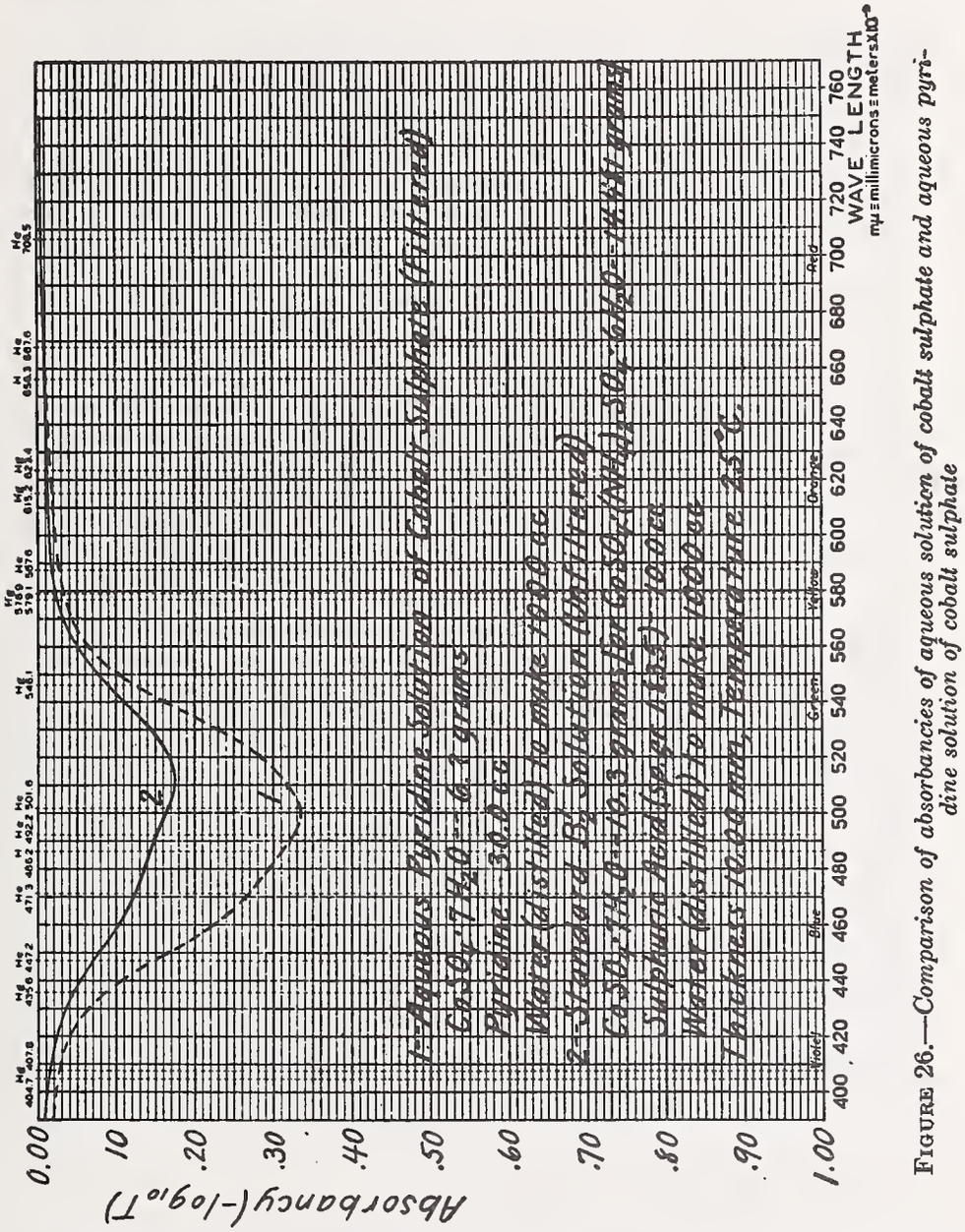


FIGURE 26.—*Comparison of absorbancies of aqueous solution of cobalt sulphate and aqueous pyridine solution of cobalt sulphate*

observational uncertainty this value is checked by consideration of the data from 546.1mμ to 620mμ, inclusive. It may be noted that the changes in absorbancy produced by the small amounts of picoline are not as great proportionately as are indicated by the 100 per cent picoline solution, even when reduced to a 30 cc basis.

The data discussed in Section VII, 8 show that α-picoline impurity in colorless commercial pyridine as great as 5 per cent has not been

518-743 O - 73 - 49

found in this investigation. α-Picoline impurity between 0 and 5 per cent will cause deviations from the standard data of less than 3 per cent of the transmittancy at 404.7mμ and of less than 1 per cent at all wave lengths in the visible spectrum above 430mμ. Under most circumstances its effect, therefore, may be ignored.

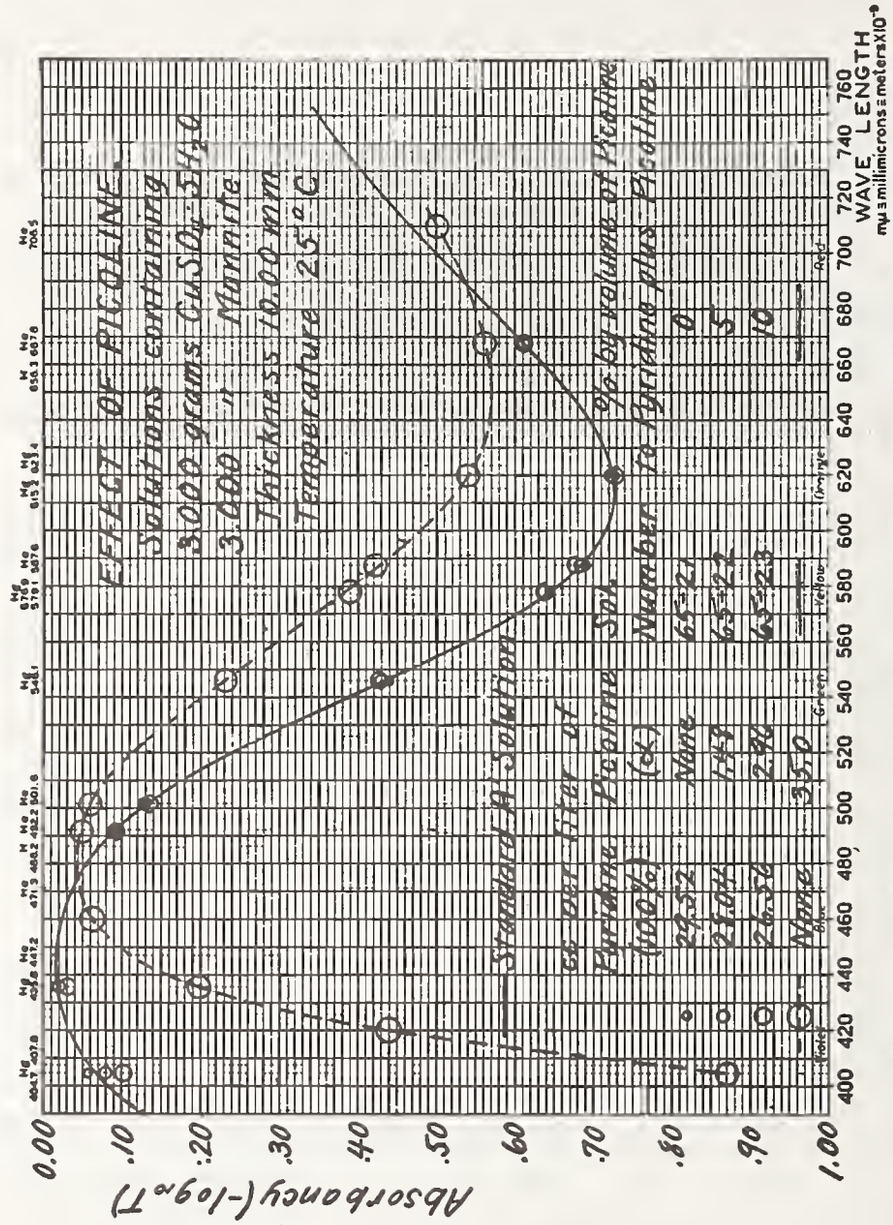


FIGURE 27.—*Effect of α-picoline on the absorbancy of the standard A′ solution*

α-Picoline is one of the next higher homologues of pyridine and may be present in small amounts as an impurity in the pyridine. The pyridine used in obtaining the standard data of this paper contained 2½ per cent of α-picoline.

(c) AQUEOUS SOLUTION OF NICKEL SULPHATE

Nickel sulphate has a well-known characteristic transmittancy curve with maximum transmission in the green and ultra-violet and maximum absorption in the extreme violet and red. As an important impurity in cobalt sulphate and cobalt ammonium sulphate, it is of interest to determine its maximum selectivity in very weak solution; that is, the ratio of its transmittancy in the violet to that in the green.

Accordingly a solution, No. 61–1, containing 1.000 g of $NiSO_4.6H_2O$ per liter was prepared and measurements made visually, photoelectrically, and thermoelectrically. The measurements were not extensive, but sufficient data were obtained to indicate that the ratio of minimum to maximum transmittancy was certainly not less than 0.950.

If nickel were present in cobalt ammonium sulphate in the proportion of nickel to cobalt of 1 to 100, the concentration of nickel sulphate present in the standard B_2' solution would be approximately 0.1 g per liter. The value of the ratio above indicated will in this case probably not be less than 0.995, amounting to one-half per cent selectivity, and for the cobalt as analyzed and used about one-quarter per cent. This small amount is, of course, negligible, so that if the nickel to cobalt (metals) proportion varies from zero to 1 per cent, no important error is introduced.

(d) AQUEOUS SOLUTION OF PYRIDINE

Two samples of pyridine were examined spectrophotometrically. One of these samples was from the stock used throughout most of the investigation, practically colorless and shown by analysis to be of 98.4 per cent strength. (Sec. IV, 2.) In a 10 mm thickness of a 3 per cent aqueous solution (that is, 30 cc per liter, as used in the filters), this shows a maximum variation in transmittancy between 404.7mμ and 700mμ of but 0.2 per cent. This value was derived from measurements made on the full-strength material in a 100 mm thickness. Values for 10 mm are plotted in Figure 28.

The other sample tested was the yellow pyridine used in solution 3A–2. (Sec. VII, 8.) It was examined in a 10 and 20 mm thickness, full strength, and in a 100 mm thickness, 3 per cent strength. Values for 3 per cent strength were computed from the full-strength measurements, assuming Beer's law, and showed good agreement with those obtained directly on the 3 per cent strength solution. The average values are plotted in Figure 28. A maximum selectivity of 3.3 per cent may be noted.

It is obvious that if the strongly colored yellow pyridine be avoided and only that used in preparing A solutions which is practically colorless, showing no more than a trace of yellow, no error of any importance can result from the color of the pyridine.

(e) AQUEOUS SOLUTION OF SULPHURIC ACID

Sulphuric acid is present in the B_1, B_2, and B solutions in the proportion of 10 cc per liter. It was of interest to see if a sulphuric-acid solution of this strength showed any selectivity. A solution was prepared containing 100 cc per liter, or ten times the strength used in the standard solutions. A single set of measurements was made, at 404.7mμ, 435.8mμ, 546.1mμ, 578.0mμ, and 700mμ. Values of transmittancy thus obtained, in a 10 mm cell, differed from unity nowhere more than 1 per cent. The selectivity of the dilute sulphuric acid as used in the B_1, B_2, and B solutions is therefore negligible.

10. USE OF SOLUTIONS AND DATA IN SPECTROPHOTOMETRIC CALIBRATION

The discrepancies which may be obtained when the same or duplicate samples are measured spectrophotometrically on different apparatus has been well illustrated in a recent publication (53). To one who has available but a single instrument or method of measurement it may not be apparent that, along with the excellent precision and

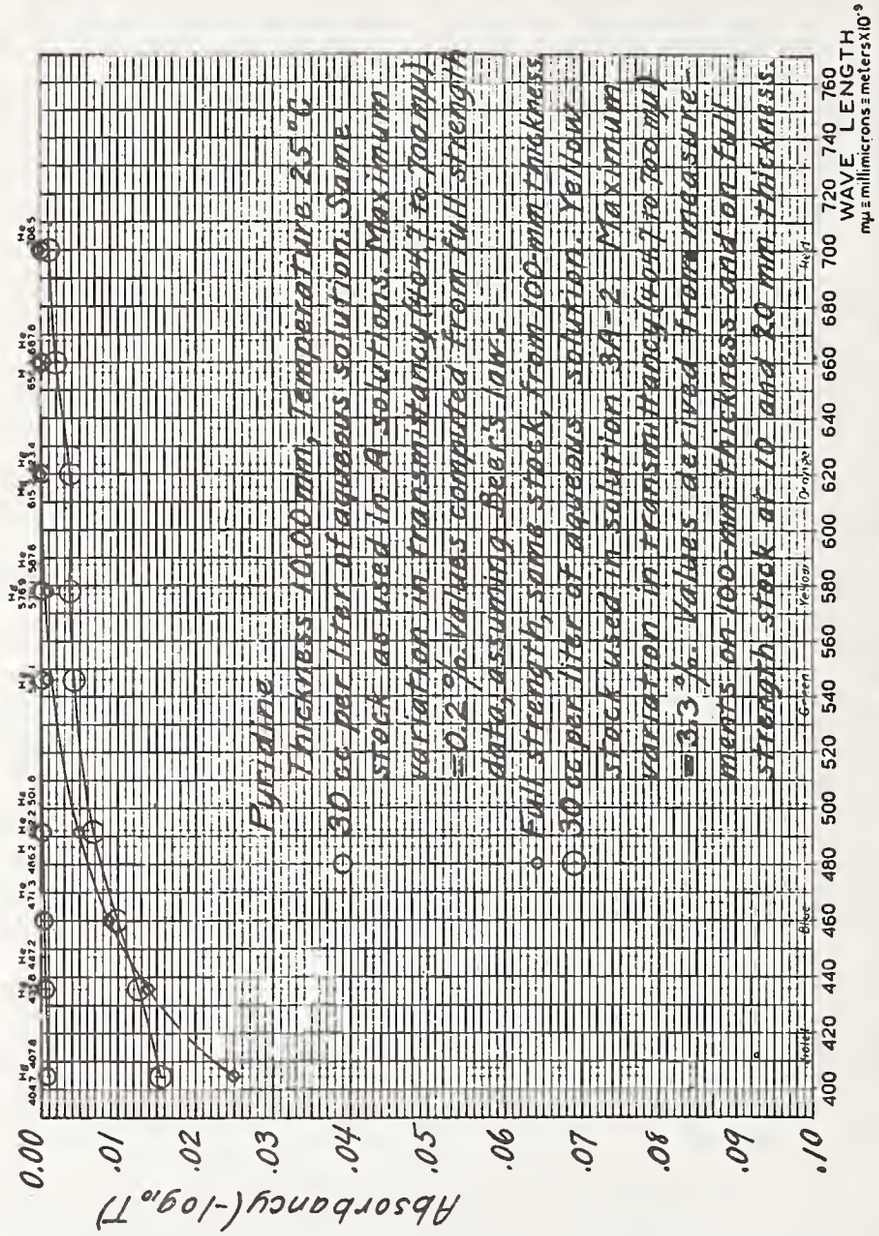


FIGURE 28.—*Absorbancy of pyridine or aqueous solution of pyridine, without copper sulphate or mannite*

Yellow pyridine should not be used in preparing filters.

the high degree of reproducibility which he may be obtaining, serious errors of measurement are, perhaps, entering into his data.

Such errors may result from stray light or energy, from impure spectra caused by the use of too wide slit widths, from inaccurate wave-length calibration, and from temperature changes occurring in

the sample during measurement. The magnitude of such errors depends greatly on the particular sample being measured. In addition the photometric scale—nicol prisms, variable sector, moving wedge, inverse-square law, etc.—may for one reason or another not be functioning in its nominal manner. It may be inaccurate throughout or at certain places only. In measuring transmission samples the introduction of the sample may change the focus or optical distance or displace the beam so that, in effect, a different source is being used when the sample is in the beam than when it is removed. Unless the source is of uniform brightness and of sufficient extent this may cause considerable error. Reflection measurements made on different apparatus will differ because of differing conditions of illumination and observation, and permanent changes may take place in the sample during measurement if subject to excessive radiant energy. Furthermore, practically every instrument or apparatus has troubles or uncertainties peculiar to itself which require constant attention to avoid the introduction of error.

General familiarity with the subject of spectrophotometry is, of course, of great help in securing reliable transmissive and reflective data. The Optical Society of America Report on Spectrophotometry (35) should assist in this respect. Detailed study of one's particular apparatus should, of course, also be made. A third method often used in attempting to verify the reliability of spectrophotometric data is to send colored glass samples to the bureau for calibration. The spectral transmissions reported by the bureau are then taken as standard values, and by measuring the samples on his own apparatus one can readily locate serious errors of measurement.

It is believed that the standard solutions used in the present investigation can be used in this manner. They have been very carefully analyzed and measured as indicated elsewhere in the paper. The series is incomplete for this purpose in that no solution is present which absorbs strongly in the blue. With this exception, by varying the thickness of the solutions, values of absorbancy or transmittancy may be obtained of almost any magnitude at any wave-length region in the visible spectrum.

The data of Tables 4, 5, and 6 only should be used for this purpose. These data do not involve any Beer's law failure. They are standard at 25° C., and any deviation from this temperature may be corrected for by the data of Table 16. The uncertainty of the data is indicated in Section VII, 7. The data apply to 10 mm thicknesses, and the uncertainty in the absorbancy values will increase with the thickness. If one is unable to analyze his chemicals or solutions, the resulting uncertainty must be allowed for. The questions of choice, analysis, and reproducibility of materials is discussed in Sections IV, VII, 8, and IX.

VIII. EXAMINATION AND TRIAL OF COMPLETED FILTERS

In Section VII was given in considerable detail a description of the methods used and the data outlined in the spectrophotometric study of the filter components. This included absorbancy measurements on solutions A′, B_1', B_2', and B′ and density measurements on the double glass cell filled with water, a study of Beer's law over the range of concentrations used, the measurement of time and tempera-

ture effects, and of the effect of variation of pyridine content (on the A′ solution), a discussion of the reliability of the spectrophotometric data, and tests of the spectrophotometric reproducibility of the chemicals. The information therein given has permanent value so far as these solutions are concerned not only as regards their use in filters, but also in connection with their use as standards in spectrophotometric calibration or chemical analysis.

In the present section the influence on the completed filters of the time and temperature effects is computed along with certain other effects which can be most advantageously considered in this manner, and, in addition, various direct experimental checks on the completed filters are described.

1. SPECTROPHOTOMETRIC MEASUREMENTS

Using the same chemicals and the same glass plates as were used in the measurements described in Section VII, three different filters were prepared and tested, viz, 2,360° K. to mean sun, 2,848° K. to Abbot-Priest sun, and 4,000° K. to mean sun, the respective solutions being approximately, 3, 3, and 7 weeks old at the time of the measurements. The concentrations thus used are fairly representative of the two extremes and the medium of the concentrations used in the filters of Series I and II, and the first of these filters is, furthermore, of special interest in its application.

Measurements were made solely on the Koenig-Martens spectrophotometer using the mercury and the incandescent light sources, with other experimental conditions as usual. It was not possible, however, to maintain any temperature control over the filters as had been done with the solutions, because the apparatus for temperature control was not designed to accommodate specimens as large as the filters. The temperature of the air near the filter deviated in extreme cases by 5° or 6° from 25° C., but in such cases measurements were repeated on days when the temperature of the air deviated from the standard in the opposite direction, so that the temperature effect was partially eliminated. No attempt was made to correct the data on the basis of the temperature corrections established in Section VII, 4, as it was uncertain just how truly the temperature of the air near the cell represented that of the filter solutions. Probably on the average the filter temperatures were somewhat in excess of 25° C.

The results of this experimental check are given in Table 26. For each of the three filters are given as a function of wave length, the percentage differences between the transmissions as computed and tabulated in the respective charts and those experimentally determined; that is,

$$\frac{T_s - T_{exp.}}{T_s} \times 100$$

The value given may come from one or two determinations as is the case for the 4,000° K. filter or from one to five determinations as is the case for the other two filters. In averaging two or more determinations at any wave length, equal weights were used.

Deviations less than 1 per cent can not be considered of much significance unless possibly when carrying consistently over a certain

wave-length range for all three filters. The following cases may be noted particularly:

1. *At 404.7mμ and 435.8mμ.*—These values undoubtedly result from lack of temperature control. For example, at 404.7mμ with the 2,360° K. filter, values of transmission were obtained varying from 0.661, air temperature 31° C., to 0.704, air temperature 23.5° C. Furthermore, the differences at 420mμ (single determinations, at air temperatures from 24.5° to 28° C.) of opposite sign in two of the three cases, although of less precision, make practically certain that no important error can be present in this region.

2. *From 480mμ to 510mμ, inclusive.*—The temperature effect is very small in this region. The deviations average about 0.8 per cent, all of the same sign without exception. This may be significant, but can not be considered important.

3. *From 610mμ to 660mμ.*—The deviations average about 1.3 per cent, all of one sign without exception, those for the 2,360° K. filter running over 2 per cent. From 610mμ to 630mμ the temperature effect is small, changing sign near 620mμ.

TABLE 26.—*Comparison of standard computed values of transmission (T_s) as given in the charts with values experimentally determined ($T_{exp.}$), using the same chemicals and glass plates in the latter case as were used in determining the standard values*

Wave length in millimicrons	Values of $\frac{T_s - T_{exp.}}{T_s} \times 100$ for the following filters—				Wave length in millimicrons	Values of $\frac{T_s - T_{exp.}}{T_s} \times 100$ for the following filters—			
	2,360° K. to mean sun	2,848° K. to Abbot-Priest sun	4,000° K. to mean sun	Mean		2,360° K. to mean sun	2,848° K. to Abbot-Priest sun	4,000° K. to mean sun	Mean
Hg 404.7	+1.6	+1.9	+0.9	+1.5	570	−.6	−.4	+.6	−.1
420	−1.0	+.8	−.2	−.1	Hg 578.0	−.6	−.4	+.7	−.1
Hg 435.8	+1.0	+1.1	+.6	+.9	580	+.4	−.9	0	−.2
450	+.3	+.1	+.7	+.4	590	+.8	−1.0	+.7	+.2
460	+1.1	−.5	+.6	+.4	600	+1.2	0	+.4	+.5
470	−.4	+.7	−.2	0	610	+2.5	+1.7	+.7	+1.6
480	+.5	+.9	+.1	+.5	620	+1.8	+.6	+.9	+1.1
490	+.8	+1.2	+.1	+.7	630	+2.5	+1.2	+.7	+1.5
Hg 491.6	+.8	+1.0	+.2	+.7	640	+1.8	+.6	+.8	+1.1
500	+1.3	+1.6	+.6	+1.2	650	+2.8	+.2	+.4	+1.1
510	+.8	+1.3	+.4	+.8	660	+2.5	+.4	+1.2	+1.4
520	−.5	+.3	+.5	+.1	670	+2.9	−.7	+1.2	+1.1
530	0	+.6	+.3	+.3	680	+1.8	+.6	+.4	+.9
540	+.5	+.7	+.1	+.4	690	+1.5	−.8	+.4	+.4
Hg 546.1	−.6	+.3	+.7	+.1	700	−.7	+1.2	+.4	+.3
550	0	+.4	+.1	+.2					
560	+.1	0	−.5	−.1	Mean				+.6

4. *The final mean value.*—This average, +0.6 per cent, illustrates the tendency of the values to be positive.

The discrepancies illustrated in Table 26 may be considered from two standpoints—(*a*) their reasonableness from the spectrophotometric standpoint, assuming no change in the chemicals and no error in preparing the solutions, (*b*) their connection, if any, with the results obtained with solutions prepared from other chemicals.

(*a*) The uncertainty estimated in the standard data for solutions A′ and B′ was in each case from ¼ to ½ per cent of the transmittancy from 404.7 to 667.8mμ, and ½ per cent of the transmission

for the cell filled with water. In addition, must be included the uncertainty resulting from any error in the Beer's law correction. A maximum uncertainty of approximately 1 per cent is therefore to be expected in the standard computed values (T_s) from which the percentage deviations of Table 26 were derived.

Regarding the Beer's law correction factor, it has been shown in Section VII, 6 and Figure 25 that the only region in which any reasonable error could exert an appreciable effect on the computed values was in the red. Such possible errors can not be the principal explanation of the deviations shown in the red region of Table 26, for in that case values for the 2,360° K. and the 4,000° K. filters would be of opposite sign.

The experimentally determined values (T_{exp}) can not be considered reliable in general to better than 1 per cent. With the incandescent light source but one or two determinations were made. As many as five determinations were made in some cases with the mercury lines, but because of the temperature variations the final values did not have the usual reliability.

The possible combined uncertainties of the computed and experimental values are, therefore, sufficient to explain most of the deviations in Table 26; that is, all those less than 2 per cent. In the red, with the 2,360° K. filter, the computed values are probably more reliable than the directly determined values; in fact, a 2 per cent error in these latter would not be surprising. It must be remembered that the transmission in this case is low, varying from about 0.08 at 610mμ to about 0.04 at 700mμ. The 2.5 per cent deviation at 610mμ equals

$$\frac{0.0834-0.0813}{0.0834}\times 100$$

and the 2.8 per cent deviation at 650mμ equals

$$\frac{0.0646-0.0628}{0.0646}\times 100$$

While, therefore, it is not to be expected that the computed values of spectral transmission given on the charts are without error, it would take further work to prove conclusively that they were in error by as much as 1 per cent, for the particular chemicals used in this investigation.

(*b*) The spectrophotometric uncertainty discussed above is somewhat less than may be caused by variations in the materials when obtained from various sources unless due care is taken in the selection of these materials. This question of reproducibility has been discussed in Section VII, 8. It is noted again merely to point out the fact that the tendency for the computed values of transmission to be high from 610mμ to 680mμ, as illustrated in Table 26, was not found in the case of the 2,360° K. solutions measured in the reproducibility test. In that case the deviations tend to be in the opposite direction; so that the net result of all the testing indicates that the values of spectral transmission tabulated on the charts should be a very close representation of what may be expected in the use of the filters.

2. COLORIMETRIC COMPUTATIONS

The effect of the spectrophotometric uncertainties, the time and temperature deviations, etc., on the energy distribution of a source-and-filter combination (E'' on the charts) readily follows from the tables and curves already given in the sections devoted to these subjects. It is of importance also to consider the influence of these uncertainties and deviations on the color of the light transmitted by the filter.

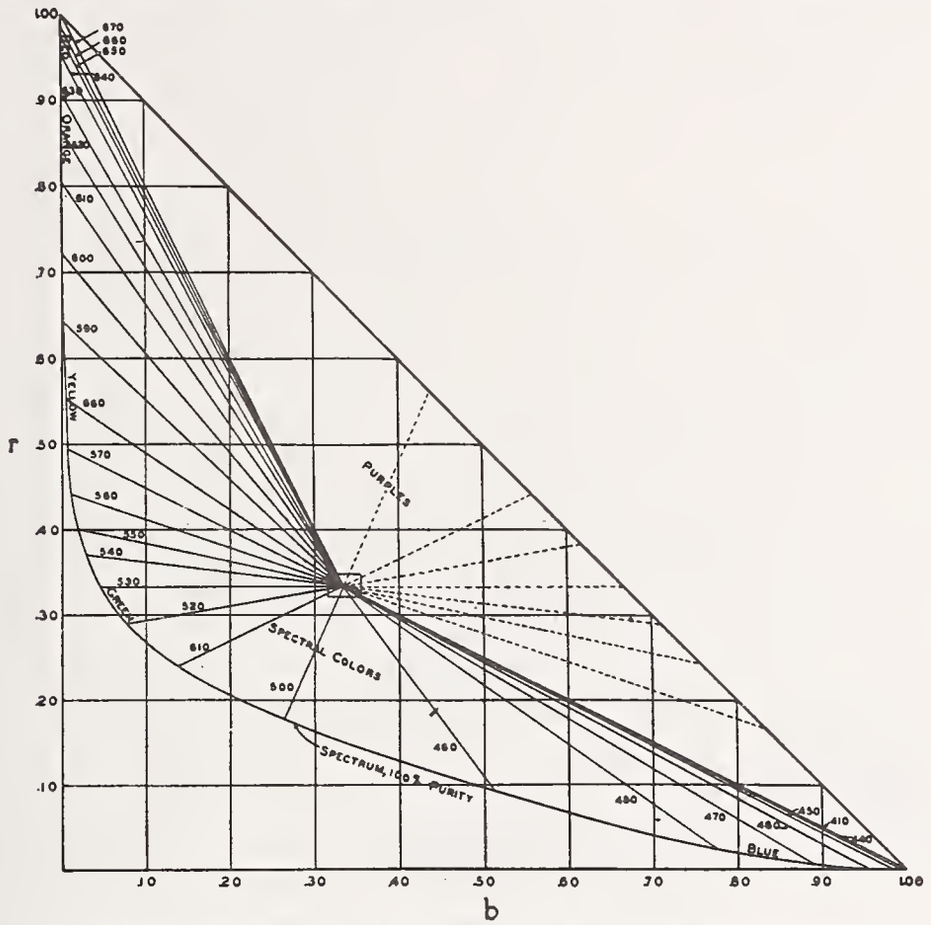


FIGURE 29.—*The complete r–b projection of the trilinear diagram showing the locus of spectrum colors*

Mean sun is the "neutral" stimulus and is represented at the point $r=b=\frac{1}{3}$. Purples (colors whose stimuli are represented by white light minus homogeneous light of some wave length) are represented in the region bounded by the hypotenuse and the radial lines labeled "410" and "670," which represent colors whose dominant wave lengths are found near the ends of the visible spectrum. Colors which may be evoked by a mixture of white light and homogeneous light are represented in the area bounded by the radial lines marked "410" and "670," and the curve marked "spectrum, 100 per cent purity."

The color of a light may be described in terms of the trilinear coordinates (r, g, b), as outlined in Section V, 4, (b). It will be recalled that the values of r, g, and b sum to unity and that the color of mean sun has, by definition, a value of one-third for each coordinate. In Figure 29 is illustrated the r–b diagram on which is also shown the location of the adopted mean sun ($r=b=\frac{1}{3}$), the locus of the spectrum from the red to the violet, and the regions of the spectral colors and purples.

As previously stated, all the filters of Series I, low color temperatures to mean sun, were adjusted to give an accurate color match with the adopted mean sun; that is, to have by computation values of $r=g=b=0.3333$. The uncertainties and deviations in which we are interested may, therefore, be illustrated by plotting the resulting r

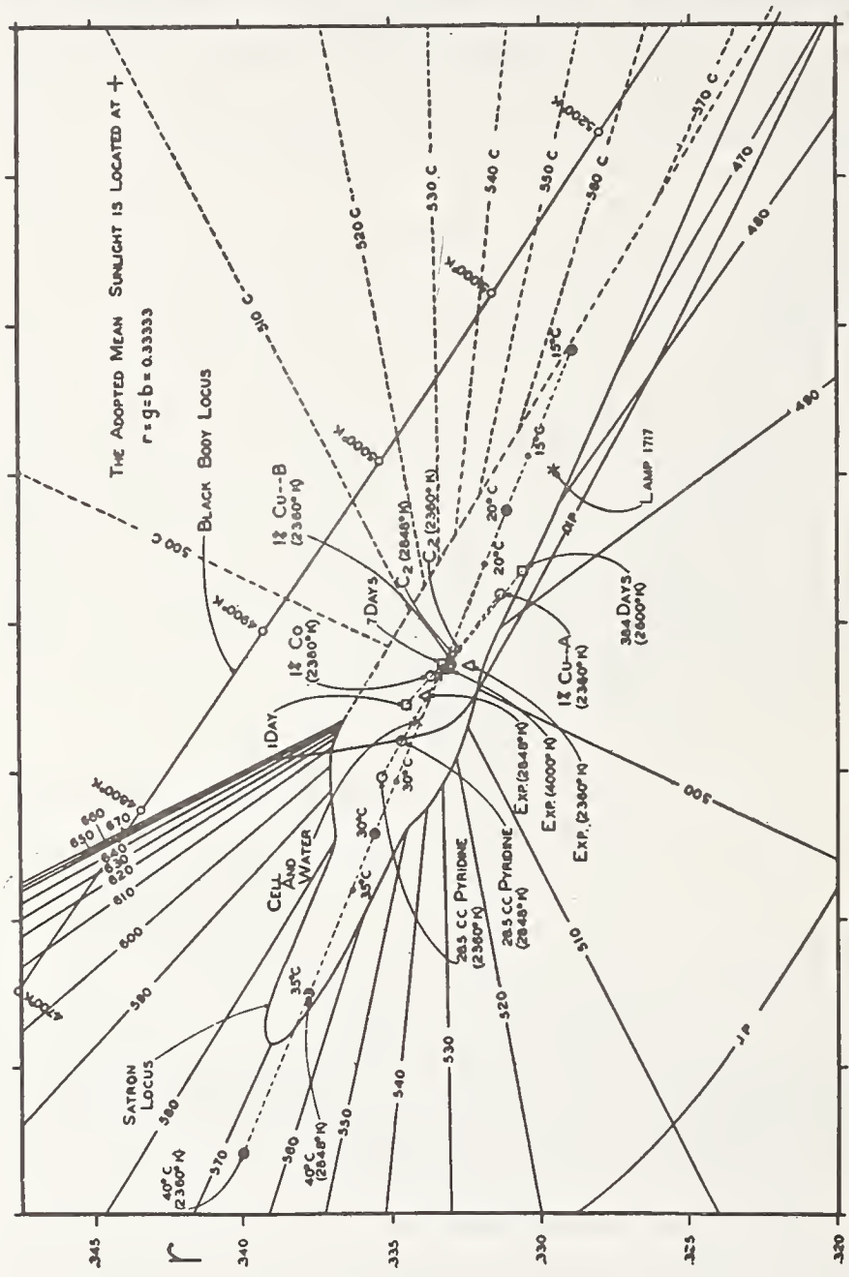


FIGURE 30.—*Enlarged section of the r–b diagram (section shown by small rectangle in fig. 29) showing the satron locus and illustrating the effects of time, temperature, C_2 of Planck's equation, variation of ingredients, and other factors as affecting the colors of Davis-Gibson filters*

and b values on the r–b diagram. The differences from 0.3333 are so small, however, that they can not well be shown to the scale of Figure 29. The small rectangular section near the center of Figure 29 is therefore shown to a greatly enlarged scale in Figure 30.

A question which immediately arises is whether the various small differences illustrated in Figure 30 and discussed in detail in the fol-

lowing sections are visually perceptible. No final answer to this can be given. Data on the "least perceptible difference" in color are very meager. Furthermore, this more or less vaguely defined quantity is so dependent upon the experimental conditions as to make any statement regarding it rather meaningless unless these experimental conditions are fully described.

One investigation has been carried out, however, which furnishes data of particular interest in this connection. This work was by Priest and Brickwedde on "The minimum perceptible colorimetric purity as a function of dominant wave length with sunlight as netural standard" (54). The complete paper has not yet been published. At each wave length they each made two determinations: (1) The least purity perceptible with certainty (P_{max}), (2) the greatest imperceptible purity (P_{min}).

In the present discussion we use a new unit of saturation derived from these data of Priest and Brickwedde and called the "satron." Since there is no previously published definition nor mention of this unit it is necessary to introduce some explanation of it. The idea of such a unit is predicated upon the concept of saturation as a psychological entity (4, p. 534) and the definition of purity as a ratio of brightnesses as defined by Priest (55, p. 175). Saturation is regarded as the minimum number of equally perceptible increments in purity between white and the color in question. The so-called just perceptible or least perceptible increment in purity is taken as the most convenient case of "equally perceptible." It is assumed that, for colors of very low saturation, and of constant dominant wave length, equal increments in purity are equally perceptible. It is, however, known (from the work of Priest and Brickwedde and others) that, for different wave lengths, equal increments in purity may be very far from being equally perceptible. It is indeed on this account that we find the data of Priest and Brickwedde particularly instructive in the discussion of the approximation to white which we now undertake. The term "satron" was tentatively proposed and defined by Priest in the spring of 1927 for the purpose of facilitating local discussion relating to colors of low saturation among those concerned with this subject at the National Bureau of Standards. The definition written by Priest at that time is as follows:

> For colors of low saturation $\left(\frac{P}{P'_o}<20\right)$ one satron is the increment in saturation due to an increment in purity equal to P'_o, where P'_o, a function of dominant wave length, is taken as $\frac{P_{max}+P_{min}}{2}$ from the mean values of Priest and Brickwedde.

The effect of this definition is that a color having a saturation of one satron would be hardly distinguishable from white.

The locus of such colors on the r–b diagram has been computed and plotted for us by D. B. Judd as shown in Figure 30. Colors whose trilinear coordinates locate points lying on this curve will be distinguished from white only by keen observers making careful observations—and even then usually with some doubt and hesitancy. Obviously, the colors corresponding to points lying within the area bounded by this curve will be distinguishable from white with still greater difficulty—and the more so the nearer they approach the center ($r=g=b=⅓$). Roughly it may be said that the area bounded

by this curve represents colors which may for most practical purposes be considered indistinguishable from white.

The experimental measurements on which its determination is based did not include the purples, so that the boundary on the purple side has been more or less arbitrarily estimated. It should be further noted that the white or "neutral" stimulus used in the experimental measurements was Abbot-Priest sunlight (Sec. II, 1), and that the present authors have assumed that the same relative values would have been obtained if the adopted mean sun had been used as the "neutral" stimulus. For the purposes of the present illustration this assumption is undoubtedly justified.

As already noted, the absolute values of the so-called least perceptible difference are subject to variations with experimental conditions, such as brightness, field size, observer, and observing technique. The relative values, and, therefore, the shape of the satron locus as shown in Figure 30, would probably, however, not be greatly affected, and it is a fortunate coincidence that the deviations to be considered are of such a nature as to lie more or less along the major rather than the minor axis of this region.

A record of the particular conditions under which Priest and Brickwedde made their observations has been published (54). Moreover, Priest has subsequently made other independent experiments which lead to the following previously unpublished conclusion which he has authorized us to quote:

> For nearly white mixtures of Prussian blue with zinc oxide, comparing areas of about 3 or 4 cm² in immediate juxtaposition in diffuse daylight, under ordinary conditions at a viewing distance of 30 to 40 cm, it has been found, by fair tests on 16 observers, that a saturation of one satron is clearly recognized by the keenest observers, recognized with some doubt and hesitancy (but correctly) by the great majority, and is quite imperceptible only to a very small minority. The perceptibility appears to be of the same order as the perceptibility of a 1 per cent difference in brightness under the best photometric conditions.

(a) SPECTROPHOTOMETRIC UNCERTAINTY

Consider first the spectrophotometric discrepancies given in Table 26. It was of interest to compute the values of r, g, and b that result if the experimental data are assumed to be wholly correct and the values given in Table 26 to be due entirely to error in the standard data. The values of r, g, and b thus obtained are as follows: For the 2,360° K. filter $r=0.33236$, $g=0.33407$, $b=0.33357$; for the 2,848° K. filter, $r=0.33384$, $g=0.33360$, $b=0.33256$ (relative to Abbot-Priest sun as neutral); for the 4,000° K. filter, $r=0.33305$, $g=0.33346$, $b=0.33349$. The results are shown in Figure 30 by the three points plotted as "Exp (2,360° K.)," "Exp (2,848° K.)," and "Exp (4,000°K.)." (The 2,848° K. point is plotted as if Abbot-Priest sun were located at $r=b=0.3333$, since this filter was designed to give Abbot-Priest sun.)

If it be remembered that the standard values are, in fact, fully as reliable as the experimentally determined values (Sec. VIII, 1), it may be assumed that the spectrophotometric uncertainty in the standard data would cause deviations of not more than half those shown in Figure 30.

(b) DEVIATIONS FROM THEORETICAL CONCENTRATIONS

In view of the small deviations of the actual content from the theoretical content which may occur in the copper and cobalt salts used in this investigation (Table 25) it is of considerable importance to know how such variations affect the color of the filters. For this purpose the trilinear coordinates were computed for the colors evoked by the 2,360° K.-to-mean-sun filter—(1) for a 1 per cent increase in the copper content of solution A, keeping the concentrations of solution B unchanged, (2) for a 1 per cent increase in the copper content of solution B, keeping solution A and the cobalt content of solution B unchanged, (3) for a 1 per cent increase in the cobalt content of solution B, keeping the copper content of solutions A and B unchanged. Values of r, g, and b were obtained in the three cases as follows:

(1) $r=0.33135$, $g=0.33267$, $b=0.33598$;
(2) $r=0.33301$, $g=0.33340$, $b=0.33360$;
(3) $r=0.33368$, $g=0.33311$, $b=0.33321$.

These values are plotted in Figure 30, the "1%" points. They indicate closely the directions of change caused by small variations of the concentrations in any of the filters. The magnitudes of such variations will be approximately proportional to the concentrations. For most of the filters (Tables 8 and 12) the concentrations are less than for the 2,360° K.-to-mean-sun filter, and the deviations caused by a 1 per cent change in the copper and cobalt salts will be proportionately less.

It is obvious that small changes or uncertainties in the concentrations are of possible importance only in the case of the copper sulphate in solution A. Since the deviations from the theoretical copper content are rarely more than a fraction of a per cent, the results plotted in Figure 30 indicate that no perceptible differences in the color of the filters may be expected to result from uncertainties in the theoretical contents of the salts used.

(c) DEVIATION FROM THEORETICAL PYRIDINE CONTENT

It will be recalled that the pyridine used in deriving the standard absorbancy data of Table 4 had a strength of 98.4 per cent and that the total variation in the strengths of 17 samples of commercial pyridine examined in this investigation varied from 97.0 to 100.0 per cent or approximately 1½ per cent from the standard. In view of the questions that have been raised regarding the use of pyridine (41, 56) it is of special interest to compute the effect on the color of the filters caused by deviations of the pyridine strength from the standard. The effect of large deviations on the spectral absorbancy has been illustrated in Figures 8 and 24.

Values of r, g, and b were computed for both the 2,360° and 2,848° K.-to-mean-sun filters with 28.5 cc of pyridine instead of the usual 30.0 cc and with the concentrations of the salts unchanged. Values were obtained as follows: For the 2,360° K. filter, $r=0.33531$, $g=0.33484$, $b=0.32984$; for the 2,848° K. filter, $r=0.33463$, $g=0.33431$ $b=0.33106$. The results are shown in Figure 30 by the points labeled "28.5 cc pyridine." It is apparent that the deviations caused by a given change in pyridine strength are approximately proportional to the copper concentration in solution A, these being 3.707 and 2.445 g respectively. It is also undoubtedly true that such deviations would

be approximately proportional to the deviations in the pyridine strength.

It should be remembered that the results shown in Figure 30 are for a 5 per cent variation in pyridine strength, which is over three times as great a variation from the standard as was found in any of the 17 samples of pyridine examined. It may be shown from Figure 30 that the 1½ per cent variation from the standard, which is about the maximum to be expected in commercial pyridine, is of no greater importance colorimetrically than one-fourth of 1 per cent variation in the copper content of the A solution, or, conversely, that it would require more than a 6 per cent variation in pyridine strength to be colorimetrically equivalent to a 1 per cent variation in the copper content of solution A. (These values are derived as follows: The relative distances from the point $r=g=b=⅓$ to the points showing the 5 per cent variation in pyridine and the 1 per cent variation in the copper of the A solution are, respectively, in the proportion of 5.65:4.65. If the lines connecting these points and the point $r=g=b=⅓$ be extended to intersect the satron locus, we may estimate to what fraction of a satron these deviations are proportional. The values are $5.65/12.40=0.455$ for the pyridine, and $4.65/8.00=0.581$ for the copper. From this we find that 5 per cent variation in pyridine is $0.455/0.581=0.783$ as perceptible as 1 per cent variation in copper in the A solution, and that it, therefore, requires $5/0.783=6.4$ times as great a percentage change in the strength of pyridine to produce a perceptible color difference as is required in the copper sulphate in solution A.)

This illustration should end all opposition to the use of commercial pyridine arising from uncertainties regarding its strength. It may again be noted that the commercial pyridine obtainable in Great Britain has been found satisfactory for use in the filters. (See quotation at end of Section VII, 8.)

(d) DEVIATION FROM STANDARD TEMPERATURE OF 25° C.

In Section VII, 4 are given data showing the change in absorbancy for each of the three component solutions as the temperature is raised from 25° to 40° C., it being further shown that this change in absorbancy is proportional to the concentration of the solution. On the basis of these data of Table 16, the colors of the 2,360° and 2,848° K.-to-mean-sun filters, in combination with their respective sources, have been computed at 40° C. and are plotted in Figure 30. The values of r, g, and b thus obtained are: For the 2,360° K. filter, $r=0.33999$, $g=0.34286$, $b=0.31715$; for the 2,848° K. filter, $r=0.33778$, $g=0.33986$, $b=0.32236$. Here again the deviations are approximately proportional to the copper concentration of solution A.

The values given at 35° and 30° C. are estimated by taking 2/3 and 1/3, respectively, of the distance along the straight line connecting the 25° C. (standard white) point and the 40° C. point; the values at 20° and 15° C. are likewise estimated by reversing the sign of the deviations shown at 30° and 35° C. No possible uncertainty in such estimations can be of any practical importance in this connection.

The importance of the temperature deviations shown in Figure 30 and the tolerance to be allowed will, of course, depend on the particular use being made of the filters. It would seem that a ±5° C. tolerance from the standard 25° C. temperature would be perfectly safe in most

colorimetric, photometric, and sensitometric work, and in many cases still larger tolerances could be permitted. It should be noted, however, that these solutions are powerful absorbers of radiant energy, due to the action of both the water and the copper sulphate, and the use of the filters in close proximity to lamps may be attended by relatively rapid increase in temperature of the solutions.

Since most of the temperature change is in the A solution, it is best to place the filter so that the A solution is on the side opposite the light source. A surer means of protection, where necessary, would be to place an auxiliary water cell between the filter and the source. A single or double water cell made with borosilicate crown glass windows, such as used in the filters, should have no perceptible effect on the color, as is illustrated in Figure 30, where the color of the double cell filled with water, computed from the standard spectral transmission data of Table 7, is shown by the point labeled "water cell." The effect of the water cell on the energy distribution is shown by its spectral transmission. In cases where two beams are being used, water cells may be placed in both beams, if considered desirable.

(e) CHANGE WITH TIME

It was shown in Section VII, 3 that the B′ solution showed no certain change over a year's time, but that solution A′ showed small but definite changes as illustrated in Figure 21. A 2,600° K.-to-mean-sun filter has concentrations close to those of the standard solutions, and it has accordingly been chosen for illustrating the aging effect. No tests have been made as to how this effect may vary with the concentration. In the computations under consideration the values shown in Figure 21 were multiplied by 3.024/3.000, the ratio of the copper content of the 2,600° K. A solution to that of the A′ solution.

The results are shown in Figure 30, the plotted values being labeled "1 day," "7 days," and "384 days." The trilinear coordinates for these three cases are: For 1 day, $r=0.33449$, $g=0.33322$, $b=0.33228$; for 7 days, $r=0.33328$, $g=0.33309$, $b=0.33362$; for 384 days, $r=0.33060$, $g=0.33266$, $b=0.33675$. Values for 24 days were not computed, as they would be practically identical with the standard values.

It is apparent that the color of any of the filters would not change perceptibly from a few days to several weeks, perhaps two or three months after preparation, but that solutions or filters a year or more old might be perceptibly different in color (but only so under the best colorimetric conditions) from newly prepared solutions or filters.

(f) CHANGE WITH C_2 IN PLANCK'S EQUATION

In Planck's equation (Sec. III, 3)

$$E_\lambda = \frac{C_1 \lambda^{-5}}{e^{\frac{C_2}{\lambda\theta}} - 1}$$

the relative values of E_λ will obviously depend upon the value of C_2 (but not upon the value of C_1) which is used. It has been stated that a value of $C_2 = 14{,}350\mu°$, as used in B. S. Misc. Publication No.

56, enters into all the Planckian energy distributions of Table 2 as used throughout the present investigation. A value of 14,330 (1.433 cm°) has been used in the International Critical Tables. It was of interest to compute the effect on the color of the filters caused by this change in the value of C_2.

Values of E_λ from 2,000 to 3,120° K., based on values of $C_2 = 14{,}330$, have been recently published by Skogland (57), and values for 2,360° and 2,848° K. were taken from his tables and carried through in the regular computation for r, g, and b, in the case of the respective filters of Series I. The following values were obtained: For the 2,360° K. filter, $r = 0.33278$, $g = 0.33309$, $b = 0.33413$; for the 2,848° K. filter, $r = 0.33292$, $g = 0.33312$, $b = 0.33395$. These values are indicated in Figure 30 by the designations "C_2 (2,360°)" and "C_2 (2,848°)."

It is more or less obvious, and has been pointed out in the appendix to M56, that differences in C_2 may be counterbalanced by differences in θ. A change of 20 in C_2 is approximately equivalent to 3° K. at 2,360° K. and 4° K. at 2,848° K. To get the same respective energy distributions with $C_2 = 14{,}330$ as were obtained with $C_2 = 14{,}350$, values of θ equal to 2.357° and 2,844° K. must be used.

If the filters of Series I had been designed with energy distributions derived on the basis of $C_2 = 14{,}330$ instead of 14,350, slightly lower concentrations of copper sulphate in the A solutions would have been necessary to secure perfect color match with the adopted mean sun. From the equivalents of C_2 given above and from Table 8, it may be estimated that such perfect color match would have been obtained by using approximately 0.01 g less of copper sulphate than has been used in this series of filters. In the filters of Series II the change required would be even less, inasmuch as both the source and the ideal are computed from Planck's equation, and any change in C_2 affects both distributions in a similar manner.

It is apparent from Figure 30 that the error introduced by using $C_2 = 14{,}350$ instead of 14,330 (assuming that the latter is really correct) in the design of the filters is of the same order of magnitude as the spectrophotometric uncertainty in the data. It is probably less than is caused by the differences between the energy distributions as given by Planck's equation and as emitted by incandescent lamps. (Sec. III, 3.) It is much less than is caused by a 5° C. deviation from the standard 25° C. temperature. In fact, probably no case will arise in the use of the filters where the uncertainty in the value of C_2 will make the slightest practical difference.

(g) EXPERIMENTALLY DETERMINED ENERGY

It was shown in Section III, 3 (fig. 7) how the energy distribution of a particular Mazda C lamp as determined experimentally differed from the energy distribution as given by Planck's equation for the same color temperature (2,848° K.). The effect on the color computed in the case of a 2,848° K.-to-mean-sun filter, using the experimentally determined values as illustrated in Figure 7 instead of as given by Planck's equation for 2,848° K., is shown in Figure 30 by the point plotted at "lamp 1717." The values of r, g, and b obtained were, respectively, 0.32954, 0.32935, and 0.34111. The shift toward the blue is caused by the excess of energy at the shorter wave lengths in the experimentally determined values.

While it was of interest to compute the magnitude of the difference caused by the two energy distributions noted above, it is not believed that any such difference can be assumed in the case of incandescent tungsten lamps in general. In the first place, it is not felt that the experimentally determined values in the blue and violet can be entirely free from error caused by stray energy. Such stray energy from longer wave lengths in the visible and infra-red would undoubtedly cause an increasing percentage error toward the shorter wave lengths where the true spectral energy is relatively very low. Furthermore, it may be noted in this connection that Forsythe found (58) that "tungsten radiates in such a manner that when its temperature is set to give the same red-blue ratio as the black body for a particular temperature the green part of the tungsten spectrum is relatively brighter than that of the black body. This difference is, however, small, being less than one-half per cent at a color temperature of 1,600° K. and increasing to about 1 per cent at a color temperature of 2,600° K." The wave lengths of the red, green, and blue light used in these tests were about 660mμ, 540mμ, and 470mμ, respectively. It is believed that, if the same comparison were made in the case of the experimental energy distribution of Figure 7 and of a Planckian radiator at 2,848° K., the experimental data would be low in the green relative to the red-blue values instead of high.

It would appear from the above evidence that the differences between the spectral energy distributions from a black body and from a tungsten lamp at the same color temperature are of no practical importance; and the reality of this small difference in the particular case illustrated in Figure 7 is subject to considerable question.

3. EFFECT OF PYRIDINE CONTENT ON THE RESULTING ENERGY DISTRIBUTION

In Section VII, 5 was illustrated the effect of the variation of the pyridine content on the spectral absorbancy of the A solution; and in Section VIII, 2 (*c*) was illustrated the effect on the color of the source-and-filter combinations of small variations of pyridine content, such as might possibly occur as a result of the use of pyridine of subnormal strength. It remains to demonstrate the statements previously made that a pyridine content of approximately 30 cc per liter, such as used in all the filters, gives the most suitable energy distribution which it is possible to obtain.

For this purpose two filters have been computed for converting 2,360° K. to mean sun, following the methods described in Section V, 4 (*a*) and Table 9, using in one case a pyridine content of 10 cc per liter and in the other a content of 100 cc per liter, and in both cases adjusting the concentrations of all three components so as to obtain the best possible energy match with mean sun.

The results are illustrated in Figure 31, where the superiority of the 30 cc content is clearly apparent. The concentrations used are therein given. The effects are readily understandable if it be remembered (figs. 8 and 24) that increasing the pyridine content increases the absorbancy over most of the spectral range and also shifts the maximum absorbancy toward shorter wave lengths. At 10 cc of pyridine the maximum absorbancy is too close to that of the copper sulphate (B_1), at 100 cc it is too close to that of the cobalt (B_2). In

518-743 O - 73 - 50

both cases it is impossible to minimize the deviations from the ideal as satisfactorily as can be done with 30 cc of pyridine, where the region of maximum absorbancy is properly spaced between those of the B_1 and B_2 components.

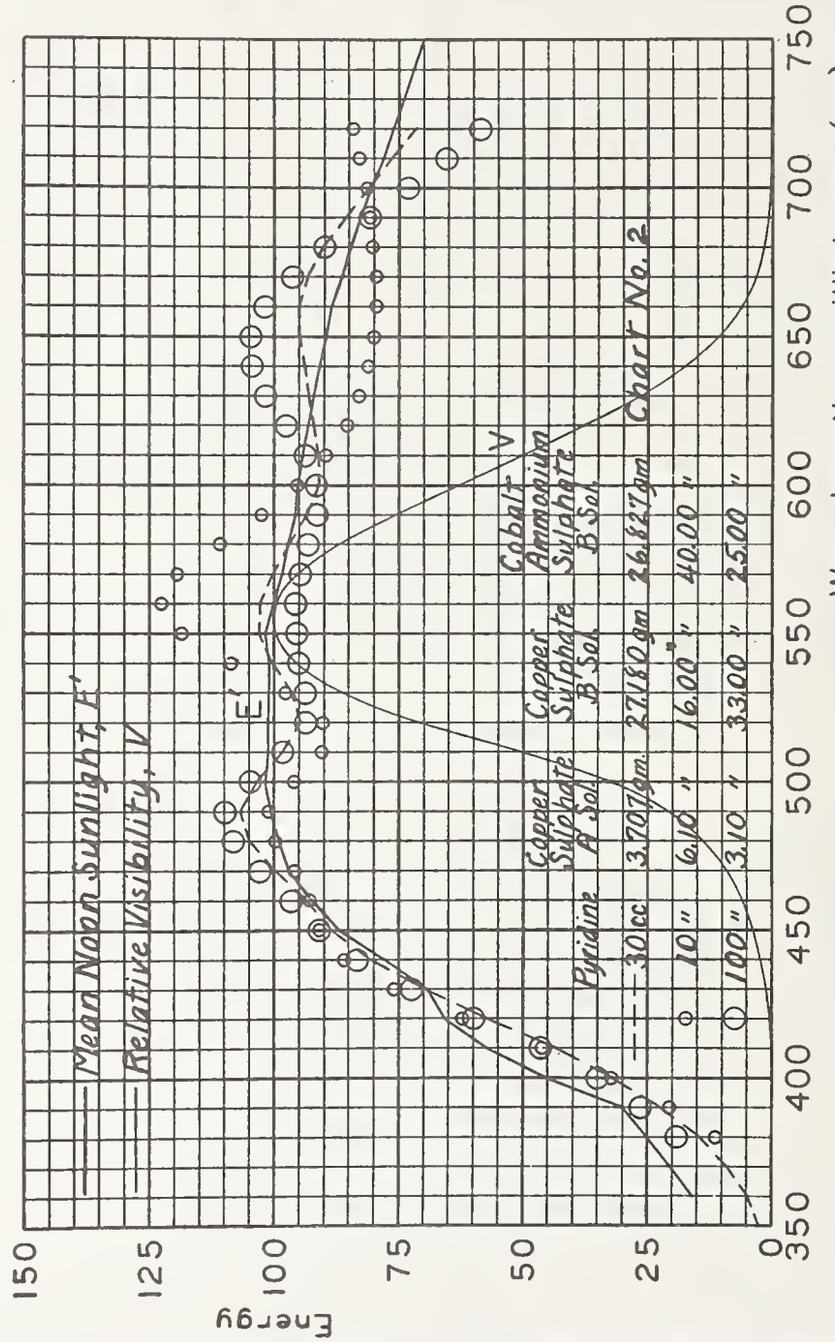


FIGURE 31.—*Effect on final energy distribution caused by using varying amounts of pyridine*

Note the marked superiority in the energy distribution match obtainable with 30 cc of pyridine as compared with 10 or 100 cc.

It is not claimed that slight improvement might not possibly still be made at some other pyridine content—for example, at 29 or 31 cc per liter—but it is obvious that no such improvement would be of any practical importance. As a matter of fact, a content of 29 cc was

the exact value derived from the measurements and computations that were made at the time the value of 30 cc was adopted.

4. PRELIMINARY TESTS OF THE COLOR OF NOON SUNLIGHT AND DAYLIGHT AT WASHINGTON

The measurements described were undertaken both as a check on the color of certain of the filters and as a test of the applicability of the filters for precise colorimetric determinations of sunlight and daylight. The general methods possible for this purpose have been outlined in Section VI, 2 (*b*).

The tests under consideration were made in a south room, the window being raised during observations so that neither the skylight nor the sunlight passed through the glass of the window. A Martens photometer was placed about 2 feet from the window. For the overcast sky, the window was opened wide so that a considerable aperture of the southern sky was used. The horizon was screened off. For the noon-sun observations the window aperture was covered with cardboard except for a small opening which transmitted the sunlight beam to the photometer.

A standard lamp (B. S. 3256) was operated at 2,848° K. This is the same lamp as that noted in the next section, and the 2,848° K. point was checked before beginning the tests in September. The light from this lamp transmitted by the Davis-Gibson filter was incident on a white diffusing surface, light from which entered one field of the Martens photometer. Light from the sun or sky was diffused into the other field of the photometer by another surface of the same white material. The artificial light was incident on the diffusing white material at approximately the same angle as the natural light.

The brightnesses in the two fields of the photometer were always equated when the colors were being compared. In the September measurements the Lovibond glasses necessary for exact color match were inserted in the sunlight or skylight beam; in the December measurements the glasses were inserted in the beam from the source and filter combination. The latter is a more accurate method, although in the method proposed in Section VI, 2 (*b*), where the light source is varied and only Lovibond red glasses of small denomination are necessary, the glass may be inserted in either beam as required with practically equal accuracy.

The results of this preliminary testing are shown in Figure 32. In addition to these experimental results, there are illustrated, in terms of the r and g coordinates used in plotting, the following data: (1) A portion of the black-body locus; (2) the sunlight locus as given by Abbot's 1917 data. Mean sun, lying on this locus, is taken as neutral, $r=g=b=1/3$; (3) the sunlight locus as derived from Abbot's 1923 data; (4) Abbot-Priest sunlight; (5) the locus obtained by varying the color temperature of the source from 2,300° to 3,000° K. in the case of the filter designed to convert 2,848° K. to sun outside atmosphere (1917 data), chart 32; and (6) the satron locus as determined by Priest and Brickwedde and reduced by Judd. (Sec. VIII, 2, and fig. 30.) In the illustration the irregular shaped figure surrounding the point representing the color of mean sunlight ($r=g=1/3$) represents this locus.

The experimental data plotted include (1) the individual observations on overcast southern sky in September, 1928; (2) the same for December, 1928; and (3) the means of the observations on direct sunlight from 11.30 to 12.30 on several days in December, 1928. The data on sunlight taken in September, 1928, are not plotted.

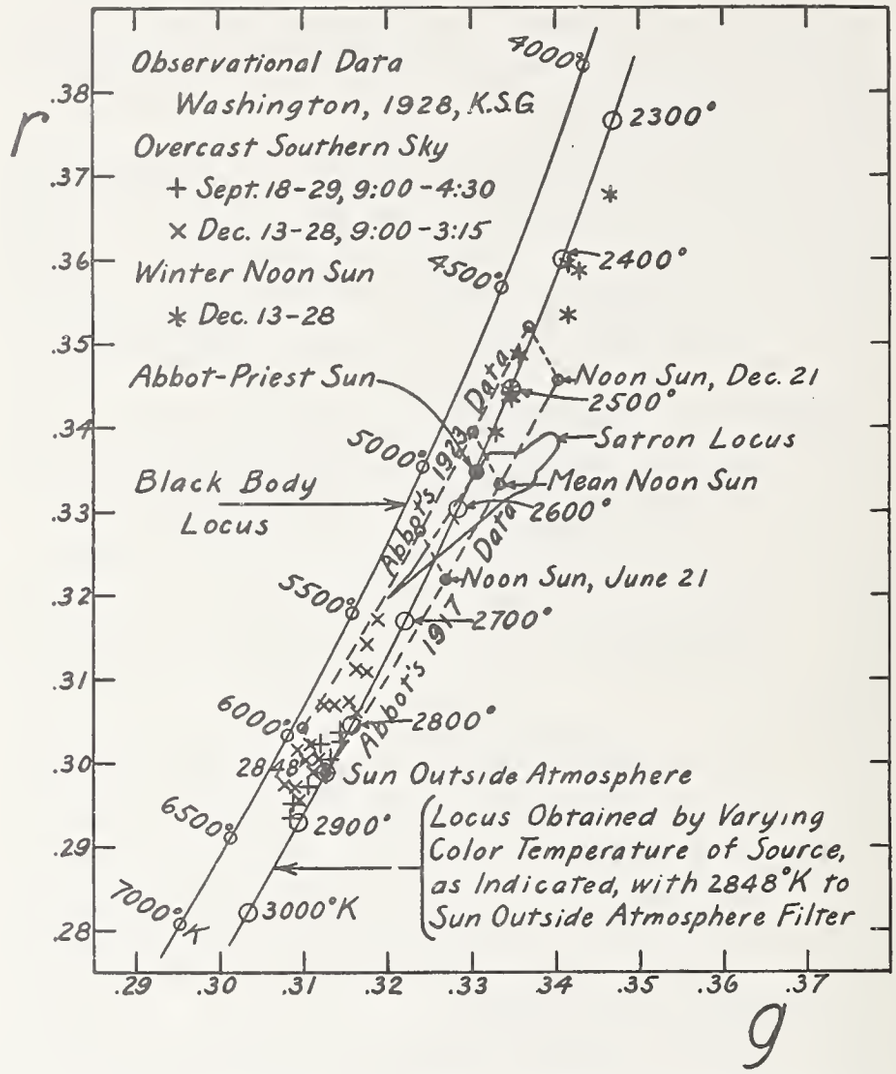


FIGURE 32.—*Section of the r–g diagram showing colorimetric positions of winter noon sunlight and of overcast sky as experimentally observed by one of the authors, using certain of the filters and Lovibond glasses (see Sec. VIII, 4 for details)*

The colorimetric positions of sunlight designated as Abbot's 1917 data were computed from the energy values of Table 1. The 1923 data for sunlight outside the earth's atmosphere were derived from Abbot's publisbed data (Bib. ref., 42, Table 10, last column); the other 1923 data were derived by the authors as follows: By dividing the 1923 energy values for sun outside atmosphere by the corresponding 1917 values, wave length by wave length, correction factors were obtained which were applied to the other 1917 energy distributions to derive the corresponding 1923 distributions.

They were rendered questionable because of insufficient attention to a certain stray-light effect which was noticed at the start of the December measurements and thereupon eliminated. The results obtained, so far as this error can be estimated for the September

sunlight measurements, would be consistent with the results shown in Figure 32. The error did not apply, at least to any important extent, to the data on overcast sky taken in September, 1928.

These data are presented more to illustrate the applicability of the filters for such measurements than as any important contribution toward the color of sunlight and daylight. Nevertheless, they are considered fairly reliable as far as they go. It is interesting to note that all the data obtained fall between the two sunlight loci as derived from Abbot's 1917 and 1923 data. The color temperature of practically all the data on overcast sky comes apparently between 6,000° and 7,000° K., confirming these figures as found by other observers. However, it may be noted that the colors observed are not exactly black-body colors, but are relatively slightly greenish. Inasmuch as Abbot's sun outside atmosphere (1923 data) is itself slightly green relative to the black body and since such true absorption as may take place in the atmosphere because of the water vapor would accentuate this greenness, it is believed the results obtained are exactly as to be expected.

The data on noon sunlight obtained in December, 1928, average nearer Abbot's 1923 data than the 1917 data, although lying between them. They tend to confirm Abbot-Priest sunlight as an excellent colorimetric representation of average noon sunlight, without con-

FIGURE 33.—*Schematic diagram showing photometer used to test the color match of the lamp-and-filter combinations*

See text for explanation.

tradicting in any way the authors' contention that their adopted mean-sun energy distribution is a more suitable choice as a standard for photographic sensitometry than either 5,000° K. or Abbot-Priest sun. The data further indicate the particular suitability of the second method proposed in Section VI, 2 (*b*), for the study of daylight and noon sunlight, viz, the use of the filter for converting 2,848° K. to sun outside atmosphere (1917 data), chart 32, with a source calibrated for color temperature from 2,300° to 3,000° K. and Lovibond red glasses of low denomination (for example, maker's numerals 0.1, 0.2, and 0.4).

5. DETERMINATION OF COLOR TEMPERATURES OF INCANDESCENT LIGHTS

The following test of the filters was made by R. P. Teele, of the photometry section. The quoted parts are from his report, B. S. 156 (e)–1–28, dated October 1, 1927:

Lamp *A* (fig. 33) is a 500-watt gas-filled standard lamp, B.S. 3256, belonging to the colorimetry section. It had been calibrated by Gibson in May, 1925, color temperature vs. applied voltage, over the range from 2,050° to 2,900° K.

Lamp *B* is a 500-watt 116-volt floodlighting lamp. * * *

The procedure followed was to have an observer vary the voltage of lamp *B* to match the color of lamp *A* at 2,360° K. The distance of *A* from *P* was

constant. The observer moved B to get a brightness match as well as color match between A and B. The voltage of B was recorded.

The cell designated as 2,360° K. to mean sun was then placed at C' and the cell being observed placed at C. The cells placed at C were 2,500° K. to mean sun, 2,700° K. to mean sun, and 2,848° K. to mean sun. The lamp A was then run at the desired temperature and the observer varied the voltage of B to obtain a color match. The voltage of B obtained with the cells in place and with the barc lamps should be the same if the cells are correct. * * *

From 6 to 8 observers took 2 observations each, both with and without the filters as described above, so that from 12 to 16 single observations went into the following averages:

	No cells	2,500° K. cell at C	No cells	2,700° K. cell at C	No cells	2,848° K. cell at C
Average voltage	71.2$_8$	71.2$_6$	70.6$_4$	70.6$_5$	70.8$_4$	70.8$_1$
Average percentage deviation	1.2$_5$	1.9$_5$	1.2$_4$	1.5$_9$	1.3$_9$	1.1$_7$

The "average percentage deviation" is an indication of the precision of the determinations, in one case at the color of 2,360° K. and in the other at the color of mean sun. The final averages are 1.29 and 1.57 per cent, respectively, and the difference, if significant, may easily be accounted for by the fact that the determinations at the color of mean sun were made at a brightness level of only 17 to 26 per cent (the transmission of the respective filters at C) of those at the color of 2,360° K.

It may or may not be significant that the average percentage deviation stays practically constant at the three trials with the 2,360° K. color (no filters), while, with the mean-sun color, it consistently decreases as the field brightness increases. (The distance AP being constant, the field brightness increases notably with increase in temperature of lamp A for two reasons, viz: (1) Increase in the candlepower of A and (2) increase in the transmission of the filter C.) This suggests that the variation in precision may be due essentially to the variation in field brightness. Since, with lamp A at 2,848° K. (relatively high brightness for the sunlight color), the average deviation is less than with the color match at 2,360° K., one is led to suspect that the deviation with lamp A at lower temperatures would also have been less than with field color 2,360° K. if the field brightness in these cases had been as high as with lamp A at 2,848° K. At any rate there would seem to be no loss of precision in determining the color temperatures of lamps by means of the filters of Series I.

So far as the accuracy of the filters for this purpose is concerned, the test is obviously satisfactory. The filters were prepared by Davis, using the same materials as used in deriving the standard data; but, as noted elsewhere, any deviations from the standard data caused by different chemicals would tend to cancel out in this use of the filters. The color matches were reported as perfect in all cases.

6. TEST OF ACETYLENE-TO-MEAN-SUN FILTER

A similar test was carried out using the filter designed to convert the acetylene flame, as represented by Coblentz's energy data (Table 3 and Sec. V, 6 (a)), to mean sun. Quoting from the same report as before:

The lamp *A* (fig. 33) was replaced by an acetylene burner and the acetylene filter placed at *C*. The 2,360° K.-to-mean-sun filter was placed at *C'* and the voltage of lamp *B* varied for color match.

Eight observers took two readings each, so that 16 observations enter into the final average. A value of 71.3$_4$ volts was obtained, compared with an average of 70.9$_2$ obtained from the three previous measurements with the bare lamps and 70.9$_1$ with the filters. This difference is equivalent to about 5° at 2,360° K. The average "percentage deviation" was 2.18, larger than obtained with the lamp and filters, and is readily accounted for by the fact that a perfect color match was not obtained with the acetylene and its filter. The conditions under which the acetylene was burned were as follows:

Temperatures	Wet bulb, 20.2° C. Dry bulb, 23.2° C.
Barometer	759 mm.
Burner	E. K. standard.
Gas	Generated from calcium carbide and water.

The pressure of the acetylene was 9 cm of water and was measured by an open manometer furnished with the burner.

Further examination of the filter by the authors, in the photometry section, confirmed the fact that the acetylene-and-filter combination did not give a perfect color match with mean sun as obtained with a standardized electric incandescent lamp operated at 2,360° K. in combination with a 2,360° K.-to-mean-sun filter. Removing the two filters showed a still larger difference in color. The acetylene was distinctly less saturated (whiter) in comparison. Next the current in the electric lamp was adjusted, without filters, in an attempt to obtain a color match. It was found, however, that a perfect color match between the acetylene light and the unfiltered electric light could not be obtained at any color temperature. The nearest approach to this was somewhat over 2,400° K. During these experiments two Eastman Kodak acetylene burners, two calibrated electric lamps, and two sets each of acetylene-to-mean-sun and 2,360° K.-to-mean-sun filter solutions were used. The acetylene gas used in this experiment was made in the same way as described above.

The observations just described were not extensive enough to justify giving any new value for the color temperature of the acetylene flame as used with the standard Eastman Kodak burner, nor are they extensive enough to justify saying that a value of 2,360° K. might never be obtained. Nevertheless, in view of the fact that Coblentz's energy data and the observations under consideration both give color temperatures definitely greater than 2,360° K. (Sec. III, 4), it is felt that this value is rendered somewhat questionable, and that further careful measurement should be carried out. So far as these observations go they show that, if one wishes to calibrate an electric lamp for 2,360° K. by means of the acetylene flame, a much more accurate value will be obtained when the filters are used than when not.

Presto-Lite acetylene was tried and found to be different from generated acetylene. The gas is absorbed by an earth impregnated with acetone in the Presto-Lite tanks. The chemistry division of the bureau has done considerable work on Presto-Lite, and finds that the acetone is carried over with the acetylene by an amount that varies from about 1.6 per cent when the tank is freshly charged to about

3.0 per cent when the tank is nearly empty. It would seem inadvisable to use Presto-Lite for this type of work as the variation in acetone present changes the color of the flame. However, the Presto-Lite gas used gave a flame that within the uncertainties of the experiment corresponded to a color temperature of 2,360° K.

IX. PREPARATION OF DAVIS-GIBSON FILTERS

1. SELECTION OF CHEMICALS

In selecting chemicals to be used for preparing these filters care should be taken to obtain those of a quality at least equal to that known to the trade as "c. p." (chemically pure), from manufacturing chemists who can certify to their purity. A higher grade of chemicals known as "reagent" quality is recommended, as this would be some protection where no facilities exist for making chemical or spectrophotometric analysis of purchased materials.

Chemical analysis of the principal materials used in the design of these filters has already been given in Section IV, 2, and further noted in Section VII, 8. The following comments, however, may be of some assistance in detecting unsuitable chemicals:

1. *Copper sulphate* ($CuSO_4 \cdot 5H_2O$)—*Appearance.*—Clean blue crystals of the pentahydrate free from white powder ($CuSO_4 \cdot 3H_2O$) or from any signs of moisture, such as small crystals (not the powder) adhering to the sides of the bottle. It should make a clear solution in water.

2. *Cobalt ammonium sulphate* ($CoSO_4(NH_4)_2SO_4 \cdot 6H_2O$).—*Appearance.*—Well-defined crystals, which should be free from any signs of moisture. It should make a clear solution in water.

If for any reason cobalt ammonium sulphate is not obtainable, cobalt sulphate ($CoSO_4 \cdot 7H_2O$) may be used (see, however, sections IV, 1 (*d*), and IV, 1 (*e*)), provided the weights of cobalt ammonium sulphate specified are adjusted to give the same cobalt content. This may be obtained by multiplying the weights given on the charts by 0.7113.

3. *Mannite* ($C_6H_8(OH)_6$), *d*-mannitol.—Colorless needles which melt at 166° C. (fine crystals, which, en masse, appear white).

4. *Pyridine* (C_5H_5N).—A colorless liquid organic compound having a strong unpleasant odor. Specific gravity 0.979 at 25° C., boiling point 115.5° C. Should be practically colorless, and when added to water the solution should be clear.

5. *Sulphuric acid* (H_2SO_4).—Should be clear and practically colorless. Specific gravity 1.84.

2. PREPARATION OF A AND B SOLUTIONS

Solutions A and B are given to milligrams on the charts. This may appear to be an unwarranted refinement. However, we have attempted to design these filters so that the color in each case, as expressed in triliner coordinates, *r*, *g*, and *b*, would coincide with those of the ideal to four decimal places by computation. In Tables 8 and 12, giving the concentrations of the ingredients for intermediate temperatures, the weights of cobalt ammonium sulphate and the copper sulphate in solution B are given to centigrams only, as this is sufficiently accurate for all practical purposes. The weights of copper

sulphate in solution A are given to milligrams because 1 mg in this case has even more significance than 10 mg of the ingredients in solution B.

1. *Solution A.*—Select several crystals of copper sulphate, enough for both A and B solutions, and crush in a mortar (not pulverize) until sufficiently fine to facilitate weighing. Weigh out in watch crystal the quantity required for solution A, and dissolve in about one-half the total quantity of distilled water in a beaker. Use a wash bottle to get all of the weighed copper sulphate off the watch crystal into the solution. With the same weights, weigh out the mannite and add to the solution. After dissolving, add the pyridine and pour into a volumetric flask. Use the wash bottle to rinse out the beaker and add this rinse to the flask and make up the full volume with distilled water. Mix thoroughly and pour into a glass-stoppered bottle, which should be labeled and dated. Most chemicals contain a small quantity of minute insoluble material (trash) which will settle to the bottom of the bottle if left undisturbed for a few days. This solution should not be filtered. The clean solution can be removed with a pipette or by careful pouring if the bottle is nearly full. For work where the highest precision is desired this solution should be made up fresh every two or three months.

2. *Solution B.*—Follow the same procedure as given for solution A, adding the sulphuric acid after the copper sulphate and cobalt sulphate are dissolved. This solution apparently does not change with time. It is difficult to state just how long the B solution will keep. It is known to be good for over a year and probably is good for a much longer period of time.

3. PREPARATION AND FILLING OF CELLS

The cell illustrated in Figure 9 for holding the filter solutions consists of three plates of optical glass, borosilicate crown,[19] 76 mm square by 2.5 mm thick, separated by two black glass[20] frames, fine ground (not polished) to 10.00 mm thickness, with a central circular aperture for the solution 43 mm in diameter. A small hole, 5 mm in diameter, is cut from one side of the black glass frame to the central aperture. This hole is closed with a waxed cork or glass stopper and may be used in filling the cell.

The black glass separator frames when new should be thoroughly scrubbed with soap and water, using a brush, to remove particles of glass mud and abrasives with which the ground glass surfaces are usually impregnated; particular attention should be paid to cleaning the small hole through which the cell is filled. To avoid lint, the frames should be air dried without wiping with a cloth. The polished optical glass windows should be cleaned and wiped dry with a soft cloth or with lens tissue. When assembling the cell the parts should be carefully brushed (with a camel hair brush) to remove any dust particles and lint that may be adhering. In case the cell is to be filled without sealing the windows on with wax, it may be assembled wet and each compartment rinsed out with the filter solution several times before final filling.

[19] Borosilicate crown glass was chosen as the most suitable of the optical glasses. It has very high and almost nonselective transmission, and as previously noted, it is also the hardest of the optical glasses and probably the most resistant to chemical action.

[20] Black glass is recommended because it prevents the passage of light through all portions of the cell other than the central aperture, thereby eliminating a mixture of filtered and unfiltered light in the apparatus in which it is used.

The cell may be held together, as shown in Figure 9, with rubber bands or may be further sealed with paraffin or ceresin before filling. White ceresin is preferable to paraffin because of its higher melting point and hardness. The cell should be warmed (70° to 80° C.) and then placed (with rubber bands and stoppers in position) alternately on each edge in a pan of melted wax about 1 cm deep and allowed to remain long enough in each position for a film of the wax, by capillary action, to fill the seam up to the compartment for the filter solutions. If the cell is to be used only a short time (a few hours or even days) and the solution changed, it is advisable not to wax the edges as it will not leak if the small filling hole is tightly stoppered. In this case, the solution is easily removed and the cell washed out by separating the parts. If the windows are on very tight, as is usually the case after standing several days, the stoppers may be removed and reinserted. In this way the stopper serves as a pump forcing the solutions between the windows and the black glass frame. When wax is used to seal the cell it is, of course, necessary that the cell be warmed to take it apart. If a cell with larger aperture is desired, it is recommended that the outside dimension (not thickness) be also increased, because the cell is likely to leak if insufficient distance (surface area) between the outside and the central hole in the black glass is not maintained.

It should be remembered that the data given for the filters are based on a temperature of 25° C. for the filter solutions. As noted before a tolerance of ± 5° C. is permissible for most work (see Secs. VII, 4, VIII, 2, (*d*), VI, 1, (*b*), and VI, 2, (*b*)).

X. CHARTS DESCRIBING FILTERS

Charts 1 to 20, inclusive, contain the filter formulas and the spectral specifications of the mean-sun group of filters, referred to throughout the paper as Series I. Chart 2 is the 2,360° K-to-mean-sun filter recommended for use in photographic sensitometry by the Seventh (1928) International Congress of Photography. The data on these charts are self-explanatory.

Charts 21 to 31, inclusive, Series II, comprise similar specifications for filters designed to convert the light from a 2,848° K. source to color temperatures between the limits 3,500° to 10,000° K.

Several other filters not in the above series are given in charts 32 to 38, inclusive, as listed in the accompanying index.

The weights of the ingredients and the light transmission for any intermediate filter, not given on the charts, may be found in Table 8 for the filters of Series I and in Table 12 for those of Series II.

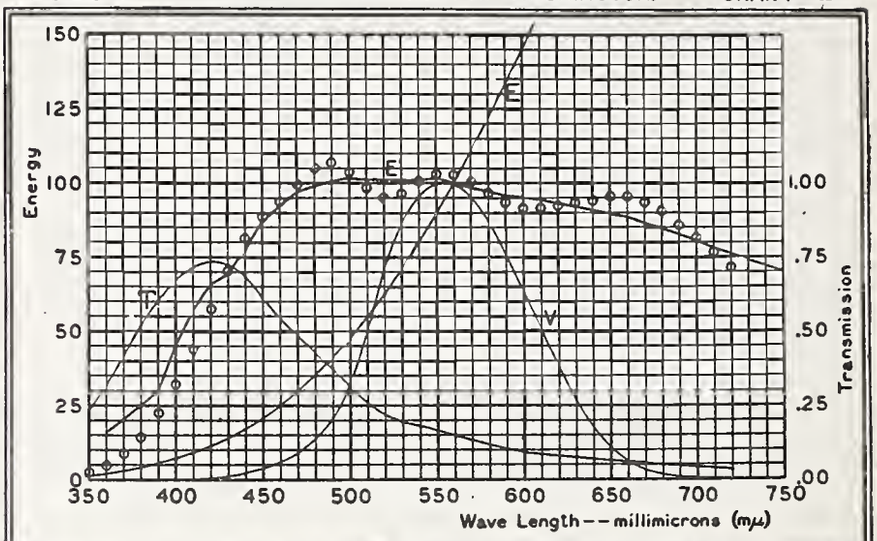
Index to charts

*1. 2,300° K. to mean sunlight.
2. 2,360° K. to mean sunlight.
*3. 2,400° K. to mean sunlight.
4. 2,500° K. to mean sunlight.
*5. 2,600° K. to mean sunlight.
6. 2,700° K. to mean sunlight.
*7. 2,800° K. to mean sunlight.
8. 2,848° K. to mean sunlight.
*9. 2,900° K. to mean sunlight.
10. 3,000° K. to mean sunlight.
*11. 3,100° K. to mean sunlight.
12. 3,200° K. to mean sunlight.
*13. 3,300° K. to mean sunlight.
14. 3,400° K. to mean sunlight.
*15. 3,500° K. to mean sunlight.
16. 3,600° K. to mean sunlight.
*17. 3,700° K. to mean sunlight.
18. 3,800° K. to mean sunlight.
*19. 3,900° K. to mean sunlight.
20. 4,000° K. to mean sunlight.
21. 2,848° K. to 3,500° K.
22. 2,848° K. to 4,000° K.
*23. 2,848° K. to 4,500° K.
24. 2,848° K. to 5,000° K.
*25. 2,848° K. to 5,500° K.
26. 2,848° K. to 6,000° K.
*27. 2,848° K. to 6,500° K.
28. 2,848° K. to 7,000° K.
29. 2,848° K. to 8,000° K.
30. 2,848° K. to 9,000° K.
31. 2,848° K. to 10,000° K.
**32. 2,848° K. to sunlight outside earth's atmosphere.
33. 2,848° K. to Abbot-Priest sun.
34. 2,848° K. to summer noon sunlight.
35. 2,848° K. to winter noon sunlight.
36. Acetylene to mean sunlight.
**37. 2,450° K. to 3,500° K.
**38. 2,450° K. to 6,500° K.

* In order to save space these charts are not published herein but any of them may be obtained from the authors on request.

** Colorimetric characteristics for these filters obtained by varying the color temperature of the source are described in Sections V, 6, (*d*) and V, 6, (*e*), Tables 13, 14 and 15.

2360°K TO MEAN NOON SUNLIGHT AT WASHINGTON. CHART 2



λ (mμ)	T	E"*	E"/E'
350	0.230	2.5	——
60	.321	4.9	.304
70	.420	8.7	.426
80	.518	14.5	.579
90	.608	22.5	.749
400	.671	32.3	.715
10	.714	44.0	.767
20	.737	57.4	.871
30	.725	70.3	1.016
40	.682	81.2	1.044
450	.612	88.6	1.021
60	.540	93.9	1.018
70	.483	99.9	1.031
80	.431	105.0	1.060
90	.376	107.1	1.065
500	.315	103.9	1.020
10	.259	98.3	.972
20	.219	95.0	.939
30	.197	96.7	.957
40	.182	100.6	.997
550	.167	103.2	1.014
60	.1495	103.1	1.031
70	.1316	100.4	1.021
80	.1151	96.9	.997
90	.1013	93.4	.977
600	.0907	91.3	.959
10	.0834	91.2	.968
20	.0778	92.3	.991
30	.0730	93.3	1.012
40	.0684	94.0	1.032
650	.0646	95.1	1.060
60	.0605	95.1	1.074
70	.0558	93.4	1.081
80	.0510	90.5	1.069
90	.0457	85.8	1.038
700	.0409	81.2	1.008
10	.0366	76.4	.979
20	.0326	71.6	.940

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2360°K
E' -- Relative Energy of Mean Noon Sunlight at Washington
E" (=T x E) ---oooo-- Relative Energy of 2360°K and Filter Combination

Light Transmission of Filter for 2360°K = 0.1352 **

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	3.707	grams
Mannite ($C_6H_8(OH)_6$)	3.707	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

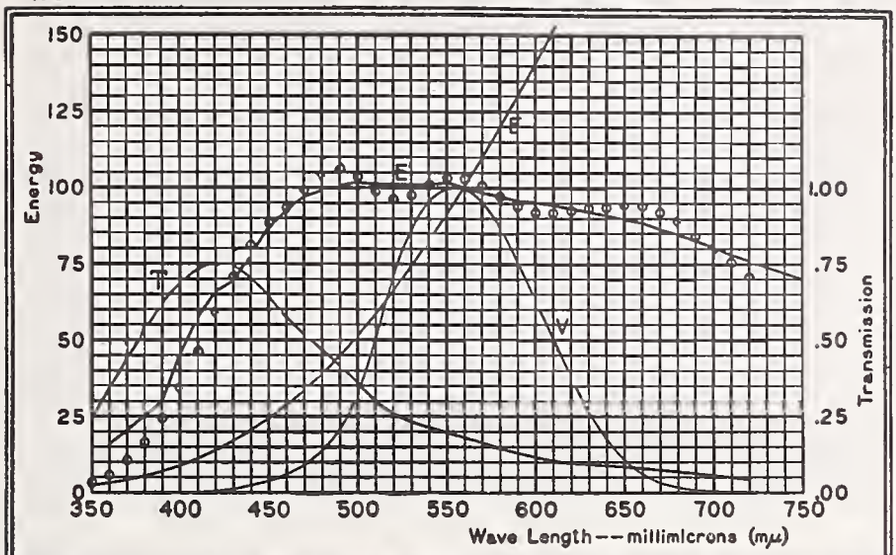
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	26.827	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	27.180	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2500°K TO MEAN NOON SUNLIGHT AT WASHINGTON. CHART 4



λ (mμ)	T	E"*	E"/E'
350	0.249	3.2	——
60	.342	5.9	.372
70	.441	10.3	.501
60	.538	16.5	.657
60	.625	24.8	.824
400	.688	34.6	.766
10	.730	46.1	.804
20	.753	58.9	.895
30	.743	71.1	1.027
40	.704	81.2	1.045
450	.641	88.2	1.016
60	.573	93.4	1.012
70	.520	99.2	1.023
80	.470	104.0	1.051
90	.416	106.3	1.057
500	.356	103.6	1.018
10	.299	99.0	.978
20	.258	96.1	.950
30	.234	97.6	.966
40	.218	101.0	1.001
550	.201	103.2	1.015
60	.182	103.0	1.030
70	.162	100.6	1.022
80	.1437	97.3	1.001
90	.1281	94.1	.984
600	.1159	92.0	.967
10	.1074	91.9	.975
20	.1009	92.7	.995
30	.0952	93.3	1.012
40	.0896	93.6	1.029
650	.0849	94.4	1.052
60	.0799	94.1	1.063
70	.0741	92.1	1.067
80	.0681	89.2	1.053
90	.0615	84.5	1.022
700	.0555	80.0	.993
10	.0501	75.5	.967
20	.0450	70.6	.928

Bureau of Standards. Davis-Gibson Filters

T — Spectral Transmission of Filter at 25°C
V — Relative Visibility Function
E — Relative Energy of 2500°K
E' — Relative Energy of Mean Noon Sunlight at Washington
E" (=T x E)* — o o o o — Relative Energy of 2500°K and Filter Combination

Light Transmission of Filter ** for 2500°K = 0.168

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	3.292	grams
Mannite ($C_6H_8(OH)_6$)	3.292	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

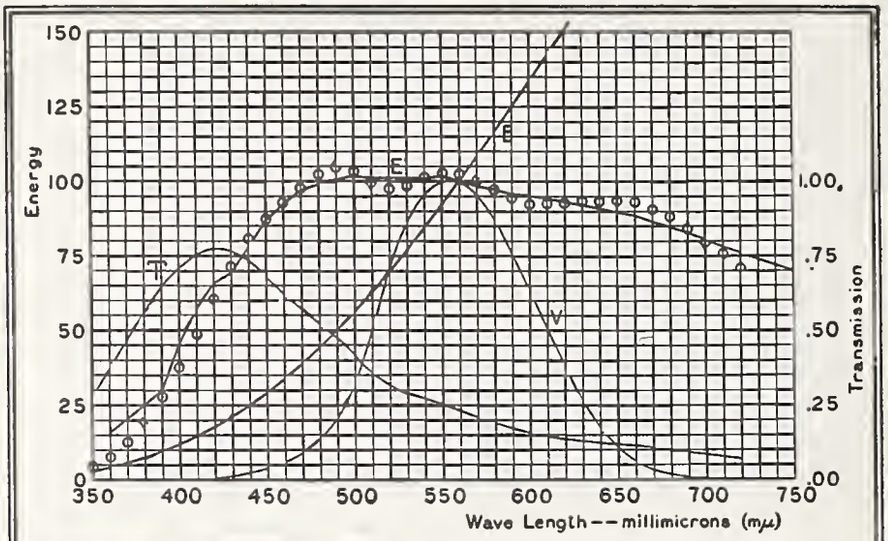
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	23.590	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	24.730	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2700°K TO MEAN NOON SUNLIGHT AT WASHINGTON. CHART 6



λ (mμ)	T	E"*	E"/E'
350	0.282	4.4	——
60	.377	7.7	.483
70	.474	12.6	.617
80	.568	19.3	.772
90	.652	27.9	.926
400	.712	37.8	.836
10	.753	48.8	.850
20	.774	60.5	.919
30	.767	71.5	1.033
40	.734	80.9	1.040
450	.680	87.5	1.008
60	.621	92.6	1.004
70	.572	97.9	1.010
80	.526	102.7	1.036
90	.476	104.9	1.043
500	.417	103.4	1.016
10	.360	99.7	.985
20	.318	97.4	.963
30	.292	98.6	.975
40	.273	101.2	1.003
550	.255	102.9	1.012
60	.233	102.4	1.024
70	.211	100.2	1.018
80	.190	97.2	1.001
90	.172	94.4	.987
600	.158	92.6	.973
10	.1481	92.3	.979
20	.1404	92.8	.996
30	.1333	93.2	1.012
40	.1264	93.3	1.025
650	.1207	93.6	1.044
60	.1144	93.1	1.051
70	.1070	90.9	1.053
80	.0995	88.3	1.042
90	.0909	84.0	1.016
700	.0831	79.8	.991
10	.0759	75.5	.968
20	.0690	71.0	.933

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2700°K
E' -- Relative Energy of Mean Noon Sunlight at Washington
E" (=T x E)* -- o o o o -- Relative Energy of 2700°K and Filter Combination

Light Transmission of Filter ** for 2700°K = 0.220

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	2.777	grams
Mannite ($C_6H_8(OH)_6$)	2.777	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

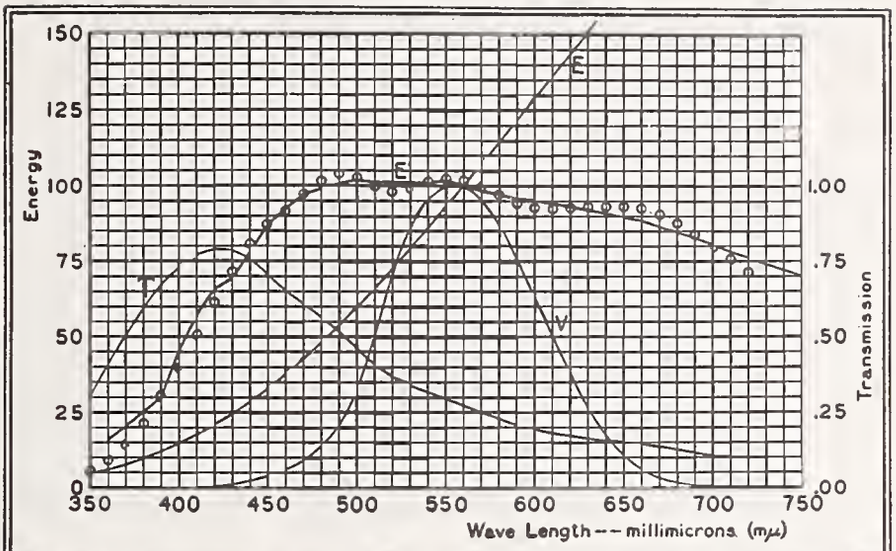
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2SO_4 \cdot 6H_2O$)	19.325	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	21.300	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848°K TO MEAN NOON SUNLIGHT AT WASHINGTON. CHART 8



λ (mμ)	T	E"*	E"/E'
350	0.309	5.5	——
60	.404	9.2	.579
70	.500	14.5	.710
80	.591	21.5	.858
90	.671	30.1	1.002
400	.728	39.8	.881
10	.768	50.4	.880
20	.788	61.5	.934
30	.783	71.8	1.037
40	.755	80.6	1.036
450	.707	87.0	1.001
60	.654	91.9	.996
70	.609	97.1	1.001
80	.567	101.6	1.026
90	.519	104.0	1.034
500	.463	102.9	1.011
10	.407	100.0	.988
20	.364	98.1	.970
30	.337	99.2	.982
40	.317	101.3	1.004
550	.297	102.5	1.008
60	.274	101.8	1.018
70	.250	99.8	1.014
80	.228	97.1	.999
90	.209	94.6	.989
600	.194	92.9	.976
10	.183	92.6	.982
20	.174	92.9	.997
30	.166	93.1	1.011
40	.158	93.0	1.022
650	.152	93.2	1.039
60	.1446	92.5	1.045
70	.1362	90.5	1.048
80	.1274	87.8	1.037
90	.1175	83.9	1.015
700	.1084	79.9	.991
10	.0998	75.8	.972
20	.0915	71.5	.940

Bureau of Standards. Davis-Gibson Filters

T — Spectral Transmission of Filter at 25°C
V — Relative Visibility Function
E — Relative Energy of 2848°K
E' — Relative Energy of Mean Noon Sunlight at Washington
E"(=T×E)* — o o o o — Relative Energy of 2848°K and Filter Combination

Light Transmission of Filter ** for 2848°K = 0.262

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	2.445	grams
Mannite ($C_6H_8(OH)_6$)	2.445	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

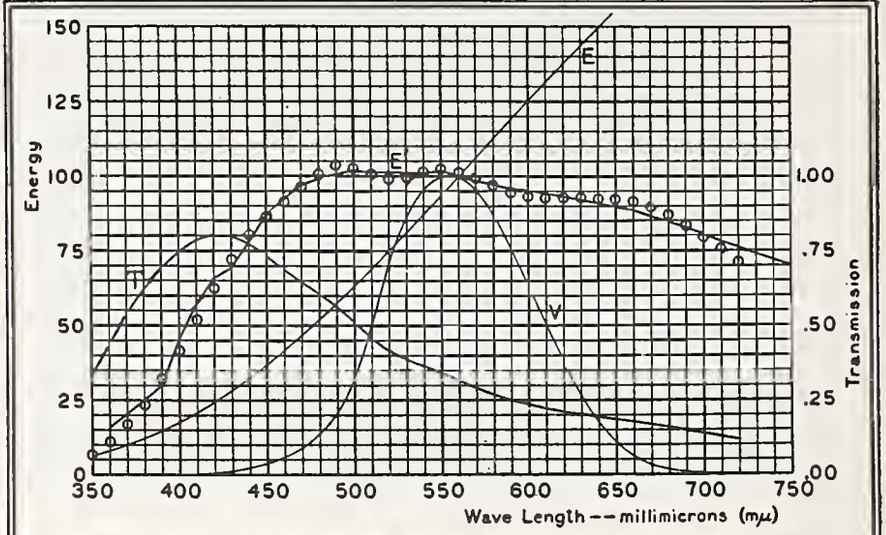
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2SO_4 \cdot 6H_2O$)	16.520	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	19.020	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

3000°K TO MEAN NOON SUNLIGHT AT WASHINGTON. CHART 10



λ (mμ)	T	E" *	E"/E'
350	0.340	6.7	—
60	.434	11.0	.696
70	.528	16.6	.812
80	.615	23.8	.951
90	.691	32.5	1.079
400	.745	41.9	.928
10	.782	52.0	.908
20	.802	62.4	.948
30	.798	72.1	1.041
40	.774	80.3	1.032
450	.733	86.4	.995
60	.686	91.4	.991
70	.647	96.5	.995
80	.607	100.8	1.018
90	.563	103.3	1.027
500	.510	102.7	1.010
10	.456	100.5	.993
20	.414	99.0	.978
30	.386	99.8	.988
40	.364	101.5	1.006
550	.343	102.3	1.006
60	.318	101.5	1.015
70	.294	99.5	1.011
80	.270	97.1	.999
90	.250	94.8	.991
600	.233	93.2	.980
10	.222	92.9	.986
20	.212	93.0	.998
30	.203	93.0	1.009
40	.195	92.7	1.019
650	.188	92.6	1.033
60	.179	91.7	1.036
70	.170	89.8	1.039
80	.160	87.2	1.029
90	.1484	83.4	1.008
700	.1378	79.6	.989
10	.1279	75.7	.970
20	.1180	71.5	.940

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 3000°K
E' -- Relative Energy of Mean Noon Sunlight at Washington
E" (=T x E) * ---oooo-- Relative Energy of 3000°K and Filter Combination

Light Transmission of Filter ** for 3000°K = 0.307

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	2.140	grams
Mannite ($C_6H_8(OH)_6$)	2.140	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

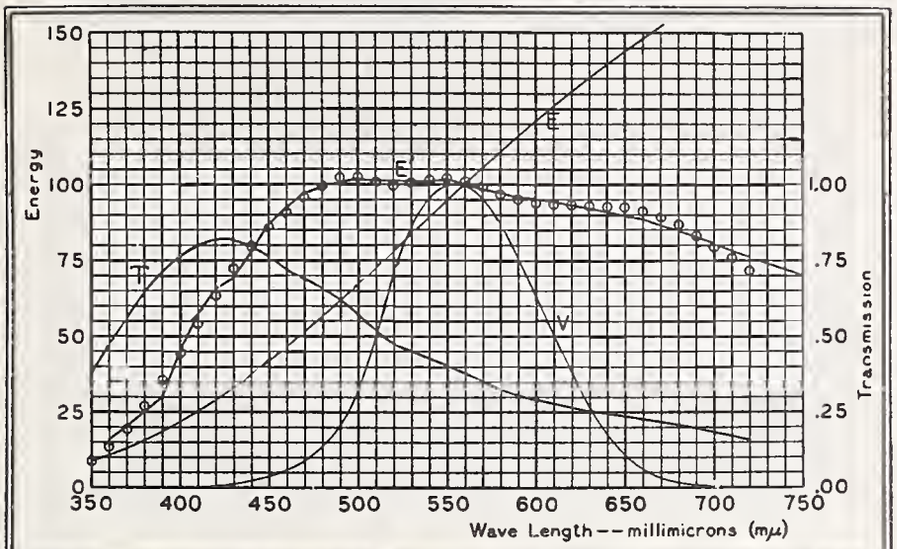
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	13.900	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	17.000	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

3200°K TO MEAN NOON SUNLIGHT AT WASHINGTON. CHART 12



λ (mμ)	T	E"*	E"/E'
350	0.382	8.7	—
60	.475	13.5	.844
70	.565	19.6	.956
80	.647	27.0	1.077
90	.716	35.5	1.180
400	.766	44.6	.988
10	.800	54.1	.944
20	.818	63.6	.966
30	.816	72.4	1.046
40	.798	79.9	1.027
450	.765	85.9	.989
60	.726	90.7	.983
70	.692	95.4	.984
80	.658	99.6	1.006
90	.618	102.3	1.018
500	.570	102.4	1.006
10	.521	100.9	.998
20	.479	99.8	.986
30	.451	100.4	.994
40	.428	101.5	1.006
550	.405	101.9	1.002
60	.379	101.0	1.010
70	.354	99.1	1.008
80	.329	97.0	.998
90	.308	95.0	.993
600	.291	93.6	.984
10	.278	93.2	.988
20	.268	93.2	1.000
30	.258	92.9	1.008
40	.249	92.6	1.017
650	.240	92.2	1.028
60	.231	91.1	1.029
70	.220	89.1	1.032
80	.209	86.7	1.023
90	.196	83.0	1.004
700	.183	79.5	.987
10	.172	75.8	.972
20	.160	71.8	.944

Bureau of Standards. Davis-Gibson Filters

T — Spectral Transmission of Filter at 25°C
V — Relative Visibility Function
E — Relative Energy of 3200°K
E' — Relative Energy of Mean Noon Sunlight at Washington
E" (=T x E)* — o o o o — Relative Energy of 3200°K and Filter Combination

Light Transmission of Filter ** for 3200°K = 0.370

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	1.784	grams
Mannite ($C_6H_8(OH)_8$)	1.784	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	10.890	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	14.570	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

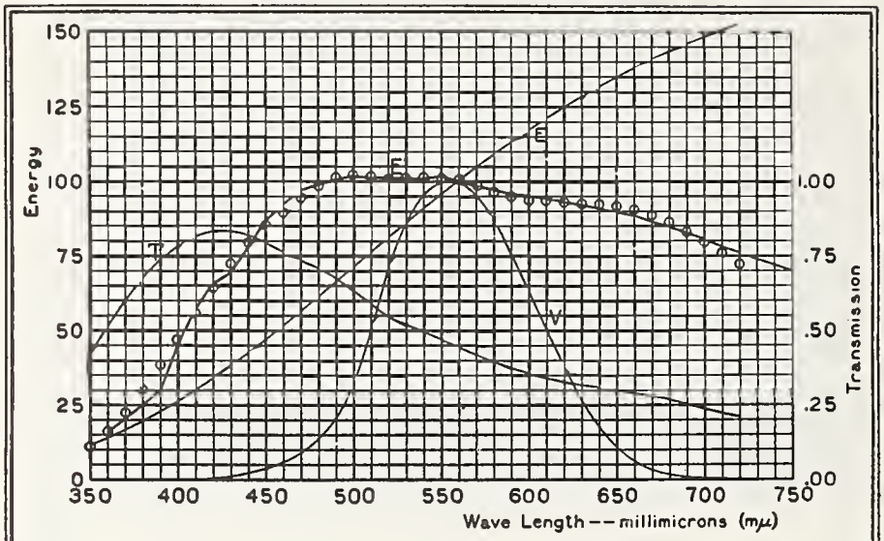
* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

783/143

518-743 O - 73 - 51

3400°K TO MEAN NOON SUNLIGHT AT WASHINGTON. CHART 14



λ (mμ)	T	E"*	E"/E'
350	0.428	11.0	—
60	.518	16.3	1.018
70	.603	22.6	1.106
80	.679	30.1	1.201
90	.741	38.4	1.276
400	.786	47.0	1.040
10	.816	55.8	.973
20	.833	64.5	.980
30	.833	72.4	1.046
40	.820	79.5	1.022
450	.795	85.2	.981
60	.764	89.8	.973
70	.737	94.7	.977
80	.708	98.7	.996
90	.674	101.5	1.010
500	.632	102.1	1.003
10	.587	101.4	1.002
20	.547	100.7	.995
30	.518	101.0	1.000
40	.494	101.6	1.007
550	.470	101.5	.998
60	.443	100.5	1.005
70	.417	98.7	1.003
80	.392	96.8	.996
90	.371	95.0	.994
600	.353	93.7	.985
10	.340	93.3	.990
20	.329	93.0	.999
30	.319	92.8	1.007
40	.309	92.3	1.014
650	.300	91.8	1.024
60	.290	90.6	1.023
70	.278	88.7	1.027
80	.265	86.3	1.019
90	.251	83.0	1.004
700	.237	79.7	.990
10	.224	76.3	.977
20	.210	72.5	.953

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 3400°K
E' -- Relative Energy of Mean Noon Sunlight at Washington
E" (=T x E)* -- o o o o -- Relative Energy of 3400°K and Filter Combination

Light Transmission of Filter ** for 3400°K = 0.437

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	1.475	grams
Mannite ($C_6H_8(OH)_6$)	1.475	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

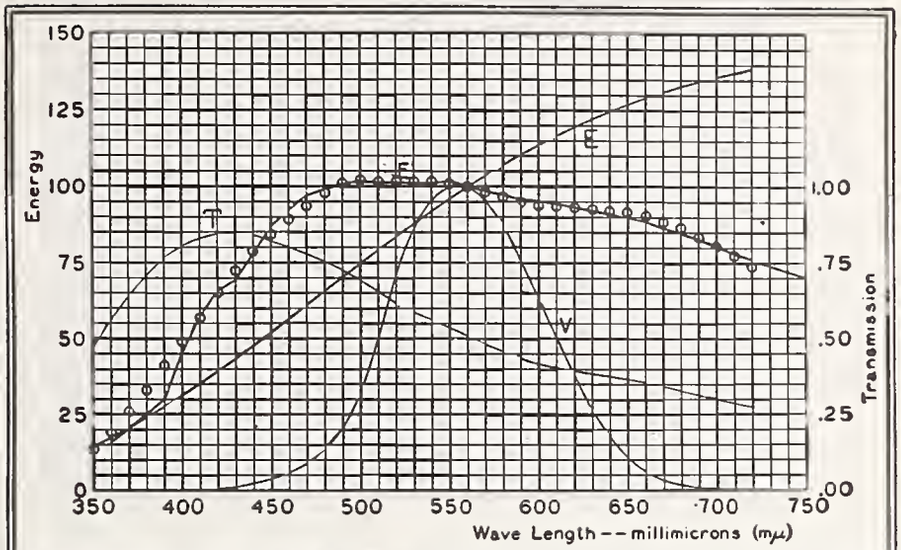
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	8.100	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	12.330	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

3600°K TO MEAN NOON SUNLIGHT AT WASHINGTON. CHART 16



λ (mμ)	T	E"*	E"/E'
350	0.476	13.7	——
60	.563	19.2	1.205
70	.642	25.8	1.262
80	.710	33.2	1.323
90	.765	41.2	1.368
400	.805	49.2	1.089
10	.832	57.2	.998
20	.847	65.1	.989
30	.849	72.4	1.047
40	.841	78.9	1.014
450	.823	84.4	.972
60	.800	89.2	.967
70	.779	93.7	.967
80	.756	97.7	.987
90	.728	100.7	1.001
500	.693	101.9	1.001
10	.654	101.7	1.005
20	.617	101.3	1.001
30	.588	101.5	1.004
40	.562	101.6	1.007
550	.536	101.1	.994
60	.509	99.9	.999
70	.483	98.1	.998
80	.459	96.4	.992
90	.437	94.9	.992
600	.420	93.7	.985
10	.407	93.3	.990
20	.396	93.0	.999
30	.386	92.6	1.005
40	.375	92.0	1.011
650	.366	91.4	1.019
60	.356	90.3	1.020
70	.343	88.5	1.024
80	.330	86.3	1.019
90	.315	83.3	1.007
700	.300	80.3	.997
10	.286	77.1	.988
20	.271	73.7	.968

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 3600°K
E' -- Relative Energy of Mean Noon Sunlight at Washington
E" (=T x E)* ---oooo-- Relative Energy of 3600°K and Filter Combination

Light Transmission of Filter ** for 3600°K = 0.506

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	1.199	grams
Mannite ($C_6H_8(OH)_6$)	1.199	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

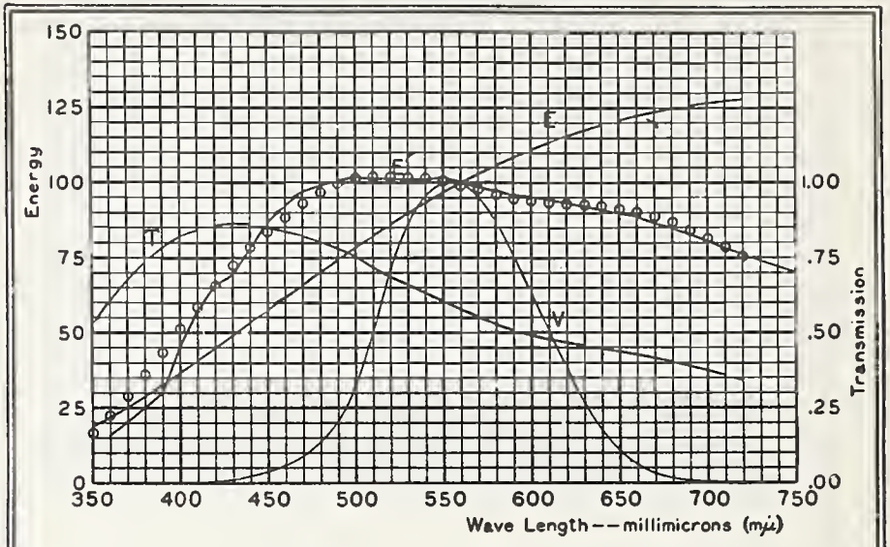
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	5.570	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	10.220	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

3800°K TO MEAN NOON SUNLIGHT AT WASHINGTON. CHART 18



λ (mμ)	T	E"*	E"/E'
350	0.526	16.6	——
60	.608	22.4	1.403
70	.680	29.0	1.419
80	.740	36.1	1.442
90	.784	43.7	1.452
400	.823	51.1	1.131
10	.846	58.5	1.021
20	.859	65.6	.996
30	.863	72.2	1.044
40	.860	78.3	1.007
450	.849	83.7	.964
60	.835	88.4	.958
70	.821	92.9	.958
80	.804	96.8	.978
90	.782	99.8	.992
500	.754	101.5	.997
10	.721	101.9	1.007
20	.688	101.9	1.006
30	.659	101.7	1.007
40	.632	101.3	1.004
550	.605	100.4	.987
60	.577	99.1	.991
70	.551	97.4	.990
80	.528	95.9	.987
90	.508	94.6	.989
600	.491	93.7	.984
10	.479	93.2	.989
20	.468	92.9	.998
30	.458	92.5	1.004
40	.448	91.8	1.009
650	.439	91.2	1.017
60	.429	90.1	1.018
70	.417	88.5	1.024
80	.404	86.6	1.022
90	.389	84.0	1.015
700	.374	81.3	1.010
10	.359	78.5	1.005
20	.342	75.1	.988

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 3800°K
E' -- Relative Energy of Mean Noon Sunlight at Washington
E" (=T x E)* ---o o o o-- Relative Energy of 3800°K and Filter Combination

Light Transmission of Filter for 3800°K ** = 0.577

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	0.955	grams
Mannite ($C_6H_8(OH)_6$)	0.955	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

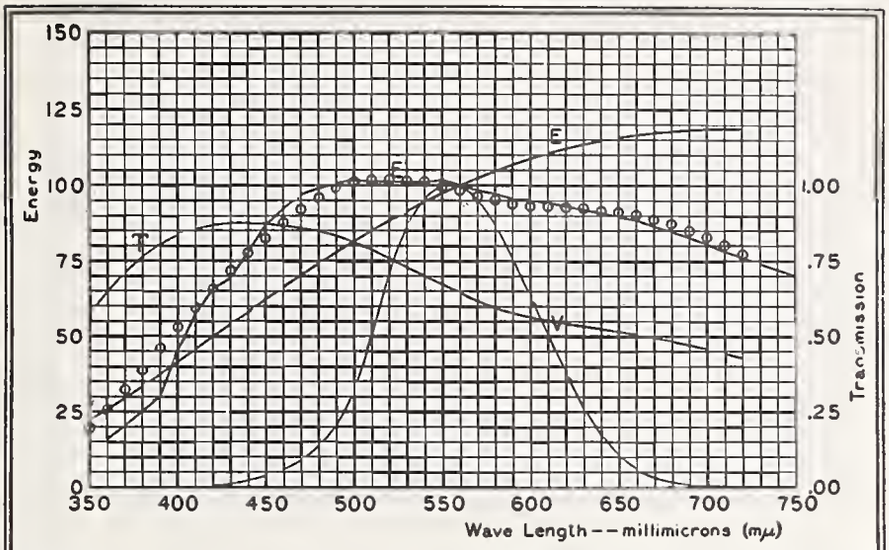
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	3.260	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	8.210	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero.

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

4000°K TO MEAN NOON SUNLIGHT AT WASHINGTON. CHART 20



λ (mμ)	T	E"*	E"/E'
350	0.577	19.7	—
60	.653	25.7	1.608
70	.717	32.4	1.581
80	.770	39.2	1.563
90	.810	46.1	1.532
400	.839	52.9	1.170
10	.859	59.6	1.039
20	.871	65.9	1.001
30	.877	72.0	1.040
40	.878	77.7	.999
450	.875	82.5	.950
60	.869	87.7	.951
70	.860	92.1	.950
80	.849	95.9	.969
90	.833	98.9	.983
500	.813	101.1	.993
10	.788	102.1	1.009
20	.759	102.2	1.010
30	.730	101.8	1.008
40	.701	100.9	1.000
550	.673	99.7	.980
60	.645	98.2	.982
70	.620	96.6	.982
80	.598	95.3	.981
90	.579	94.2	.985
600	.564	93.4	.981
10	.553	93.0	.987
20	.544	92.7	.995
30	.535	92.2	1.001
40	.526	91.7	1.007
650	.518	91.1	1.016
60	.509	90.1	1.018
70	.498	88.8	1.028
80	.486	87.1	1.029
90	.472	85.0	1.028
700	.458	82.7	1.027
10	.444	80.2	1.028
20	.427	77.2	1.015

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 4000°K
E' -- Relative Energy of Mean Noon Sunlight at Washington
E" (=T x E)* ---- Relative Energy of 4000°K and Filter Combination

Light Transmission of Filter for 4000°K = 0.649 **

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	0.741	grams
Mannite ($C_6H_8(OH)_6$)	0.741	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

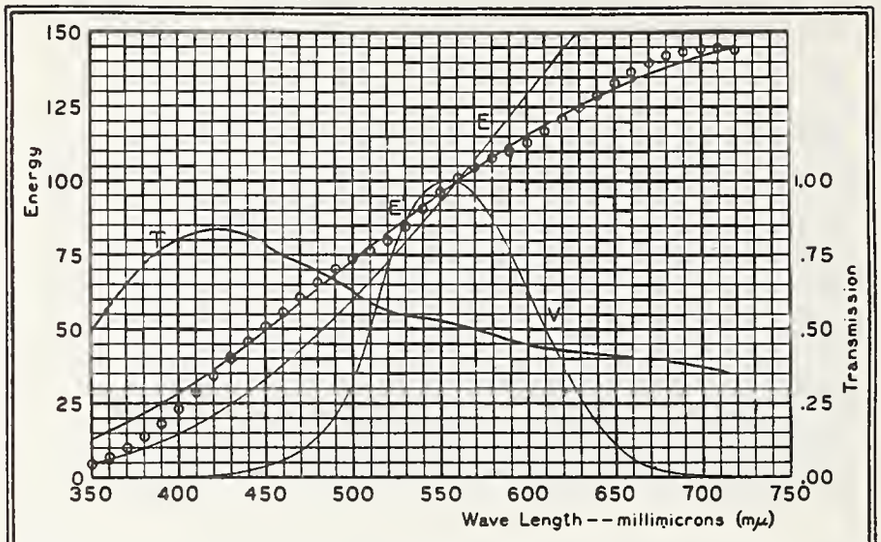
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	1.130	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	6.250	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848°K to 3500°K CHART 21



λ (mμ)	T	E"*	E"/E'
350	0.492	4.6	.353
60	.577	7.0	.447
70	.653	10.0	.541
80	.717	13.8	.635
90	.768	18.2	.726
400	.804	23.2	.806
10	.826	28.6	.878
20	.836	34.4	.938
30	.831	40.2	.982
40	.813	45.7	1.009
450	.783	50.7	1.019
60	.749	55.4	1.019
70	.721	60.5	1.025
80	.694	65.6	1.030
90	.664	70.1	1.025
500	.627	73.5	1.005
10	.589	76.3	.980
20	.561	79.7	.967
30	.546	84.7	.974
40	.538	90.5	.991
550	.528	96.1	1.005
60	.514	100.7	1.007
70	.498	104.4	1.004
80	.479	107.4	.995
90	.462	110.1	.985
600	.447	113.0	.979
10	.437	116.7	.982
20	.429	120.7	.989
30	.422	124.7	.996
40	.416	128.6	1.004
650	.411	132.9	1.015
60	.406	136.7	1.024
70	.399	139.7	1.028
80	.391	142.1	1.029
90	.381	143.4	1.024
700	.372	144.4	1.017
10	.361	144.7	1.008
20	.349	143.9	.992

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2848° K
E' -- Relative Energy of 3500°K
E" (=T x E)* -- o o o o -- Relative Energy of 2848°K and Filter Combination

Light Transmission of Filter ** for 2848° K = 0.497

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	1.113	grams
Mannite ($C_6H_8(OH)_6$)	1.113	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

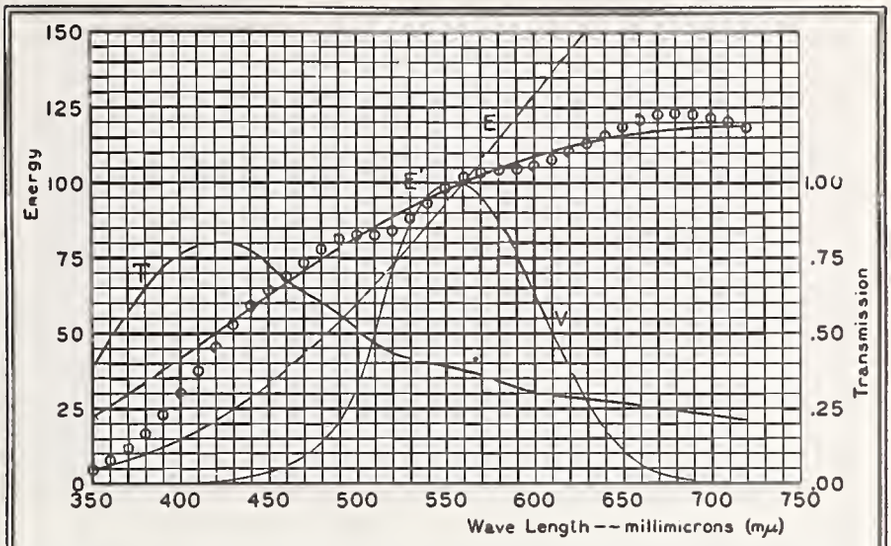
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	9.780	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	7.197	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848°K TO 4000°K CHART 22



λ (mμ)	T	E"*	E"/E'
350	0.390	5.0	.221
60	.483	8.0	.308
70	.571	11.9	.404
80	.650	17.0	.510
90	.717	23.2	.621
400	.763	30.1	.726
10	.793	37.6	.825
20	.807	45.4	.913
30	.799	52.8	.979
40	.771	59.3	1.020
450	.726	64.4	1.034
60	.678	68.6	1.035
70	.639	73.3	1.043
80	.602	77.9	1.049
90	.563	81.2	1.042
500	.514	82.4	1.010
10	.467	82.6	.971
20	.433	84.1	.951
30	.416	88.1	.962
40	.406	93.3	.988
550	.394	98.2	1.009
60	.379	101.5	1.015
70	.360	103.3	1.009
80	.340	104.1	.994
90	.321	104.6	.979
600	.305	105.5	.971
10	.295	107.7	.975
20	.287	110.3	.986
30	.280	113.0	.998
40	.273	115.5	1.009
650	.268	118.4	1.025
60	.262	120.9	1.040
70	.255	122.3	1.045
80	.248	123.1	1.046
90	.238	122.5	1.036
700	.229	121.6	1.026
10	.219	120.2	1.012
20	.209	117.9	.994

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2848° K
E' -- Relative Energy of 4000°K

E" (=T x E)* -- o o o o -- Relative Energy of 2848°K and Filter Combination

Light Transmission of Filter ** for 2848° K = 0.361

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	1.726	grams
Mannite ($C_6H_8(OH)_6$)	1.726	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

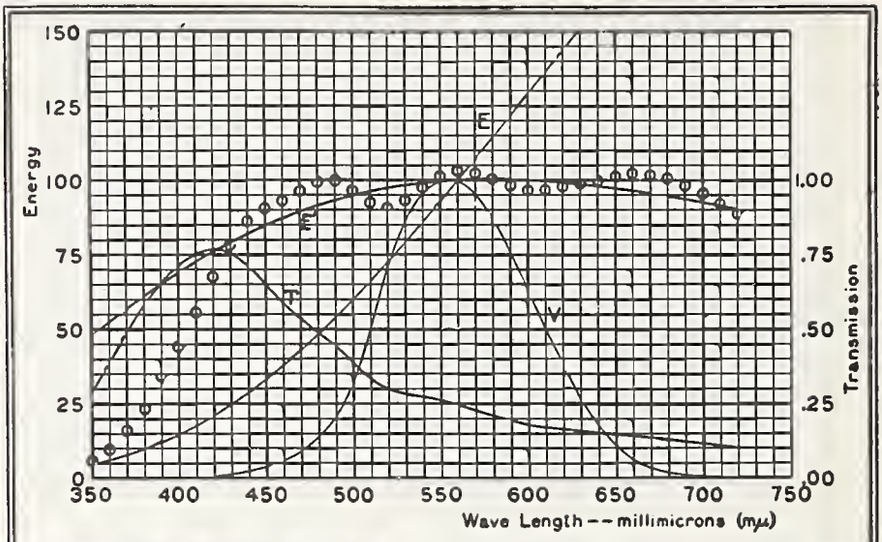
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	15.133	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	11.220	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848°K TO 5000°K CHART 24



λ (mμ)	T	E''*	E''/E'
350	0.296	5.9	.123
60	.390	10.1	.192
70	.486	16.0	.281
80	.577	23.7	.389
90	.658	33.4	.513
400	.714	44.1	.640
10	.750	55.8	.769
20	.767	67.7	.890
30	.754	78.1	.986
40	.714	86.1	1.048
450	.651	90.5	1.066
60	.586	93.0	1.064
70	.535	96.3	1.074
80	.489	99.2	1.082
90	.441	99.9	1.068
500	.385	96.8	1.019
10	.333	92.4	.959
20	.297	90.6	.929
30	.280	93.1	.946
40	.271	97.7	.987
550	.260	101.7	1.021
60	.246	103.2	1.032
70	.228	102.6	1.024
60	.209	100.5	1.002
90	.192	98.2	.980
600	.178	96.7	.966
10	.169	96.9	.973
20	.162	97.9	.987
30	.156	98.9	1.004
40	.150	99.9	1.020
650	.1462	101.4	1.043
60	.1414	102.2	1.062
70	.1355	101.8	1.068
80	.1292	100.7	1.066
90	.1215	98.1	1.050
700	.1144	95.3	1.032
10	.1073	92.2	1.011
20	.1002	88.6	.984

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2848° K
E' -- Relative Energy of 5000° K
E'' (=T x E)* -- o o o o -- Relative Energy of 2848°K and Filter Combination

Light Transmission of Filter ** for 2848° K = 0.231

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	2.599	grams
Mannite ($C_6H_8(OH)_6$)	2.599	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

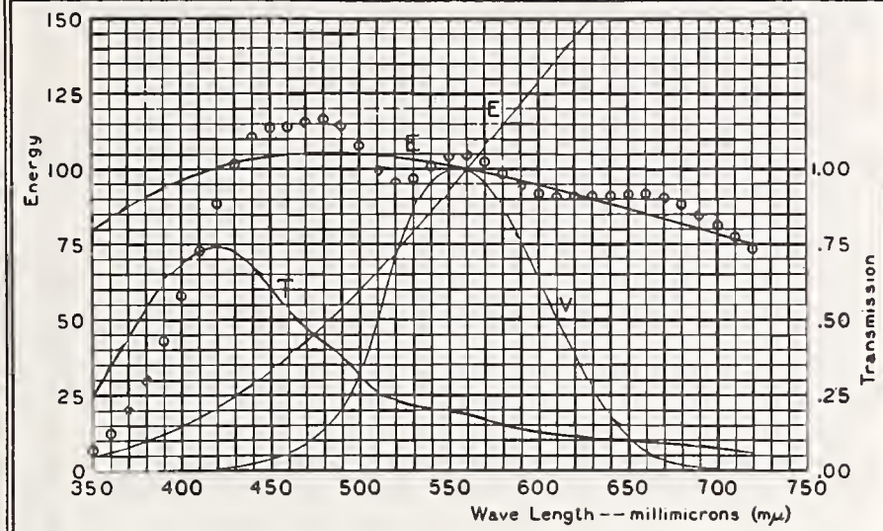
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	23.080	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	17.090	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E'' - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848°K TO 6000°K CHART 26



λ (mμ)	T	E"*	E"/E'
350	0.255	6.9	.086
60	.348	12.2	.145
70	.446	19.8	.226
80	.541	30.1	.331
90	.626	43.0	.459
400	.685	57.3	.594
10	.724	72.8	.738
20	.742	88.6	.881
30	.725	101.7	.997
40	.677	110.5	1.070
450	.604	113.8	1.093
60	.530	114.0	1.088
70	.474	115.6	1.099
80	.425	116.6	1.105
90	.374	114.6	1.089
500	.316	107.7	1.026
10	.265	99.6	.953
20	.231	95.2	.916
30	.215	96.7	.935
40	.206	100.9	.987
550	.197	104.3	1.031
60	.184	104.7	1.047
70	.168	102.3	1.036
80	.151	98.4	1.010
90	.1365	94.5	.984
600	.1248	91.6	.968
10	.1169	90.6	.974
20	.1110	90.6	.989
30	.1059	90.8	1.009
40	.1012	90.9	1.029
650	.0976	91.6	1.057
60	.0937	91.6	1.077
70	.0889	90.3	1.053
80	.0837	88.3	1.079
90	.0776	84.8	1.058
700	.0720	81.2	1.035
10	.0667	77.5	1.009
20	.0614	73.5	.978

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2848° K
E' -- Relative Energy of 6000°K
E" (=T x E)* -- o o o o -- Relative Energy of 2848°K and Filter Combination

Light Transmission of Filter for 2848° K = 0.173 **

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	3.188	grams
Mannite ($C_6H_8(OH)_6$)	3.188	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

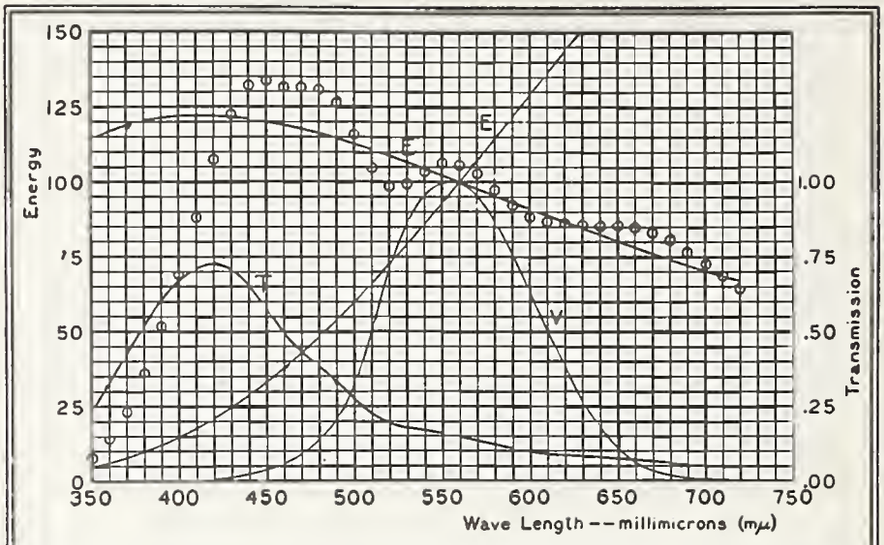
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	28.480	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	21.000	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848°K to 7000°K CHART 28



λ (mμ)	T	E"*	E"/E'
350	0.234	7.8	.069
60	.325	14.1	.121
70	.423	23.3	.197
80	.520	35.9	.299
90	.607	51.7	.427
400	.667	69.1	.567
10	.706	88.0	.720
20	.725	107.3	.878
30	.705	122.6	1.007
40	.652	132.0	1.089
450	.573	133.7	1.113
60	.494	131.6	1.106
70	.435	131.4	1.117
80	.384	130.6	1.125
90	.332	126.3	1.105
500	.275	116.1	1.031
10	.225	104.8	.948
20	.193	98.5	.906
30	.178	99.3	.932
40	.170	103.3	.988
550	.162	106.3	1.039
60	.150	105.9	1.059
70	.1356	102.5	1.048
80	.1205	97.3	1.018
90	.1073	92.2	.988
600	.0971	88.3	.970
10	.0901	86.7	.976
20	.0850	86.1	.993
30	.0806	85.7	1.015
40	.0765	85.3	1.035
650	.0735	85.5	1.066
60	.0700	85.0	1.087
70	.0659	83.1	1.093
80	.0616	80.6	1.088
90	.0565	76.6	1.062
700	.0519	72.6	1.035
10	.0476	68.6	1.004
20	.0434	64.1	.970

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2848°K
E' -- Relative Energy of 7000°K

E" (=T x E)* -- oooo -- Relative Energy of 2848°K and Filter Combination

Light Transmission of Filter ** for 2848°K = 0.1407

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	3.605	grams
Mannite ($C_6H_8(OH)_6$)	3.605	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

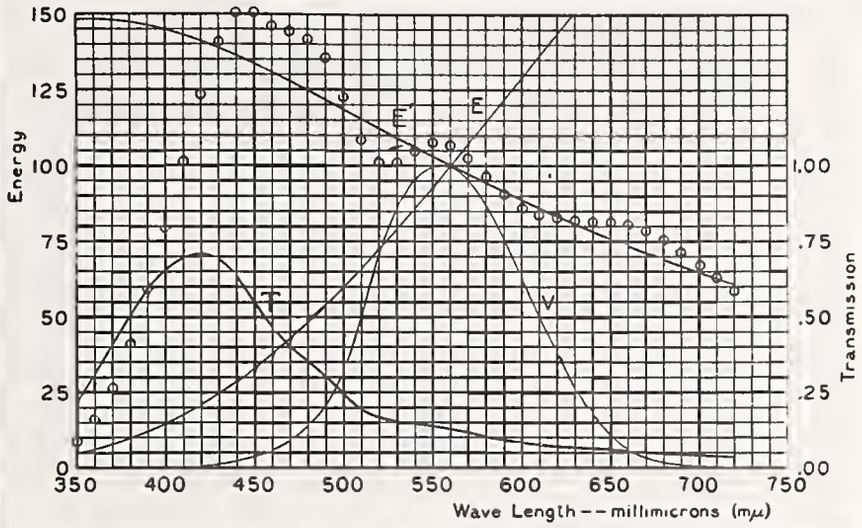
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	32.400	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	23.800	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848° K TO 8000° K CHART 29



λ (mμ)	T	E"*	E"/E'
350	0.221	8.7	.059
60	.312	15.9	.107
70	.409	26.4	.178
80	.506	41.0	.278
90	.594	59.4	.406
400	.654	79.5	.548
10	.693	101.3	.708
20	.712	123.5	.877
30	.691	140.8	1.017
40	.634	150.4	1.106
450	.550	150.5	1.128
60	.468	146.2	1.120
70	.407	144.3	1.132
80	.356	141.8	1.139
90	.304	135.4	1.115
500	.247	122.4	1.035
10	.199	108.7	.943
20	.168	100.8	.900
30	.155	101.1	.927
40	.1477	104.9	.990
550	.1402	107.7	1.045
60	.1292	106.7	1.067
70	.1157	102.4	1.056
60	.1019	96.4	1.024
90	.0899	90.5	.990
600	.0807	86.0	.972
10	.0744	83.8	.976
20	.0698	82.9	.996
30	.0660	82.2	1.019
40	.0623	81.4	1.041
650	.0596	81.3	1.074
60	.0566	80.5	1.097
70	.0530	78.4	1.102
80	.0493	75.5	1.097
90	.0449	71.3	1.068
700	.0409	67.1	1.037
10	.0372	62.9	1.004
20	.0337	58.7	.967

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2848° K
E' -- Relative Energy of 8000° K

E" (=T×E)* -- o o o o -- Relative Energy of 2848° K and Filter Combination

Light Transmission of Filter ** for 2848° K = 0.1211

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	3.916**	grams
Mannite ($C_6H_8(OH)_6$)	3.916	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

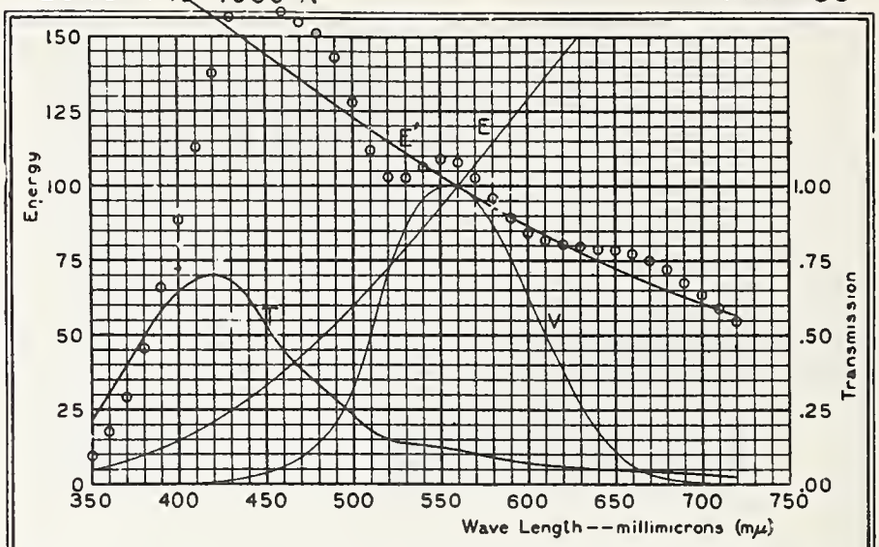
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	35.355	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	25.800	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848° K to 9000° K CHART 30



λ (mμ)	T	E"*	E"/E'
350	0.213	9.4	.052
60	.303	17.4	.097
70	.400	29.2	.166
80	.498	45.4	.263
90	.586	66.0	.391
400	.645	88.4	.535
10	.684	112.8	.700
20	.703	137.5	.875
30	.680	156.4	1.023
40	.620	166.1	1.118
450	.534	164.8	1.142
60	.449	158.4	1.132
70	.358	155.1	1.143
80	.336	151.2	1.151
90	.284	143.0	1.124
500	.229	127.7	1.038
10	.182	111.9	.940
20	.152	102.8	.894
30	.1390	102.6	.923
40	.1326	106.3	.991
550	.1255	108.8	1.050
60	.1152	107.4	1.074
70	.1026	102.5	1.062
80	.0897	95.7	1.028
90	.0786	89.2	.992
600	.0700	84.3	.971
10	.0643	81.8	.978
20	.0602	80.6	.998
30	.0567	79.7	1.023
40	.0534	78.6	1.045
650	.0509	78.4	1.080
60	.0482	77.3	1.105
70	.0450	75.0	1.110
80	.0416	72.0	1.104
90	.0377	67.6	1.073
700	.0342	63.3	1.041
10	.0310	59.0	1.006
20	.0279	54.5	.967

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2848° K
E' -- Relative Energy of 9000° K
E" (=T x E)* -- o o o o -- Relative Energy of 2848° K and Filter Combination

Light Transmission of Filter for 2848° K = 0.1082 **

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	4.154	grams
Mannite ($C_6H_8(OH)_6$)	4.154	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

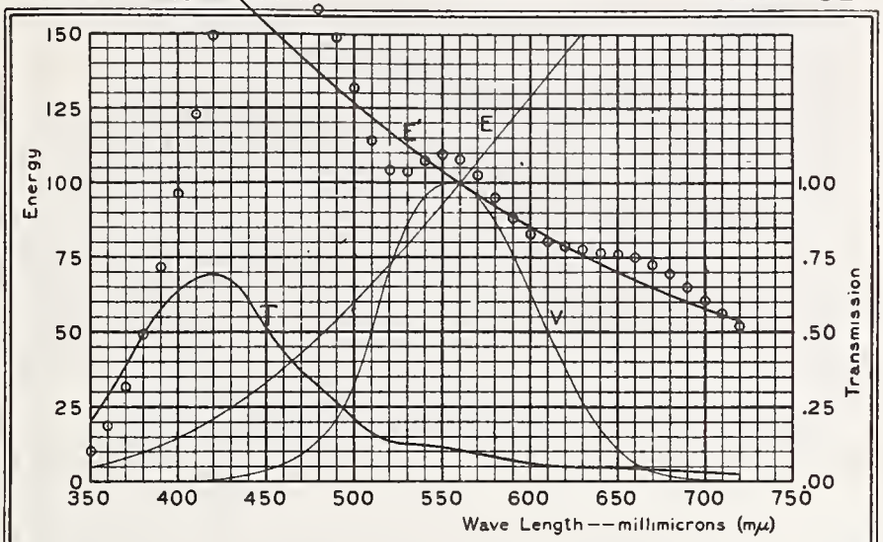
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	37.550	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	27.290	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848°K TO 10000°K CHART 31



λ (mμ)	T	E″ *	E″/E′
350	0.208	10.1	.048
60	.297	15.8	.091
70	.394	31.6	.157
80	.491	49.3	.253
90	.580	71.8	.380
400	.638	96.2	.526
10	.677	122.7	.693
20	.695	149.6	.875
30	.671	169.7	1.028
40	.609	179.4	1.127
450	.521	176.7	1.152
60	.435	168.5	1.141
70	.373	163.8	1.151
80	.321	158.7	1.160
90	.269	148.9	1.131
500	.214	131.7	1.039
10	.169	114.2	.937
20	.1401	104.1	.890
30	.1277	103.5	.920
40	.1216	107.2	.991
550	.1150	109.6	1.054
60	.1052	107.9	1.079
70	.0933	102.5	1.068
80	.0811	95.2	1.031
90	.0707	88.2	.994
600	.0627	82.9	.973
10	.0574	80.2	.978
20	.0535	78.8	.999
30	.0503	77.7	1.026
40	.0472	76.5	1.049
650	.0450	76.2	1.086
60	.0425	75.0	1.111
70	.0396	72.6	1.118
80	.0365	69.4	1.111
90	.0330	64.9	1.080
700	.0298	60.5	1.046
10	.0269	56.3	1.009
20	.0241	52.1	.969

Bureau of Standards. Davis-Gibson Filters

T — Spectral Transmission of Filter at 25°C
V — Relative Visibility Function
E — Relative Energy of 2848° K
E′ — Relative Energy of 10,000°K

E″ (=T x E)* — o o o o — Relative Energy of 2848°K and Filter Combination

Light Transmission of Filter ** for 2848° K = 0.0989

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	4.342	grams
Mannite ($C_6H_8(OH)_6$)	4.342	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

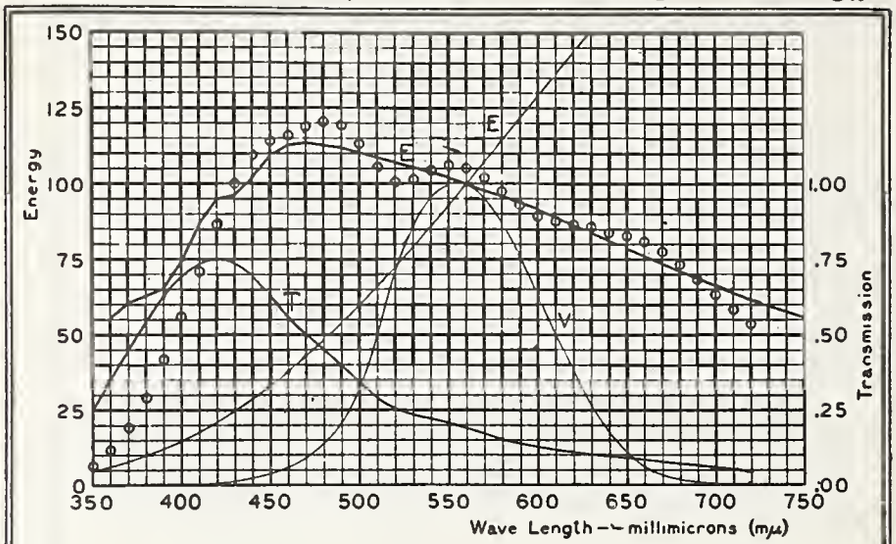
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	39.353	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	28.420	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E″ - E′ from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848°K TO SUNLIGHT OUTSIDE EARTH'S ATMOSPHERE CHART 32



λ (mμ)	T	E"*	E"/E'
350	0.258	6.7	——
60	.351	11.9	.215
70	.449	19.3	.320
80	.545	29.3	.463
90	.630	41.9	.641
400	.690	55.8	.753
10	.731	71.0	.816
20	.751	86.6	.911
30	.737	100.0	1.039
40	.694	109.6	1.075
450	.627	114.1	1.042
60	.558	115.9	1.031
70	.503	118.6	1.045
80	.454	120.5	1.067
90	.403	119.3	1.067
500	.344	113.1	1.026
10	.290	105.3	.969
20	.253	100.6	.940
30	.233	101.3	.961
40	.221	104.2	1.005
550	.208	106.1	1.039
60	.191	105.2	1.052
70	.173	101.9	1.041
80	.155	97.3	1.016
90	.1386	92.8	.991
600	.1259	89.3	.976
10	.1168	87.5	.982
20	.1096	86.6	1.000
30	.1031	85.5	1.018
40	.0967	84.0	1.034
650	.0913	82.8	1.055
60	.0854	80.8	1.063
70	.0788	77.4	1.055
80	.0720	73.5	1.035
90	.0646	68.2	.996
700	.0580	63.2	.958
10	.0520	58.4	.917
20	.0464	53.7	.872

Bureau of Standards. Davis-Gibson Filters

T — Spectral Transmission of Filter at 25°C
V — Relative Visibility Function
E — Relative Energy of 2848° K
E' — Relative Energy of Sunlight Outside Earth's Atmosphere
E" (=T x E)* — o o o o — Relative Energy of 2848°K and Filter Combination

Light Transmission of Filter ** for 2848° K = 0.180

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	3.125	grams
Mannite ($C_6H_8(OH)_6$)	3.125	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

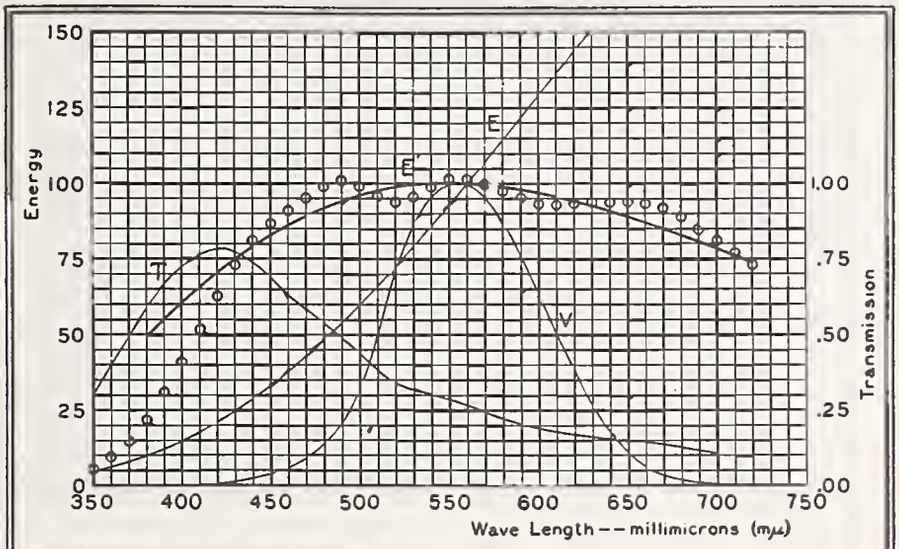
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	25.470	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	25.000	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848°K to ABBOT-PRIEST SUNLIGHT CHART 33



λ (mμ)	T	E"*	E"/E'
350	0.306	5.5	.—
60	.401	9.4	.—
70	.497	14.8	.—
80	.587	22.0	.442
90	.667	30.8	.559
400	.724	40.7	.674
10	.762	51.4	.786
20	.781	62.6	.891
30	.772	72.8	.972
40	.739	81.1	1.025
450	.686	86.7	1.045
60	.628	90.6	1.048
70	.581	95.1	1.063
80	.537	99.0	1.073
90	.489	100.7	1.065
500	.433	99.0	1.026
10	.379	95.6	.976
20	.339	93.9	.948
30	.316	95.6	.957
40	.301	98.8	.985
550	.285	101.1	1.008
80	.266	101.4	1.014
70	.244	99.9	1.002
80	.223	97.5	.983
90	.204	95.0	.967
600	.189	93.3	.962
10	.179	93.0	.972
20	.170	93.4	.994
30	.163	93.7	1.017
40	.155	93.6	1.032
650	.1489	93.8	1.058
60	.1419	93.2	1.071
70	.1338	91.4	1.078
80	.1254	88.8	1.072
90	.1157	84.9	1.054
700	.1069	80.9	1.036
10	.0985	76.9	1.013
20	.0904	72.7	.984

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2848° K
E' -- Relative Energy of *Abbot-Priest Sun*

E" (=T x E)* -- o o o o -- Relative Energy of 2848°K and Filter Combination

Light Transmission of Filter for 2848° K = 0.251 **

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	2.477	grams
Mannite ($C_6H_8(OH)_6$)	2.477	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

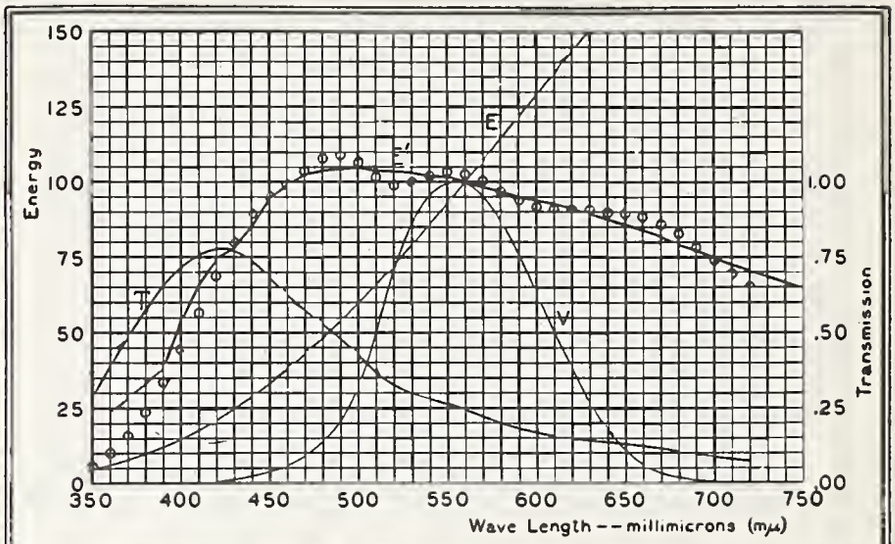
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	18.955	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	19.020	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848°K TO NOON SUNLIGHT AT WASHINGTON JUNE 21 CHART 34



λ (mμ)	T	E"*	E"/E'
350	0.290	5.8	
60	.385	10.0	.440
70	.481	15.9	.559
80	.574	23.7	.715
90	.657	33.5	.875
400	.715	44.4	.837
10	.755	56.4	.858
20	.775	68.8	.925
30	.768	80.0	1.035
40	.734	89.1	1.047
450	.680	95.1	1.011
60	.621	99.2	1.006
70	.572	103.7	1.015
80	.527	107.5	1.037
90	.478	108.8	1.044
500	.420	106.2	1.016
10	.364	101.7	.982
20	.323	98.9	.959
30	.299	99.8	.976
40	.281	102.2	1.004
550	.264	103.6	1.017
60	.243	102.8	1.028
70	.221	100.3	1.022
80	.200	97.0	1.004
90	.182	93.8	.989
600	.168	91.5	.976
10	.157	90.7	.980
20	.1492	90.5	.997
30	.1417	90.3	1.013
40	.1344	89.7	1.027
650	.1252	89.4	1.045
60	.1214	88.3	1.051
70	.1135	85.8	1.052
80	.1054	82.6	1.042
90	.0963	78.2	1.011
700	.0851	73.8	.983
10	.0804	69.4	.957
20	.0730	65.0	.922

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2848° K
E' -- Relative Energy of Noon Sunlight at Washington June 21
E" (=T x E)* -- o o o o -- Relative Energy of 2848°K and Filter Combination

Light Transmission of Filter ** for 2848° K = 0.231

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	2.671	grams
Mannite ($C_6H_8(OH)_6$)	2.671	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

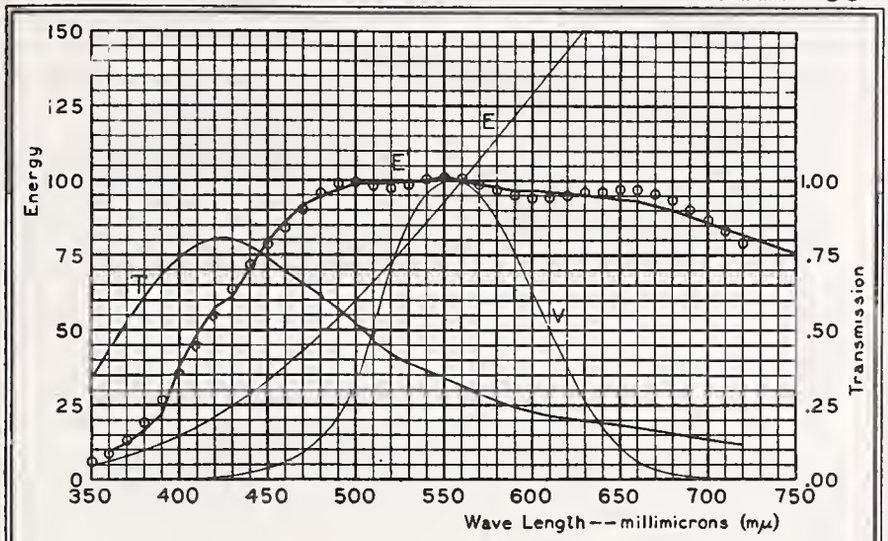
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	19.430	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	21.000	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2848°K TO NOON SUNLIGHT AT WASHINGTON DEC. 21 CHART 35



λ (mμ)	T	E"*	E"/E'
350	0.334	5.1	.——
60	.429	8.5	.936
70	.523	13.1	1.063
80	.612	19.2	1.144
90	.688	26.7	1.230
400	.744	35.1	.946
10	.782	44.4	.910
20	.803	54.1	.945
30	.801	63.4	1.040
40	.779	71.8	1.021
450	.740	78.6	.988
60	.695	84.3	.982
70	.656	90.2	.985
80	.617	95.5	1.012
90	.572	98.9	1.021
500	.518	99.6	1.006
10	.464	98.3	.996
20	.419	97.4	.982
30	.388	98.5	.987
40	.364	100.3	1.003
550	.340	101.2	.996
60	.314	100.6	1.006
70	.288	98.9	1.003
60	.264	96.9	.991
90	.244	95.1	.987
600	.227	94.1	.974
10	.216	94.4	.983
20	.207	95.1	.996
30	.198	95.9	1.008
40	.190	96.3	1.017
650	.183	97.0	1.033
60	.175	96.9	1.040
70	.166	95.5	1.045
80	.157	93.4	1.036
90	.1461	90.0	1.022
700	.1360	86.5	1.005
10	.1264	82.9	.992
20	.1167	79.0	.966

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2848° K
E' -- Relative Energy of Noon Sunlight at Washington Dec. 21
E" (=T x E)* -- o o o o -- Relative Energy of 2848°K and Filter Combination

Light Transmission of Filter for 2848° K = 0.302 **

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	2.194	grams
Mannite ($C_6H_8(OH)_6$)	2.194	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	13.060	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	16.890	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

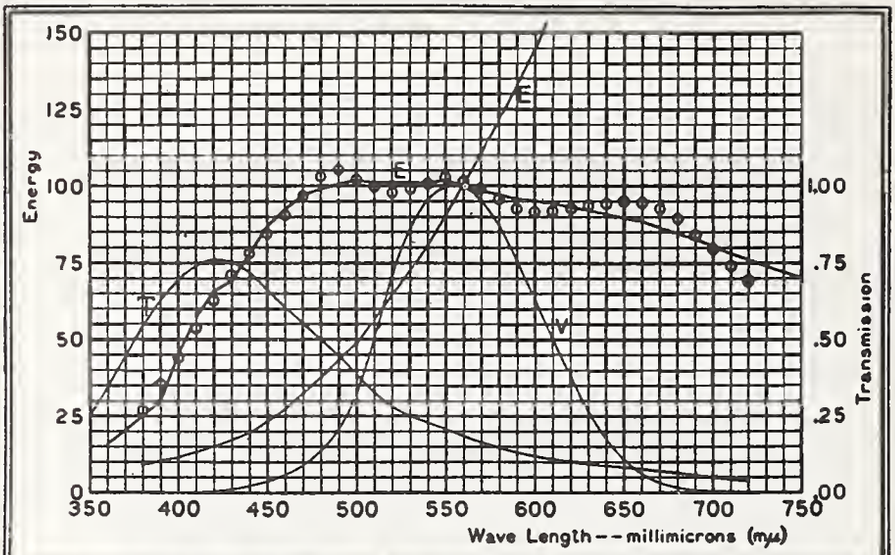
These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

518-743 O - 73 - 52

ACETYLENE TO MEAN NOON SUNLIGHT AT WASHINGTON. CHART 36



λ (mμ)	T	E″ *	E″/E′
350	0.250	—	—
60	.343	—	—
70	.442	—	—
80	.539	26.8	1.068
90	.628	35.1	1.168
400	.692	44.0	.974
10	.736	53.4	.931
20	.761	62.9	.955
30	.755	71.1	1.027
40	.721	78.0	1.003
450	.665	84.5	.973
60	.602	90.4	.980
70	.550	96.6	.997
80	.501	103.2	1.042
90	.447	105.1	1.045
500	.384	102.2	1.004
10	.325	99.6	.985
20	.280	97.8	.966
30	.251	98.9	.979
40	.229	100.9	1.000
550	.208	103.0	1.012
60	.186	101.4	1.014
70	.164	99.0	1.007
80	.1443	95.8	.986
90	.1280	93.3	.975
600	.1154	91.7	.964
10	.1064	91.8	.973
20	.0994	92.7	.995
30	.0931	93.7	1.017
40	.0870	94.1	1.034
650	.0818	94.9	1.058
60	.0761	94.5	1.067
70	.0697	92.4	1.069
80	.0633	89.3	1.054
90	.0563	84.1	1.017
700	.0501	79.1	.983
10	.0446	74.1	.950
20	.0395	69.1	.908

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of Acetylene
E′ -- Relative Energy of Mean Noon Sunlight at Washington
E″ (=T×E)* ---- Relative Energy of Acetylene and Filter Combination

Light Transmission of Filter ** for Acetylene = 0.170

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	3.287	grams
Mannite ($C_6H_8(OH)_6$)	3.287	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

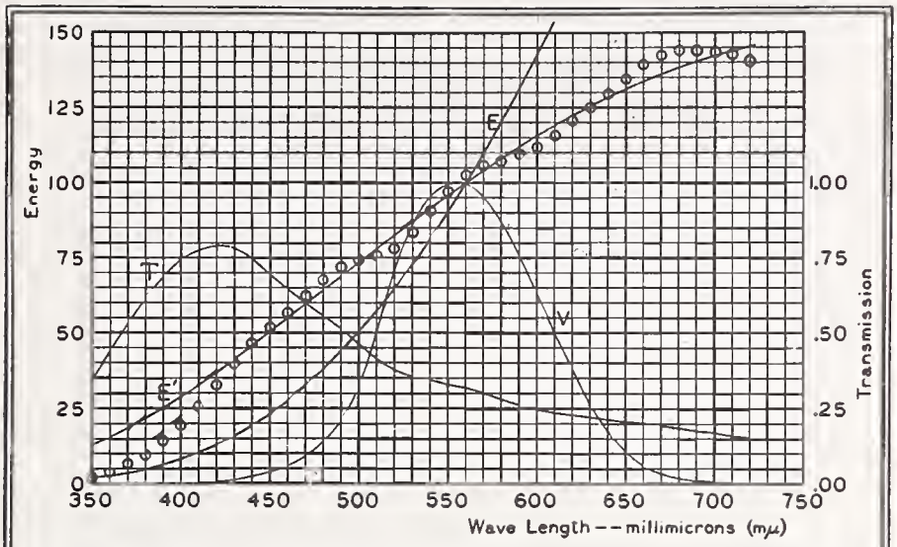
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	20.580	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	26.480	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E″ - E′ from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

2450°K to 3500°K CHART 37



λ (mμ)	T	E″ *	E″/E′
350	0.347	2.2	.169
60	.441	3.9	.248
70	.533	6.3	.341
80	.619	9.7	.448
90	.692	14.2	.565
400	.743	19.5	.677
10	.776	25.7	.788
20	.792	32.7	.892
30	.782	39.7	.971
40	.749	46.2	1.020
450	.698	51.7	1.038
60	.642	56.5	1.040
70	.598	62.0	1.050
80	.557	67.5	1.059
90	.514	72.0	1.051
500	.461	74.1	1.013
10	.411	75.3	.968
20	.375	77.9	.945
30	.357	83.2	.957
40	.347	90.3	.988
550	.335	97.1	1.014
60	.319	102.3	1.023
70	.300	105.6	1.015
80	.280	107.7	.997
90	.261	109.6	.980
600	.246	111.9	.969
10	.235	115.7	.973
20	.227	120.4	.986
30	.220	125.0	.999
40	.214	129.5	1.010
650	.208	134.6	1.028
60	.202	134.1	1.042
70	.195	142.2	1.046
80	.157	143.9	1.042
90	.178	143.8	1.027
700	.169	143.7	1.012
10	.160	142.6	.993
20	.151	140.6	.969

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2450° K
E′ -- Relative Energy of 3500°K
E″ (=T x E)* -- o o o o -- Relative Energy of 2450°K and Filter Combination

Light Transmission of Filter for 2450° K = 0.295 **

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	2.071	grams
Mannite ($C_6H_8(OH)_6$)	2.071	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B.

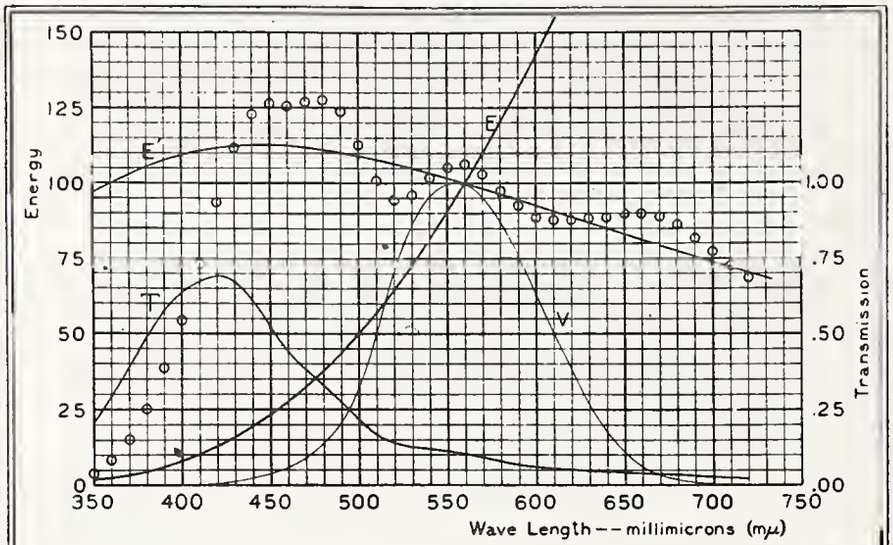
Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	18.045	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	14.000	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of E″ - E′ from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filtec combination

2450°K TO 6500°K CHART 38



λ (mμ)	T	E"*	E"/E'
350	0.206	4.2	.044
60	.295	8.4	.084
70	.392	15.1	.146
80	.489	25.0	.237
90	.578	38.6	.359
400	.637	54.4	.497
10	.676	73.0	.660
20	.695	93.7	.841
30	.672	111.4	.993
40	.611	123.0	1.093
450	.523	126.3	1.124
60	.437	125.6	1.119
70	.375	126.9	1.136
80	.322	127.3	1.147
90	.271	123.8	1.125
500	.215	112.8	1.035
10	.169	100.9	.937
20	.1395	94.5	.889
30	.1264	96.0	.916
40	.1196	101.6	.983
550	.1122	106.0	1.043
60	.1020	106.5	1.065
70	.0898	103.1	1.050
60	.0778	97.7	1.013
90	.0675	92.5	.979
600	.0598	88.7	.958
10	.0546	87.6	.965
20	.0508	87.8	.989
30	.0477	88.3	1.015
40	.0447	88.6	1.042
650	.0426	89.8	1.051
60	.0402	90.0	1.109
70	.0373	88.6	1.117
80	.0343	86.0	1.110
90	.0309	81.5	1.078
700	.0279	77.2	1.047
10	.0251	72.8	1.010
20	.0225	68.2	.971

Bureau of Standards. Davis-Gibson Filters

T -- Spectral Transmission of Filter at 25°C
V -- Relative Visibility Function
E -- Relative Energy of 2450° K
E' -- Relative Energy of 6500°K
E" (=T x E)* -- o o o o -- Relative Energy of 2450°K and Filter Combination

Light Transmission of Filter ** for 2450° K = 0.0917

FILTER FORMULA

A

Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	4.428	grams
Mannite ($C_6H_8(OH)_6$)	4.428	grams
Pyridine (C_5H_5N)	30.0	cc
Water (distilled) to make	1000.	cc

B

Cobalt Ammonium Sulphate ($CoSO_4 \cdot (NH_4)_2 SO_4 \cdot 6H_2O$)	38.930	grams
Copper Sulphate ($CuSO_4 \cdot 5H_2O$)	29.000	grams
Sulphuric Acid (sp. gr. 1.835)	10.0	cc
Water (distilled) to make	1000.	cc

These data are for a one centimeter layer each of solutions A and B in a double cell with three plates of borosilicate crown glass (refractive index, D line, = 1.51), each 2.5 mm thick.

* Adjusted to make sum of "E" - E' from 400 to 720 mμ equal practically to zero

** Factor to be used to multiply the candle-power of the light source to obtain the candle-power of the source-and-filter combination

XI. BIBLIOGRAPHY [21]

1. Ives, H. E., Color Measurement of Illuminants—A Résumé, Trans. I. E. S., **5,** pp. 189–207; 1910.
2. Priest, I. G., A Precision Method for Producing Artificial Daylight,[22] Phys. Rev. (2), **11,** pp. 502–507; 1918.
3. Priest, I. G., Standard Artificial Sunlight for Colorimetric Purposes,[22] J. Opt. Soc. Am. and Rev. Sci. Inst., **12,** pp. 479–480; 1926. (For correction of printer's errors, see also J. Opt. Soc. Am. and Rev. Sci. Inst., **14,** p. 70; 1927.)
4. Report of Optical Society of America Committee on Colorimetry for 1920–21, L. T. Troland, chairman, J. Opt. Soc. Am. and Rev. Sci. Inst., **6,** pp. 527–596; 1922.
5. Guild, J., A Critical Survey of Modern Developments in the Theory and Technique of Colorimetry and Allied Sciences, Proceedings of the Optical Convention, **1,** pp. 61–146; 1926.
6. Report of Optical Society of America Committee on Unit of Photographic Intensity, L. A. Jones, chairman, J. Opt. Soc. Am. and Rev. Sci. Inst., **12,** pp. 567–586; 1926.
7. Rapport de la Commission de l'"Optical Society of America" Sur l'Etalon Lumineux pour Sensitometrie, Proceedings, Sixth International Congress of Photography, Paris, 1925, pp. 60–70; Société Francaise de Photographie; 1926.
8. Proceedings, Sixth International Congress of Photography, Paris, 1925, pp. 23–24; Société Francaise de Photographie; 1926.
9. Frehafer, M. Katherine, and Snow, C. L., Tables and Graphs for Facilitating the Computation of Spectral Energy Distribution by Planck's formula, B. S. Misc. Pub. No. 56; 1925 (35 cents).
10. Proceedings, International Commission on Illumination, Sixth Meeting, Geneva, 1924, p. 67; Cambridge University Press; 1926.
11. Gibson, K. S., and Tyndall, E. P. T., The Visibility of Radiant Energy, B. S. Sci. Pap. No. 475; 1923 (15 cents); Trans. I. E. S., **19,** pp. 176–196; 1924.
12. Gibson, K. S., The Relative Visibility Function, Proceedings, Sixth Meeting, International Commission on Illumination, pp. 232–238. (See 10, above.)
13. Ives, H. E., Artificial Daylight, J. Frank. Inst., **177,** pp. 471–499; 1914.
14. Luckiesh, M., and Cady, F. E., Artificial Daylight—Its Production and Use, Trans. I. E. S., **9,** pp. 839–872; 1914.
15. Brady, E. J., The Development of Daylight Glass, Trans. I. E. S., **9,** pp. 937–952; 1914.
16. Gage, H. P., Daylite Glass, Sibley J. Engineering, **30,** pp. 247–250; 1916.
17. Macbeth, N., Color Temperature Classification of Natural and Artificial Illuminants, Trans. I. E. S., **23,** pp. 302–324; 1928.
18. Mees, C. E. K., and Sheppard, S. E., On the Sensitometry of Photographic Plates, Photo. J., **44,** pp. 282–303; 1904.
19. Sheppard, S. E., and Mees, C. E. K., Investigations on the Theory of the Photographic Process, Longmans, Green & Co.; 1907.
20. Mees, C. E. K., The Illuminating Engineer, London, **5,** pp. 79–80; 1912.
21. Wratten Light Filters, 6th ed. rev., 1924, Eastman Kodak Co.
22. Naumann, H., Ein Spektrales Weisslichtfilter, Zeits. f. Wiss. Photogr., **23,** pp. 303–319; 1924.
23. Sinden, R. H., Studies Based on the Spectral Complementaries, J. Opt. Soc. Am. and Rev. Sci. Inst., **7,** pp. 1123–1153; 1923.
24. Guild, J., A White Light for Colorimetry, Trans. Opt. Soc., **27,** pp. 122–124; 1925–26.
25. Hilger, Adam (Ltd.), Standard Artificial Daylight, Nature, p. **13,** January 14, 1928.

---

[21] National Bureau of Standards official publications may be obtained from the Superintendent of Documents, Government Printing Office, Washington, D. C., postpaid, at the prices given.

[22] The values of spectral energy distribution, E, introduced by Priest into the work of the colorimetry section, National Bureau of Standards, some years ago, now in current use and reproduced in Figure 3 of the present paper, are not exactly the same as proposed in his earliest publication cited above. The values given here are as shown in Figure 4 of the Standards Yearbook for 1927. They are given by:

$$E \propto E_{2,848} \sin^2 \alpha/2$$

where $E_{2,848}$ is energy per unit wave length as a function of wave length for the Planckian radiator at 2,848° K. for C_2 at 14,350μ°, and α is the rotation of the plane of polarization of light, as a function of wave length, per mm in crystalline quartz.

26. Davis, Raymond, B. S. Sci. Paper No. 409; 1921 (5 cents). Walters, Francis M., and Davis, Raymond, B. S. Sci. Paper No. 422; 1921 (15 cents). Davis, Raymond, and Walters, Francis M., B. S. Sci. Paper No. 439; 1922 (35 cents). Davis, Raymond, B. S. Sci. Paper No. 511; 1925 (15 cents). Davis, Raymond, B. S. Sci. Paper No. 528; 1926 (20 cents).
27. Coblentz, W. W., Distribution of Energy in the Spectrum of an Acetylene Flame, B. S. Sci. Paper No. 362; 1920 (5 cents).
28. Priest, I. G., Preliminary Data on the Color of Daylight at Washington, J. Opt. Soc. Am. and Rev. Sci. Inst., **7**, pp. 78–79; 1923.
29. Kimball, H. H.; The Distribution of Energy in the Visible Spectrum of Sunlight, Skylight, and the Total Daylight, Proc., Int. Com. on Illumination, Seventh Meeting, Saranac Inn, N. Y., 1928, pp. 501–515.
30. Priest, I. G., The Colorimetry and Photometry of Daylight and Incandescent Illuminants by the Method of Rotatory Dispersion, J. Opt. Soc. Am. and Rev. Sci. Inst., **7**, pp. 1175–1209; 1923.
31. Forsythe, W. E., and Worthing, A. G., The Properties of Tungsten and the Characteristics of Tungsten Lamps, Astrophys. J., **61**, pp. 146–185; 1925.
32. Hyde, E. P., Forsythe, W. E., and Cady, F. E., On the Distribution of Energy in the Visible Spectrum of an Acetylene Flame, Phys. Rev. (2), **13**, pp. 379–388; 1919.
33. Ives, H. E., and Kingsbury, E. F., Experiments with Colored Absorbing Solutions for use in Heterochromatic Photometry, Trans. I. E. S., **9**, pp. 795–813; 1914.
34. Ives, H. E., and Kingsbury, E. F., Additional Experiments on Colored Absorbing Solutions for use in Heterochromatic Photometry, Trans. I. E. S., **10**, pp. 253–258; 1915.
35. Report of Optical Society of America Committee on Spectrophotometry for 1922–23, K. S. Gibson, chairman, J. Opt. Soc. Am. and Rev. Sci. Inst., **10**, pp. 169–241; 1925.
36. Peters, H. H., and Phelps, F. P., Color in the Sugar Industry, B. S. Tech. Paper No. 338; 1927 (20 cents).
37. Brode, W. R., The Spectral Absorption of Certain Monoazo Dyes, B. S. Research Paper No. 47; 1929 (15 cents).
38. Trans. I. E. S., **20**, pp. 629, 632; 1925.
39. Mees, C. E. K., and Sheppard, S. E., New Investigations on Standard Light Sources, J. Roy. Photo. Soc., **50**, pp. 287–292; 1910.
40. Report of Optical Society of America Committee on Unit of Photographic Intensity, L. A. Jones, chairman, Proceedings, Seventh International Congress of Photography, London, 1928, pp. 152–161; W. Heffer & Sons (Ltd.), Cambridge; 1929.
41. Proceedings, Seventh International Congress of Photography, London, 1928, pp. 170–173.
42. Abbot, C. G., Fowle, F. E., and Aldrich, L. B., The Distribution of Energy in the Spectra of the Sun and Stars, Smithsonian Miscellaneous Collections, **74**, No. 7, Pub. No. 2714; 1923.
43. Priest, I. G., A New Study of the Leucoscope and its Application to Pyrometry, J. Opt. Soc. Am., **4**, pp. 448–495; 1920.
44. Gibson, K. S., Spectral Centroid Relations for Artificial Daylight Filters, J. Opt. Soc. Am. and Rev. Sci. Inst., **11**, pp. 473–478; 1925.
45. McNicholas, H. J., Equipment for Routine Spectral Transmission and Reflection Measurements, B. S. Research Paper No. 30; 1928 (20 cents).
46. Gibson, K. S., Direct-Reading Photoelectric Measurement of Spectral Transmission. J. Opt. Soc. Am. and Rev. Sci. Inst., **7**, pp. 693–704; 1923.
47. Gibson, K. S., McNicholas, H. J., Tyndall, E. P. T., Frehafer, M. K., with the cooperation of Mathewson, W. E., The Spectral Transmissive Properties of Dyes, B. S. Sci. Paper No. 440; 1922 (15 cents).
48. McNicholas, H. J., Use of the Under-water Spark with the Hilger Sector Photometer in Ultra-violet Spectrophotometry, B. S. Research Paper No. 33; 1928 (5 cents).
49. Gibson, K. S., Spectral Characteristics of Test Solutions Used in Heterochromatic Photometry, J. Opt. Soc. Am. and Rev. Sci. Inst., **9**, pp. 113–121; 1924.
50. George, E. F., The Absorption of Light by Inorganic Salts, Dissertation, Ohio State University; 1920.
51. Jones, H. C. and Anderson, J. A., The Absorption Spectra of Solutions, Carnegie Institute of Washington, Pub. No. 110; 1909.

52. Müller, E., Über die Lichtabsorption Wässeriger Lösungen von Kupfer- und Nickelsalzen, Ann. d. Phys., **12**, pp. 767–786; 1903.
53. Crittenden, E. C. and Taylor, A. H., An Interlaboratory Comparison of Colored Photometric Filters, Trans. I. E. S., **24**, pp. 153–207; 1929.
54. Priest, I. G. and Brickwedde, F. G., The Minimum Perceptible Colorimetric Purity as a Function of Dominant Wave Length with Sunlight as Neutral Standard, J. Opt. Soc. Am. and Rev. Sci. Inst., **13**, p. 306; 1926.
55. Priest, I. G., Apparatus for the Determination of Dominant Wave Length, Purity and Brightness, J. Opt. Soc. Am. and Rev. Sci. Inst., **8**, pp. 173–200; 1924.
56. Calzavara, E., L'Etalon lumineux pour Sensitométrie, Observations sur la composition du filtre coloré, Science et Industries Photographiques, p. 113; November 1, 1928.
57. Skogland, J. P., Tables of Spectral Energy Distribution and Luminosity for use in Computing Light Transmissions and Relative Brightnesses from Spectrophotometric Data, B. S. Misc. Pub. No. 86; 1929 (10 cents).
58. Forsythe, W. E., Color Match and Spectral Distribution, J. Opt. Soc. Am. and Rev. Sci. Inst., **7**, pp. 1115–1121; 1923.

WASHINGTON, March 21, 1930.

PHOTOGRAPHIC SCIENCE AND ENGINEERING
Volume 3, Number 6, November-December 1959

A Nomograph for Selecting Light Balancing Filters for Camera Exposure of Color Films

C. S. McCamy, *National Bureau of Standards, Washington, D.C.*

A nomograph has been designed for rapid selection of filters to adjust the spectral quality of the illumination to the spectral sensitivity of the film for color photography. The nomograph provides for a very large number of combinations of films and light sources in a simple form, permits the easy interpolation of new films, light sources, or filters, indicates which filters would nearly satisfy when an ideal filter is not available, indicates the general nature of abnormal combinations for special effects, and provides a convenient conversion from color temperature to reciprocal color temperature. The nomograph is based on the assumption that the color balance of color films, the chromaticity of the illumination, and the effect of filters can be characterized adequately on a scale of reciprocal color temperature and that the change of reciprocal color temperature by a given filter is a constant.

Generally, the most satisfactory color photographs are obtained if the spectral sensitivity of the film is balanced for the spectral quality of the illumination. However, it frequently becomes necessary to use film which is not properly balanced, either because there is no available film balanced for the light to be used or because it is not practical to change films once the camera is loaded. It is common practice in such cases to place a colored filter over the camera lens to adjust the spectral quality of the light to the spectral sensitivity of the film. There are many such filters commercially available to adjust a variety of light sources to the various color films. Manufacturers of films and filters publish tables indicating which filter to use with a given combination of film and light source. The nomograph described herein serves the same purpose as such tables but is more useful in that it provides for a very large number of combinations in a simple form; permits the easy interpolation of new films, sources, or filters without altering the rest of the chart; gives an approximate solution to the problem when there is no exact solution; and indicates the nature of abnormal combinations which might be employed for special effects.

The Nomograph and Its Use

The nomograph is shown in Fig. 1. Various light sources and their correlated color temperatures in degrees K are listed on the left, various filters and their correction effects in micro-reciprocal-degrees (mireds) are listed in the center, and various films are listed on the right, with the type of source for which they are balanced given in mireds. If a straight edge is placed on the points corresponding to a film and a light source, the edge crosses the center line at the point corresponding to the appropriate filter. The point on the filter scale corresponding to zero-correction indicates that no filter is needed since it lies between the points representing any film and the light source for which it is balanced.

Filters above the zero-correction point (+) are yellowish or orange in color; those lying below (−) are bluish. If the point corresponding to the filter actually used lies above the indicated ideal point, the resulting photograph will be generally yellower than normal. If the filter actually used falls below that indicated, the photograph will be bluer than normal. The extent of these effects can be visualized approximately before the picture is taken by viewing the scene momentarily through a filter having the same correction effect in mireds as the difference between the indicated ideal correction and that of the filter to be used on the camera. The correct algebraic sign for the viewing filter is obtained by subtracting the value for the ideal filter from that for the filter to be used on the camera.

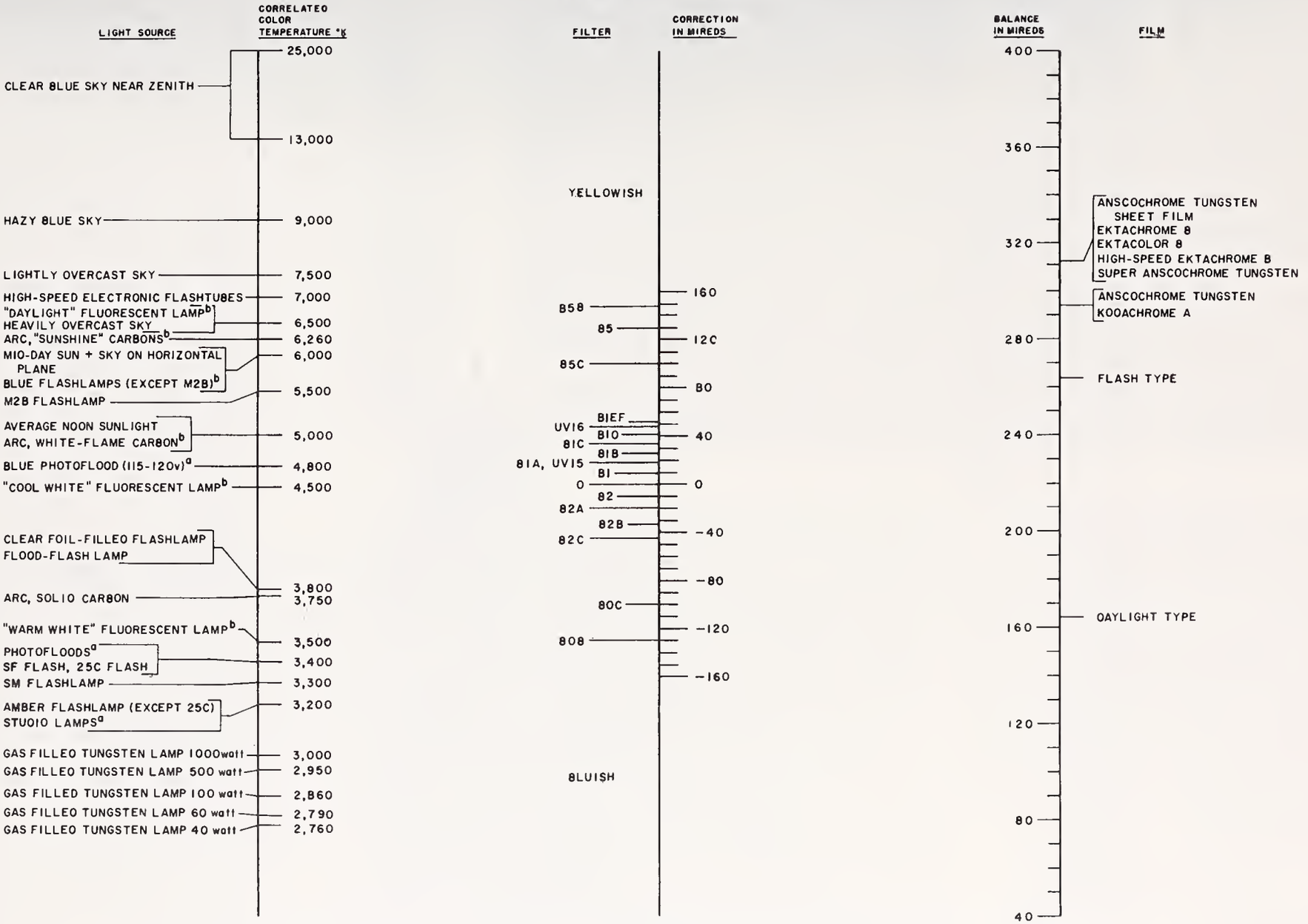
The nomograph may be used to convert color temperature to mireds and vice versa. The sources are listed on a graduated scale of color temperature, the films are listed on a scale of mireds, and any line through the zero-correction point connects corresponding points on these scales.

Sources having spectral radiation characteristics markedly different from those of ideal radiators are indicated with a superscript "b." In such cases, correlated color temperature is not an adequate specification of color for use in color photography but gives an approximation which may be useful.

The data on which Fig. 1 is based were taken from numerous sources, mostly the published recommenda-

Presented at the National Conference, Chicago, 27 October 1959
Received 28 September 1959.

COLOR FILTER NOMOGRAPH



a- The correlated color temperature of these lamps increases about 11°K for each volt increase in applied potential, in the neighborhood of 115v. As lamps are used, the correlated color temperature (at a given voltage) decreases, often from 50°K above to 50°K below the rated value during the life of the lamp.

b- Color temperature is only an approximate specification of these sources.

tions of manufacturers of films, lamps, and filters. The Wratten filter designations were used, for the most part, because of their general usage in photographic technology. Filters designated according to reciprocal color temperature may be selected by use of the filter correction scale graduated in mireds. Although the effects of the various filters could be computed from the known spectral transmittances, the values assigned in Fig. 1 were not obtained in this way. Rather, values were computed from the light source and film combinations for which the filters are recommended. For this purpose, daylight was assumed to have a correlated color temperature of 6100°K. No effort was made to achieve completeness in the nomograph shown, as it was intended merely to illustrate the method.

Theory

Although the colors of bodies are generally described and specified by multidimensional systems, it is sometimes practical to use a one-dimensional scale. One such is the scale of color temperature on which the color of a body is specified by the temperature of an ideal radiator which is of the same color. Since an ideal radiator does not assume all possible colors as the temperature is varied, it follows that not all colors can be described adequately in terms of color temperature. However, the colors of the illuminants generally used in color photography happen to match the colors of ideal radiators sufficiently well that color temperature is an adequate specification for the purpose. (There are a few exceptions to this rule, as noted earlier.) Priest[1] proposed the use of reciprocal temperature, instead of temperature itself, for specifying the chromaticity of incandescent illuminants and various phases of daylight and recommended the micro-reciprocal-degree Kelvin as the most convenient unit. He pointed out that a given small interval on this scale represents an equally perceptible difference in chromaticity regardless of the color temperature.

Gage[2] showed that if the logarithm of the transmittance of a filter is a linear function of wave number, the filter will alter the radiant energy distribution from that corresponding to Wien's equation for one temperature to that for another temperature. Fabry[3] and Langmuir and Orange[4] had suggested earlier that with a given filter designed to alter color temperature, the reciprocal color temperature is changed by the same amount regardless of the color temperature of the source. The present nomograph is based on this fact. It is assumed that the color balance of color films, the color of light sources, and the effect of real filters can be characterized adequately on a scale of reciprocal color temperature.

The accuracy of the method depends, in general, on the degree to which the assumed conditions are met. However, because filter correction values were assigned on the basis of recommended usage rather than on a theoretical basis, the answers obtained necessarily correspond with manufacturers' recommendations in the most-often encountered cases, those cases for which the filters were designed.

The mathematical validity of the nomograph can be shown by the following geometrical considerations. The three scales on the nomograph are parallel and are linear with respect to reciprocal color temperature. The scale for light sources, if given in micro-reciprocal-degrees, would increase downward, that for films increases upward, and that for filters has graduations half the size of the other two scales and passes through zero at the center. The filter scale lies equidistant from the other scales. Given a point A corresponding to some film, a line may be drawn through the "no filter" point D to a point B corresponding to the source for which the film is balanced. Another line may be drawn from A to some other source C, passing through point E on the filter scale. Since the scales are parallel and equidistant, DE is one half BC, but since the filter scale has graduations half as large as those on the light source scale, the number of mireds difference between the light sources is equal to the number of mireds change shown on the filter scale. The same kind of geometrical construction holds for all combinations of films and sources.

References

1. Priest, I. G., *J. Opt. Soc. Am.*, 23: 41 (1933).
2. Gage, H. P., *ibid.*: 46 (1933).
3. Fabry, C., *Trans. Illum. Eng. Soc.* (N.Y.), 8: 302 (1913).
4. Langmuir, I., and Orange, J. A., *Trans. Am. Inst. Elec. Engrs.*, 32: 1946 (1913).

UNITED STATES DEPARTMENT OF COMMERCE • Sinclair Weeks, *Secretary*
NATIONAL BUREAU OF STANDARDS • A. V. Astin, *Director*

Techniques for Ruling and Etching Precise Scales in Glass and Their Reproduction by Photoetching With a New Light-Sensitive Resist

Raymond Davis and Chester I. Pope

National Bureau of Standards Circular 565
Issued August 26, 1955

Contents

Techniques for Ruling and Etching Precise Scales in Glass and Their Reproduction by Photoetching With a New Light-Sensitive Resist

Raymond Davis and Chester I. Pope

Two methods are described for the making of very precisely graduated circles for theodolites. In one method the graduations and numbers are ruled in a wax resist coated on the glass circle by means of a ruling engine followed by etching. In the other method the circle is coated with a light-sensitive resist and exposed in contact with a glass master negative followed by etching. The requirements of precise circles and the mechanics of engraving the resist on both circles and reticules are discussed. The development of resist coatings and their application on the glass and the methods of glass etching are reviewed. A new mild etching solution containing phosphoric acid and hydrofluoric acid was developed. The synthesis of phenol-formaldehyde resin for use as a light-sensitive resist coating and the procedure for its use are given in detail.

1. Introduction

Instruments used for the measurement of angles are equipped with a reference scale graduated in degrees, minutes, and seconds, or in grads and its subdivisions. Sometimes only a part of a circle (an arc) is graduated, as in the case of the sextant. With other instruments, such as the theodolite, a complete circle is graduated. The graduated portion of a circle is commonly referred to as the limb. It has been the practice for many years to use brass with a silver inlay for the graduation of arcs and circles. In recent years aluminum alloys without the silver inlay have been used for this purpose, and also certain of the stainless steels have been recommended [1].[1] The precise graduation of circles on different materials, the techniques used, and the operation of the dividing engine have been discussed by Page [1].

A type of precision theodolite, having glass circles with the graduations etched in the glass, has come into use during the past 20 to 30 years. Only a few experimental models of these instruments have been made in this country. One of these experimental theodolites is shown in figure 1. The supply of such instruments is limited to a few European manufacturers. Theodolites equipped with glass circles have several advantages over the older types. They are smaller than conventional instruments of equal accuracy; they have greater compactness of design; and, in most cases, are appreciably lighter in weight. Moreover, illumination of the graduations of the circles and micrometer scales is superior, and the graduations have higher contrast with the background than is generally obtained with metal limbs. The most important feature is the simplicity of reading both circles with the micrometer microscope (magnifications up to $\times$ 40) located alongside the eyepiece of the telescope. Some instruments have separate micrometer microscopes for the horizontal and vertical circles. Each reading of the optical micrometer gives the mean of two diametrical points of the circle, thus automatically correcting for eccentricity of mounting of the circles [2, 3, 4, 5,

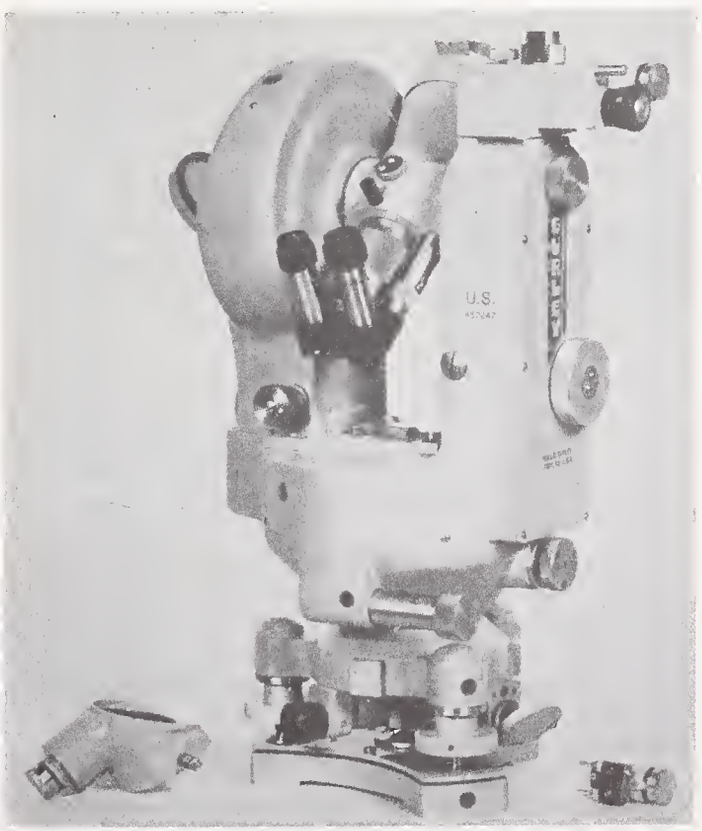


FIGURE 1. *Pilot model of a one-second theodolite.*

This theodolite was made for the U. S. Army Engineer Board by W. & L. E. Gurley, Troy, N. Y. These instruments were equipped with glass circles made at the National Bureau of Standards by the processes described in this paper.

[1] Figures in brackets refer to literature references at the end of this paper.

6, 7]. Theodolites with circles 3.5 to 5.5 inches in diameter, graduated to 20 minutes, may be read to 1 second and, in some cases, estimated to 0.1 second. Larger instruments with 10-inch circles, graduated to 2-minute intervals, may be read directly to 0.1 second [8]. It should be pointed out that the production of graduated circles having errors not exceeding the resolution of these optical micrometers poses a problem because such accuracy is not common. An error of 1 second of angle on a circle 4 inches (10 cm) in diameter would amount to a displacement of a graduation by 0.00001 inch (0.25 μ).

Graduated glass circles and scales also have been used in other fields. Dividing heads, indexing tables, cam-rise gages, and other units for machine shop use are now available, and it is expected that the use of glass circles and scales with etched graduations will find wide application in precise work.

Sometime ago the Army Engineer Board requested the National Bureau of Standards to make a limited number of precise glass circles and reticules and to develop a method for their production, with particular emphasis on a photographic method of reproduction that would facilitate quantity production. However, no light-sensitive coating was known that, under suitable exposure and development, would form a mask, or resist, that would also withstand the action of hydrofluoric acid used to etch the graduations into the glass. The project was undertaken and the immediate needs of the Engineer Board were met. However, the techniques used were not fully satisfactory, and further work was done toward a more satisfactory solution of the problem. The details of methods used by manufacturers for producing these circles have not been made generally available. However, examination of a number of graduated circles of European manufacture indicated that each is an original, separately coated with a resist, engraved, and etched.

The production of circles require the making of master circles, engraved on a circular dividing engine and etched in a more or less conventional manner; the development of a satisfactory photosensitive resist; and a contact-printing and etching process to produce replicate circles of the required accuracy. The production of reticules involve similar techniques and the design and construction of a ruling machine. Work on the project was divided between the NBS Length Section and the Photographic Technology Section. The former was responsible for engraving graduations and numbers on the master circles and testing the master circles and the replicas for accuracy [1]. The remainder of the work, that is, the development of resists and of techniques for applying them, design of the master circle blanks, construction of a pantograph for numbering the graduation lines on the master circles, production of replicas; and the production of precise reticlues, was the responsibility of the Photographic Technology Section.

This publication discusses the problems that were encountered and the techniques devised in the production of a limited number of experimental glass circles and reticules. Circles of the desired accuracy have been produced and the major problems have been solved. In quantity production of such circles, modifications in minor details are, of course, to be expected. The first six sections of this paper deal with the production of master circles, theodolite circles and reticules by engraving wax resists, etching the glass, and filling the etched graduations.

Section 7 deals with the production of etched-glass limbs, etc., from a master by photoetching techniques. Some of the observations and techniques are discussed at greater length than would be necessary if similar information were available in the technical literature.

2. Design of Graduated Circles and Master Circles

2.1. Graduated Circles

Two graduated circles, one horizontal and one vertical, are required for each theodolite. The horizontal circle is usually somewhat larger than the other because greater precision of measurement is required in the horizontal plane for geodetic triangulation, and, furthermore, a larger vertical circle makes the instrument more bulky, or less "streamlined." In one well-known instrument, the diameter of the horizontal circle is 3.5 inches, and the vertical circle is 2.7 inches, although both have the same 20-minute graduations and can be read to 1 second with the optical micrometers. The specifications for the graduated circles will, of course, be dictated by the design of the instrument. This design will include the position, length, and width of lines; position, orientation, and size of numbers, as well as the dimensions of the glass, type of etching, and other pertinent information.

All graduation lines should be of equal width, usually between 5 and 10μ. With one exception, in all of the circles of European manufacture examined, the degree lines were longer than the 20-minute subdivisions and were numbered. The numbers were about 0.005 inch high. The thickness of the glass was found to be about from ⅛ inch (3 mm) to slightly over ¼ inch (6 mm) for circles of from 2.5 inches (64 mm) to 4 inches (102 mm) in diameter. Figure 2 shows a much enlarged view of about 1°20′ of a 3.5-inch (89 mm) circle with 20-minute graduations. Two types of graduations were found: (1) more or less deep-etched

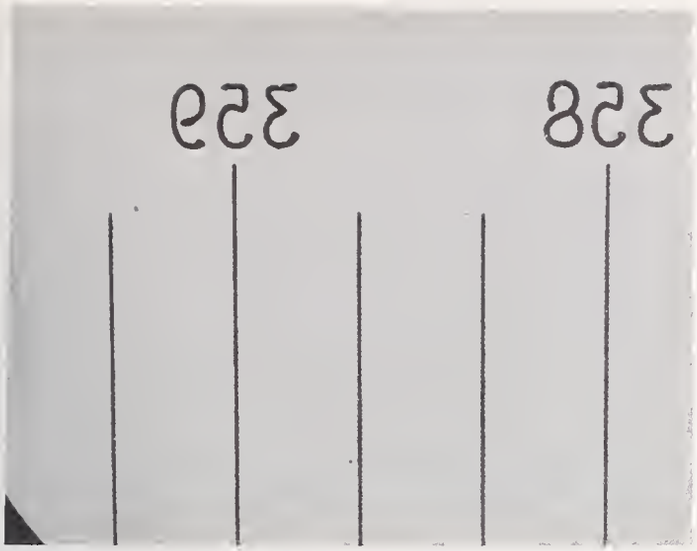


FIGURE 2. *Enlarged view of graduations on a master circle.*

Photographic reproductions were made from these master circles with graduations etched into the glass. These graduations, at 20-minute intervals, are subdivided with an optical vernier to 1-second intervals.

and filled lines on clear glass, (2) frosted, usually shallow-etched lines of fine texture, not filled, but mirror-silvered on the etched side. The silvering either covered the whole surface of the circle or only a narrow band, or ring, covering the graduations. Circles with silver over the etched graduations are viewed through the unsilvered side of the glass. Figure 3 shows a horizontal circle and its mount, disassembled, and also a vertical circle in its mount. Another type of circle examined was a small (2.5 inch) ring of relatively thick glass (¼ inch) with etched graduations on the outside edge of the ring. Obviously, this design would not be a satisfactory one for photographic duplication by contact printing.

One maker of theodolites uses a double line (two lines slightly separated) for each graduation. The circle may have been first graduated as usual with a single-point engraving tool, then the glass circle turned nearly 180 degrees on the dividing engine and a second set of lines engraved close to the first set. Alternatively, both graduations could have been made simultaneously with two tools set nearly 180 degrees apart. These double-line graduations should have the effect of reducing the dividing-machine errors by a factor of 2, provided both sets of graduations are used.

Selected plate or spectacle crown glass can be used for making blanks for the circles, and also for reticules, except, of course, where the specifications require the use of an optical glass, such as borosilicate crown. The blanks are annular rings cut from sheet glass with tube cutters. They are accurately ground on the outside diameter, then they are fitted into a special lathe collet, and the inside diameter is ground to exact size with a diamond grinding wheel on a tool-post grinder. The diameter of the hole is checked with a plug gage before removing the blank from the collet.

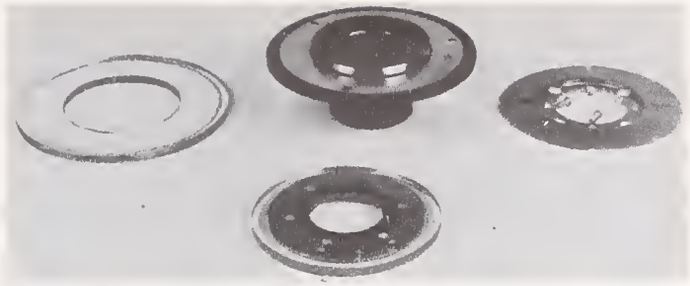


FIGURE 3. *Horizontal and vertical circles from a theodolite of European manufacture.*

In the foreground is the vertical circle with its clamp ring. In the back, on the left, is the horizontal circle. The band, or ring, just inside the edge of the circle is mirror silvering, which covers the graduated area. In the center is the mount for this circle, and on the right is a clamp ring used to hold the circle in position.

Following this, the front and back surface of the rings are ground and polished and the edges are beveled.

The quality of the glass surface required for reticules and etched scales is higher than that necessary for lenses. Very small imperfections, such as occasional pits or scratches, which are relatively unimportant on the surface of most optical components, are disturbing on surfaces containing scales or fiduciary marks because in use considerable magnification may be required for the observing or measuring operations for which they were designed. Further, if the tool runs into a pit or similar imperfection in the glass during the engraving, both the work and the tool are almost certain to be ruined.

As already stated, the physical dimensions of the glass for circles will be specified by the instrument design in which the completed circle or reticule is to be used. However, thick glass (about ¼ inch) is preferable to thin glass (for example, 1/10 inch). Uneven or local pressure in clamping a circle to the dividing engine table or in mounting it in the instrument in which it is to be used may cause distortion of the circle and thereby introduce several seconds of error. With thicker glass the errors from this cause will be smaller, other things being equal. In the instrument a thick glass has the further advantage that any dust or specks on the back side of the glass will be more completely out of focus.

2.2. Master Circles

The master circle consists of a thick clear glass disk, somewhat larger than the theodolite circle, with etched and filled graduations. Its purpose is similar to that of a photographic negative used for contact printing. (The details of contact printing from the master are discussed in sections 7.3 and 8.1.)

Glass blanks for master circles were made of either spectacle crown or borosilicate crown glass. Those for master circles 4 inches (102 mm) or smaller were made ½ inch (12.7 mm) thick and 6

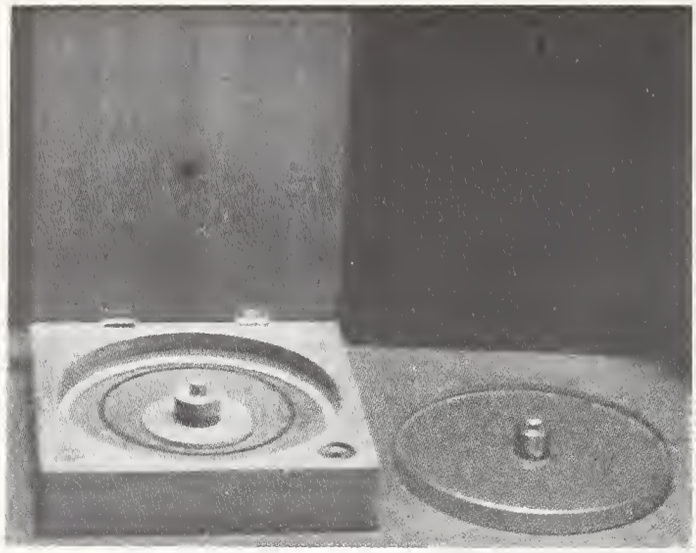


FIGURE 4. *Storage box and clamping plug for master circle.*
The box on the left is used for safely transporting and storing the master during all stages of its graduation and while not in use. On the right is a master with its centering and clamping plug in position.

inches (152 mm) in diameter. Figure 4 shows a glass master and its centering plug in its storage case. In the center of the glass disk a 60-degree conical hole was drilled and ground, having its larger diameter (about ¾ inch) at the top surface of the glass. The cylindrical steel centering plug is ¾ inch (19 mm) in diameter and about the same length, with a ¼-inch (6 mm) hole through its center. One end of this plug was machined to a 60-degree taper for about ⅜ inch (10 mm) of its length. The plug was fitted to the conical hole in the glass blank by lapping with suitable abrasives. Following this, both sides of the glass blank were ground and polished, the face side to a good optical plane free from scratches. Freedom from scratches is important, as very small ones not ordinarily noticed will fill about as well as the etched graduations, thus spoiling a master. The glass must also be free from seeds and bubbles in the vicinity of the graduations.

The steel plug is used to clamp the master blank, by means of a knurled-head screw through the hole in the plug, first to the table of the numbering machine for the engraving of the numbers and then to the table of the graduating engine for the engraving of the graduations.

The master blank is centered by means of a sensitive dial indicator in contact with the plug. Both centering and leveling are very important for accurate graduating. The centering and leveling screws on the table of the graduating machine provide this adjustment. The glass blank is leveled optically by lowering the ruling point until it is a few thousandths of an inch above the coated surface and by observing the ruling point and its image from the surface of the blank with a microscope at near grazing incidence (about 30 degrees). When the blank is level (at right angles to the axis of the spindle), the space between the ruling point and its mirror image remains constant while the table of the ruling or numbering machine is rotated. The back side of the glass blank is prevented from touching the metal table by a paper washer.

As mentioned in section 2.1, two types of graduations on the theodolite circles are in common use: (a) etched and filled lines on clear glass, and (2) frosted etched graduations, mirror-silvered on the etched side. When planning to engrave the master circle for contact printing, one must remember that the engraving on the master should be a mirror image of the graduations for type (1) circles. On the other hand, because type (2) circles are viewed through the glass, the graduations on the master will be the same as that on the finished theodolite circle.

3. Resists

3.1. General Composition and Characteristics

The term "resist" is commonly used in several industries, for example, calico printing, electroplating, photoengraving, and glass etching. A resist is a coating that protects the areas it covers from chemical, electrical, or physical action.

Several publications [9, 10, 25] describe processes used for the production of etched glass scales and reticules in a general way. However, no publication was found with specific information in regard to the composition of the resist. The subject is generally so lightly dealt with that several assumptions can be made: first, the whole matter is so simple that no difficulty should be expected; second, the composition and coating details are trade secrets; third, none of the resists is completely satisfactory. Probably both the second and third assumptions are true. Another factor that must be considered is that in some instances the materials (asphalts, waxes, resins, and the like) used in making resists are of unknown composition, identifiable only by trade name. Furthermore, natural materials frequently vary in properties according to the source, and synthetic materials may also vary because of differences in manufacturing and purifying processes. Finally, some materials cannot easily be obtained.

Resists used for engraving and etching of designs and scales in glass with hydrofluoric acid are usually made from a combination of such substances as asphalts, beeswax, paraffin, Burgundy pitch, and rosin. These materials are resistant to the action of hydrofluoric acid. Waxes are the basis for most resist formulations because of their inherently good ruling characteristics. Other ingredients (asphalts, resins) are added to improve this latter characteristic and to promote adhesion to the glass. Further modifiers some-

times used include oils, fats, varnish, and pigments. Most, if not all, of the ingredients mentioned are obtainable in several grades, which may differ in hardness, melting point, color, purity, and solubility, and one grade may not always be a substitute for another. This is particularly true with waxes and asphalts, because, in addition to the natural products, a wide variety of synthetic materials is available.

In planning resist formulations, compatibility of the components must be taken into account and if a solvent is used, all components (except pigments, if any) must be readily soluble in the solvent. Compatibility is, of course, related to the rule that materials of similar composition are miscible with one another. Certain waxes are incompatible with other waxes, resins, asphalts, or other material [11]. In some cases compatibility of a mixture is obtained by the introduction of a third material that is compatible with the other materials. Compatibility is a relative term expressed in different cases as "excellent", "good", or "poor." The subject of compatibility, solubility, and other properties of waxes is too involved to be more than mentioned here, and reference should be made to the books by Bennett [11] and Worth [12].

Most waxes are poor adhesives, although they are among the best resists for hydrofluoric acid. The adhesiveness can, in some cases, be improved mixing with other materials such as rosin, resins, pitches, and asphalts.

If given sufficient time, waxes will absorb hydrogen fluoride. For example, wax bottles (mainly ceresin), which have been used for hydrofluoric acid, will not lose all of the absorbed acid even after washing and boiling the wax in water.

The asphalts available have a wide variety of properties. In general, asphalt from any one geographic location is likely to be fixed in composition, appearance, melting point, etc. To a large extent natural asphalts for many commerical uses have been displaced by coal-tar pitches, petroleum asphalts, etc., derived from the oil industry. For specific information the reader is referred to Abraham [13], who gives a detailed history, description, and sources of various asphalts, both natural and synthetic.

Resists are sometimes applied to the glass hot, no solvents being used. More often they are dissolved in a solvent, and this solution is applied by several methods, which will be described in another section. Several solvents are available that will dissolve waxes, pitches, and asphalts. The solvent should have a rapid rate of evaporation. Benzene is considered one of the best because of its good solvent power and desirable rate of evaporation. Carbon tetrachloride is a good solvent, but it has a tendency to introduce moisture, thereby spoiling the bond between the glass and the resist. As both of these solvents are very toxic, care must be taken to avoid breathing the fumes. The use of a hood with an exhaust blower is strongly recommended. Turpentine and chloroform are sometimes mentioned in the literature as suitable solvents. Such resists coated on glass will not dry completely but will retain a small quantity of the solvent for an indefinite time; therefore, the ruling characteristics may change with time. Resists containing turpentine would be expected to increase in hardness with age.

In practice a thin layer of the resist is coated on the glass to be etched. The thickness of the layer depends to a large extent upon the type of work to be done. For very coarse work, where the line or pattern is observed without magnificaton, the resist can be as much as 0.1 mm in thickness, but for lines of 1 to 6 μ in width, it may be only about the equivalent of 1 or 2 wavelengths of light in thickness. There is no standard or gage to use for determining the proper thickness for a specific job; it will vary with the composition of resist, the etching techniques, and the user's experience and judgment.

The most important qualities of an ideal resist may be itemized as follows:

(1) Impermeability to either gaseous hydrogen fluoride or to hydrofluoric acid for a length of time at least equal to the time required for etching.

(2) Firm adhesion to the glass surface.

(3) Insensitivity to changes in the mositure content of the atmosphere (relative humidity).

(4) Ability to be engraved cleanly (that is, not be soft and stick to the tool, giving a line of variable width, or be hard and brittle, producing a ragged-edged line).

(5) Retention of its characteristics for at least a week after coating.

(6) Retention of its engraving characteristics over a moderate range of temperatures.

(7) Resistance to loosening of its bond with the glass under moderate thermal shock (10 deg).

The listing of these desirable characteristics will give the reader an idea of the nature of the difficulties commonly encountered with resists. The following is a brief discussion of difficulties, given in the same sequence as the desirable qualities were enumerated.

(1) *Permeability to hydrogen fluoride or hydrofluoric acid.* Practically all of the ingredients and solvents previously mentioned will take up several percent of hydrogen fluoride if given sufficient time; however, the rate of permeability is so low that no difficulty is likely to be encountered. Certain synthetic waxes, such as chlorinated naphthalenes, are so permeable to hydrogen fluoride that only a light etching can be made before the whole surface of the glass is mildly etched. Considerably better etchings can be obtained if hydrofluoric acid is used instead of the dry gas.

(2) *Lack of a good bond to the glass.* This is a common defect, and when present to a small degree is sometimes overlooked. After etching, it can be observed with a microscope as a gradual sloping of the surface of the glass toward the etched line, or as a rounding of the edge of the etched line

518-743 O - 73 - 53

at the surface of the glass. It should not be confused with the tapering sides of a deeply etched line. We have termed this defect "surface undercutting." Surface undercutting appears to be caused by a progressive lifting of the resist during the etching, probably a weakening of the bond due to stress during the ruling or engraving. Another form of undercutting is the normal removal of the glass in deep etching so that the resist layer extends beyond the side of the deeply etched line, the adjacent surface of the glass being free from etching. This is illustrated in figure 5.

(3) *Adverse effects of moisture.* Moisture on the surface of the glass or within the resist reduces the adhesion of the resist to the glass. Most resist materials are moisture-permeable or moisture-absorbent to some extent [11, p. 160]. The ruling characteristics of very thinly coated resists containing materials with appreciable hygroscopicity may be expected to change if subjected to an atmosphere of high humidity.

The difficulty of obtaining a strong bonding of the resist coating to the glass is probably due in part to a layer of moisture at the surface of the glass. Hubbard, Cleek and Rynders [14] have definitely established that glass surfaces exposed to water have a measureable hygroscopicity which varies with the kind of glass. It is very likely that this taking up of water is not merely an adsorption on the surface, but involves chemical combination within the surface layer of the glass. Indications are that the amount of water in glass surfaces follows changes in the relative humidity.

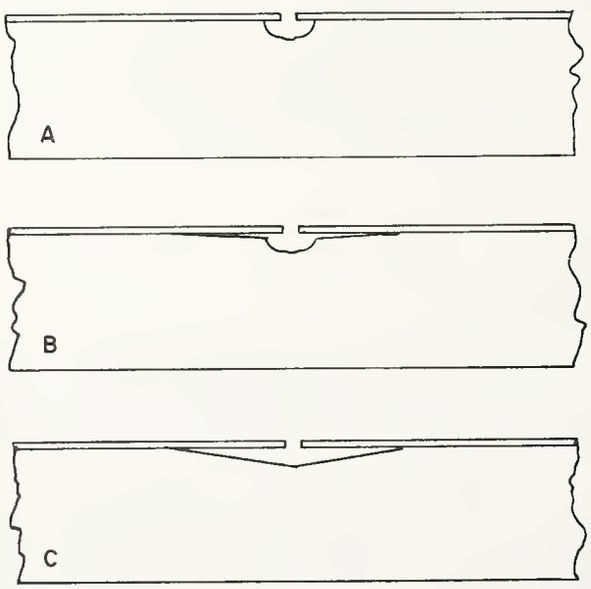


FIGURE 5. *Sketches showing three kinds of etched lines.*

The sketches were drawn as sectional views from visual impressions received with the microscope by observing normally to the wax resist coating. The top view, A, indicates a good etching of a deep line with normal undercutting of the wax resist. The resist coating retained its adhesion to the glass during the etching. The middle figure, B, indicates "surface undercutting." The adhesion of the wax resist loosened progressively during the etching, which resulted in a gradual slope of the surface of the glass toward the etched line. The bottom figure, C, represents an "excessive", or "abnormal", undercutting, resulting in a very widely etched area that slopes toward the etched line so excessively that the boundary of the etched line is sometimes difficult to find. Resists that have a poor bond to the glass will show this effect.

Heating a glass to drive off moisture is only temporarily effective; when the glass is allowed to fall to room temperature, the moisture is again taken up in some proportion to the relative humidity.

(4) *Unsatisfactory engraving or ruling quality of a resist.* This is of the greatest importance, since the quality of the final product, the etched lines, is dependent upon it. Faulty positioning of the engraving tool can, of course, spoil the work after a good resist coating has been applied. The point to be made here is that the resist layer must not be too soft or it will stick to the tool, giving a nonuniform line which changes in width somewhat periodically. Furthermore, at intersecting lines trails can be seen that show plainly which line was ruled last. On the other hand, a resist coating is sometimes too brittle. In this case the edges of the lines are ragged, giving a saw-toothed effect. In extreme cases flakes of resist are broken out, sometimes to a considerable distance (several line widths) from the line. Paradoxically, the same resist can sometimes show the effect of being both too hard and too soft; that is, the rulings show both effects to a small degree. This condition is generally due to a change of ruling speed; at low speeds, the resist tends to behave as if soft, and at high speed, it tends to behave as if brittle. Toughness is another characteristic sometimes encountered in resists. Such a resist can be satisfactorily engraved with a sturdy, properly shaped tool, but not equally well with a stylus or a delicately pointed round tool. Toughness can sometimes be remedied by addition of a plasticizer or a soft component.

(5) *Change of the ruling properties of a resist with time.* When a volatile solvent such as benzene is used it will evaporate rather quickly, but, while the resist may be ruled shortly after its application, it appears preferable to let it season a few hours or overnight before ruling. Resists containing drying oils behave like paint and are slow in drying: the best time for ruling has to be determined by trial. Compositions containing mainly soft asphalts often flow, causing the engraved lines to close up slowly; consequently, they cannot be used when the engraving requires several days to complete. Such a resist could be used when both ruling and etching can be completed in a short time, as in some reticule work.

(6) *Change of the ruling properties of a resist with temperature.* When resists contain a high percentage of wax the quality of rulings is markedly affected by the temperature at the time of ruling. It is, of course, desirable to have a resist that can be engraved well at room temperature (22° to 25° C), but this has proved to be a difficult problem. A resist that is too soft to be engraved satisfactorily at ambient temperature is sometimes much improved by lowering the temperature 2 to 5 deg. In general, compositions containing much soft wax can be engraved best at temperatures somewhat below 15° C. The addition of a compatible hard ingredient will usually improve a resist that is too

soft to rule at room temperature. However, too much of a hard material may make the resist sensitive to thermal shocks, resulting in some loss of adhesion, as explained in (7).

(7) *Loosening of the resist from the glass caused by large (10- to 20-deg.) changes in temperature.* The harmful effect of large fluctuations in temperature can be readily understood when one bears in mind that few, if any, of the wax-containing resist coatings are bonded very tightly to the surface of the glass, that the rigidity of these coatings increases with a decrease in temperature, and that the coefficient of expansion of waxes, asphalts, and resins is much larger than that of the glass on which it is coated. These factors combine to cause a reduction and, in extreme cases, a complete loss of the adhesion of the resist to the glass. Data on the coefficients of linear thermal expansion of waxes and asphalts are meager and widely scattered in the literature, but the examples in table 1 give some idea of their relative magnitudes. These data may not apply precisely to all samples of wax or asphalt.

TABLE 1. *Coefficients of linear thermal expansion*

Substance	Temperature range	Coefficient of expansion
	°C	
Plate glass [1]	0 to 100	0.089×10^{-4}
Beeswax (white) [2]	10 to 26	2.3
Paraffin [2]	16 to 38	1.3
Spermaceti [3]	25	3.2
Asphalt [4]		[5] 2
Halowax, [6] No. 1014		0.49
Halowax, No. 1013		.64
Halowax, No. 1012		.82

[1] Handbook of chemistry and physics, 34th ed., p. 1901 (Chemical Rubber Publishing Co., Cleveland, Ohio).
[2] Smithsonian physical tables, 7th ed., p. 219 (Smithsonian Institution, Washington, D. C.).
[3] A. H. Worth, Chemistry and technology of waxes, p. 489 (Reinhold Publishing Co., New York, N. Y., 1947).
[4] Handbook of chemistry and physics, 34th ed., p. 1906 (Chemical Rubber Publishing Co., Cleveland, Ohio).
[5] One-third of cubical expansion.
[6] H. Bennett, Commercial waxes, p. 118 (Chemical Publishing Co., Brooklyn, N. Y., 1944).

Use is made of the difference in expansion between glass and pitch in optical shops to remove reticule blanks from the pitch plate after polishing. This plate, covered with a number of reticule blanks, is placed in a refrigerator, and when sufficiently cold, the glass blanks will become detached from the pitch.

Some resists are much less sensitive to thermal shocks than others. For example, Halowax is much less sensitive to thermal shocks than beeswax; ceresin is about the worst in this respect; on the other hand, Halowax is much more permeable to hydrogen fluoride than ceresin. Except in extreme cases, no evidence of loosening of the bond to the glass is observable visually. Apparently the bond becomes so weakened that the stress caused by the engraving tool in cutting through the resist breaks the bond in the immediate vicinity of the engraving. Then upon etching, this area will exhibit "surface-undercutting."

In an attempt to secure stronger adhesions of wax resists to the glass, several substrata were tried but without much success. Waterproofing the glass with Amino Silane prior to coating gave lines with sharper edges. This method was used on circles; details are given in section 3.2.

3.2. Cleaning and Waterproofing the Glass Surface Before Applying the Resist

The glass must be cleaned with the same degree of care as in mirror-silvering. In the work under discussion, chromic acid cleaning solution (30 g of powdered potassium dichromate in 1 liter of concentrated sulfuric acid) was used. Some systematic method for handling the glass during the cleaning is needed, as otherwise the surfaces are likely to become scratched and nicked. For handling glass rings, tripods with glass legs were constructed from glass rod and rubber stoppers, as shown in figure 6. A ring blank was mounted on the tripod by squeezing the legs together and putting them through the hole in the ring. Upon releasing the pressure, the legs separate and hold the glass. The feet on the legs of the tripod prevent the ring blank from falling off and keep it from touching any flat surface on which the tripod may rest. After the cleaning solution had been washed off, the ring was placed under distilled water and swabbed lightly with absorbent cotton. After a final rinsing, it was placed on a whirler [2] and spun until dry. When dry, if the glass circle was to be waterproofed, it was placed in an oven at room temperature and heated to about 100° C. The circle was then removed from the oven and

[2] The whirler was similar to that described in footnote 4, except that the circle holder consisted of a disk of methyl methacrylate 3¾ inches in diameter and ½ inch thick. The top of the disk was reduced in diameter to slightly less than the inside diameter of the circle (3½ inches) for a depth of ⅛ inch, to form a shoulder upon which the circle blank rested during the whirling. The disk, fitted with a central bushing, was mounted directly on the motor shaft.

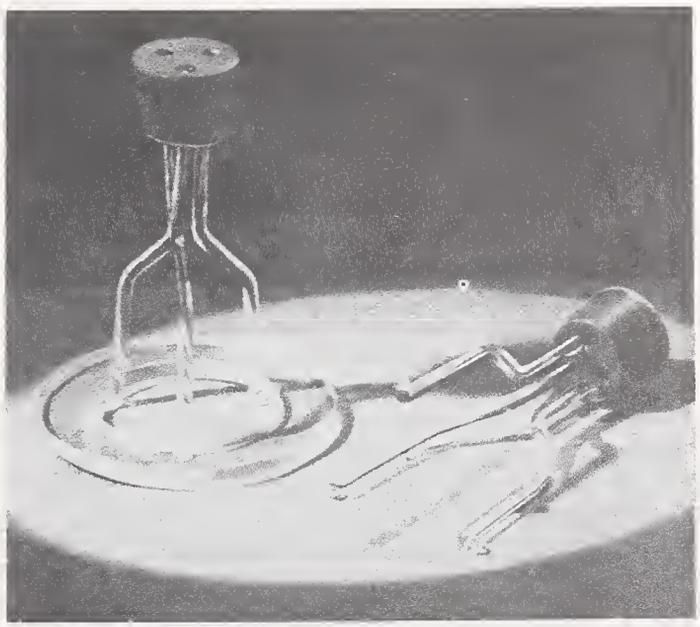


FIGURE 6. *Glass tripods.*

Tripods with rubber (stopper) heads and glass legs used for handling of rings during cleaning and other operations.

immediately waterproofed, by exposing it, while hot, to the vapors of Amino Silane[3] for 90 seconds. This was done by suspending the tripod and ring in a large desiccator at room temperature, with the surface of the glass about 3 inches above the surface of the Amino Silane. It is not considered necessary to waterproof the glass to get good results, however, it was found to improve the results. The magnitude of the improvement resulting from the use of Amino Silane is demonstrated in section 5.4.

3.3. Methods of Applying Resist Coatings

In the experimental work that preceded the adoption of specific formulas for resists, the coating of the glass was done by flowing the resist solution on the glass while it was spinning on a whirler;[4] this method is rapid and gives a coating of uniform thickness. However, it was concluded that withdrawing the glass from the resist solution at a constant rate was a superior method because less trouble was experienced with trash and lint.

The withdrawing, or "pull-out," method is no more difficult than the whirler method when used for reticule coating. A small (500-ml) jar or can, having a tight closure, half-filled with resist, is kept in readiness in an incubator at 45° C. The container is removed from the incubator, the top opened, and the reticule, in a holder attached to one end of a rod, is dipped into the resist and, after a moment or two, is withdrawn slowly with a uniform motion. When large pieces of glass are to be coated, it is advisable to use a mechanical withdrawing method to avoid coating irregularities. The mechanical method used in later work will now be described.

The mixture of beeswax, gilsonite, and Burgundy pitch (formula P–1), described in section 3.5, was filtered into a clean stainless-steel tank 3½ by 7 by 20 inches, having a flat top fitted with a plate-glass cover, as shown in figure 7. In use, the stainless-steel tank was immersed in a constant-temperature bath maintained at 45° ±1°C. A completely inclosed reduction-gear synchronous motor, mounted above the tank, was used to drive a V-pulley, which pulled the glass circle blank or master circle blank out of the resist solution. The glass blank was supported by a wire hook fastened to one end of a cord. On the other end was fastened a counterweight, the cord being supported by the aforementioned pulley, as shown in figure 7. Before being coated, the glass blank, supported on a tripod, was submerged in filtered

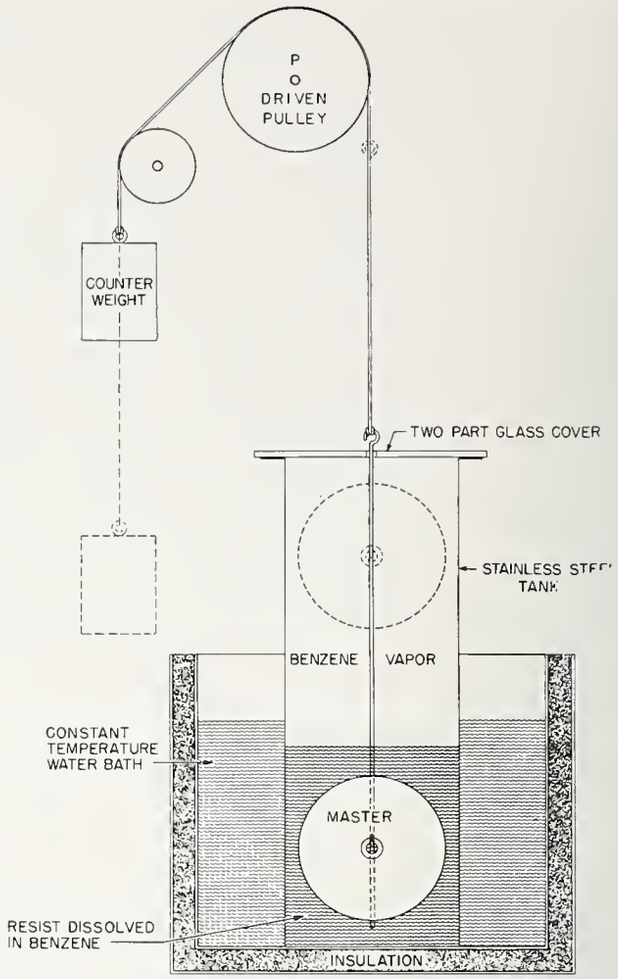


FIGURE 7. *Schematic view of the apparatus used for coating the glass with wax resist by the pull-out, or withdrawing, method.*

The glass, immersed in a solution of the wax resist, is withdrawn by the pulley, P, driven at a uniform rate. The speed of the pulley, the temperature of the solution, and its composition control the thickness of coating.

benzene, gently swabbed with cotton, and rinsed with filtered benzene; it was then immediately placed on the wire hook attached to the cord and completely immersed in the resist solution. A two-piece glass tank cover, each piece with a small notch cut in the center of one edge, was put in place with the edges butting (that is, with notched places together so that the supporting cord could pass through without touching the glass cover). The glass covers protected the tank from drafts and helped to maintain a high concentration of benzene vapor above the resist solution; these conditions were necessary to obtain uniform and reproducible coatings. After the glass blank had been completely immersed and the tank covered for a period of about 3 minutes, the motor was started and the glass drawn out of the solution at a uniform rate of approximately 1 foot in 80 seconds. The circle should not twist or turn when it is being drawn out of the resist solution as this may cause variations in the thickness of the coating. The motor was stopped just before the glass touched the tank covers. The tank covers were removed and the resist-coated glass was lifted out

[3] The chemical composition of Amino Silane is di-tert.-butoxydiaminosilane. See Miner et al, Ind. Eng. Chem. **39**, 1368 (1947).

[4] The whirler, used for coating the master with a resist, was made with four arms 6 inches long, 90 degrees apart, each with an adjustable stop. The position of the stops was adjusted to hold the glass in the center of rotation. The arms were attached directly to the shaft of a 110-volt, ⅛-horsepower, shunt-wound, direct-current motor. The motor was mounted with its shaft vertical; thus the glass rotated in a horizontal plane. The field of the motor was connected across the line. A slide-wire rheostat of about 150-ohm resistance was also connected across the line. One of the armature leads was connected to one side of the line, and the other armature lead was connected to the rheostat slider. By adjusting the position of the rheostat slider, any speed from zero to full motor speed (1,750 rpm) could be obtained.

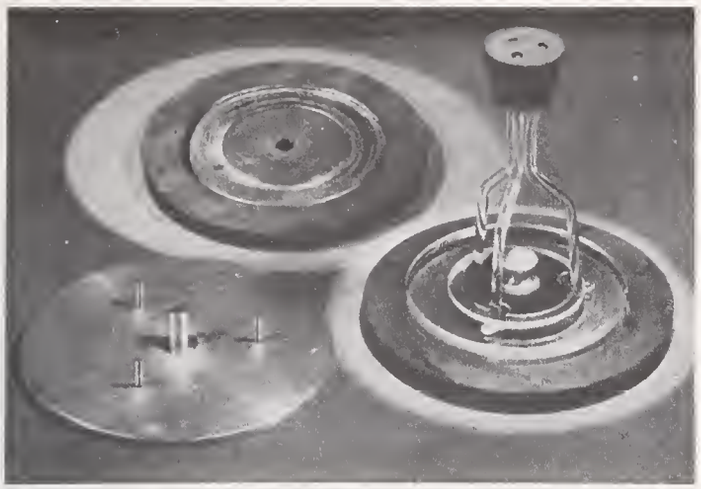


FIGURE 8. *Bakelite holders for resist-coated circles.*

The right-hand figure shows the method of placing the circle into the Bakelite holder preparatory to ruling, numbering, and etching. In the background, the circle is positioned in the holder. The metal plate with center peg and three pins, on the left, is used to remove the circle from the holder. After the etching is complete and the resist has been dissolved off the glass, the holder and circle are placed on the metal plate. The three pins fit into three mating holes in the holder and push the circle out of the recessed groove.

of the tank. This procedure gave very clean and uniform coatings. The resist solution may be used for an indefinite period before it needs refiltering. If the resist solution becomes too cold on standing without the temperature control, the beeswax will precipitate; it is then necessary to warm the tank of solution on the steam bath to about 60° C until it is certain that the beeswax has redissolved before replacing it in the constant-temperature bath. If the temperature of the resist solution rises to 50° C during coating, the resist becomes sticky; this is evidently due to an increase in the concentration of beeswax coating out on the glass. Because of the poisonous nature of benzene, the coating should be done in a well-ventilated room, preferably under a hood. Because benzene is highly flammable, care must be taken that no flames are near and that no sparking electric devices are operating in the immediate vicinity.

If the glass was a master, it was removed from the hook by grasping it at the edge of the glass with the fingers; if a circle, a tripod was used to take it off the hook [5] and to place it in a Bakelite holder designed to contain the circle during the engraving and etching. This Bakelite holder, ½ inch thick and 6 inches in diameter, is shown in figure 8. It had a recessed band milled out as a seat for the ring. The depth of this band, or groove, was about 1/16 inch less than the thickness of the ring. Consequently, the surface of the circle was slightly above the surface of the holder. The circle was fastened in this holder with 3 or 4 small pieces of wax, which were melted at the inside edge of the glass, using a short, electrically heated, U-shaped piece of resistance wire on a handle.

Resists are also applied to the glass hot without the use of solvents. This method is best suited for small-sized glass such as reticules. The resist ingredients are melted and well mixed (see formula H–1, section 3.5). The glass is heated in an oven or over a Bunsen burner, and a small amount of resist is spread over the surface of the glass while hot. A strip of paper somewhat wider than the reticule is placed on top of the melted resist, then slid off with a continuous motion. The glass is then reheated (not too hot) and finally placed on a warm, level surface to allow any fine streaks in the resist to level out. When cooled to room temperature, it is ready to rule. Some excellent etched lines have been made with this kind of resist.

[5] The area to be engraved on circles was about 1/8 inch from the outside edge. Consequently this edge should not be touched with the hand.

3.4. Procedure Followed in Developing and Testing Experimental Resists

The general procedure used for finding suitable resists was to make up solutions, composed of one or more of the ingredients in a suitable solvent, to coat them on glass, and to test the resist for its ruling and etching properties. The principal solid materials used in this work were:

Asphalts	Resins	Waxes
Asphaltines	Burgundy pitch	Beeswax.
Gilsonite	Congo resin	Candelilla wax.
Roofing asphalt	Coumar resin	Ceresin.
Trinidad asphalt	Pine tar	Halowax.[a]
Asphalt paints	Rosin	Ozokerite.
	Piccolyte resin	Paraffin.
		Spermaceti.

[a] Chlorinated naphthalene.

In addition to those listed above, a number of waxes synthesized from petroleum were tried, but these proved too insoluble to be of use. Carnauba and candelilla waxes are soluble only after heating, and Halowax [11, p. 117] is toxic.

Benzene was used as the solvent in most of the work, because of its good solvent power and its rapid evaporation from the resist coating. Other solvents tried were carbon tetrachloride, ethylene dichloride, toluene, turpentine, and xylene.

Most experimental formulations contained approximately 6 g of solid material per 100 ml of solution. Each formulation was coated on a 3¼- by 4-inch glass plate by pouring from 3 to 5 ml of the resist solution onto the center of the glass plate while it was rotating at from 400 to 600 rpm. The resist dries almost immediately, and, in most cases test rulings can be made in a few minutes, but it is preferable to test the resist after a lapse of a few hours and to retest it after storage overnight to note any changes caused by aging. If the resist had good characteristics, additional tests were made after further aging.

Test lines were engraved in the resist with the reticule-ruling machine. The ruled patterns generally consisted of 6 or 8 lines about 1 cm long, spaced about 100 μ apart; a second set of lines was ruled through the first set of each pattern at an angle of 90 degrees. Each such pattern was

numbered (by hand scribing) for identification. The patterns were examined with a microscope (16-mm objective and a ×10- or ×15-eyepiece) for quality of lines, with particular attention to edges, the beginning and ending of lines, and trails at intersecting lines. After this examination, the individual patterns on the glass were treated separately with experimental etching solutions or with dry hydrogen fluoride. The etched patterns were again examined with the microscope, the observations being compared with those previously made. Inspection also was made for any new defects and for surface-undercutting. This latter defect can be seen, when present, by adjusting the microscope in and out of focus on the surface of the glass at and near the edge of a line. If the surface of the glass slopes toward the etched line and the edge is not sharp, surface-undercutting is present.

3.5. Formulas Adopted

Some very high-quality lines were ruled and etched on masters in the beginning of this work with a Halowax and gilsonite resist. This resist was coated on the master with the whirler at 600 rpm. The etching was done with dry gas, using the wax pot described in section 5.2. The particular sample of Halowax used was very old, and its source could not be identified. This resist, thinly coated (about 0.00001 inch thick) had one fault; it was somewhat permeable to the hydrogen fluoride, and the glass surface was faintly etched. However, this mildly etched surface was easily removed by polishing. Polishing of the glass was necessary when lead sulfide was used to fill the lines and is described in section 6.1. The following resist is equivalent to that mentioned:

Resist Formula W–1

Halowax No. 1014	33.5 g.
Gilsonite	22.5 g.
Burgundy pitch	3 g.
Benzene	1,000 ml.

This resist can be used with liquid etch No. 2, section 5.3, to etch lines considerably deeper than can be obtained with gas etching.

Beeswax is one of the most commonly used substances of those investigated. It bonds to the glass somewhat better than most waxes and is not attacked by the etching solution, but it is too soft to use by itself. Beeswax is compatible with asphalts and resins. Gilsonite, one of the asphalts of fixed composition is less likely than other asphalts to vary appreciably from one lot to another.[6] It was found that a mixture of about 40 percent of white, bleached beeswax and 60 percent of gilsonite engraved well when the temperature was not higher than 22° C, but was improved by the addition of about 5 percent of Burgundy pitch. The Burgundy pitch used was soluble in ethyl ether, ethyl alcohol, benzene, acetone, 10-percent sodium carbonate solution and 10-percent borax solution. Hard pine tar may be substituted for the Burgundy pitch but it not quite as satisfactory. The following resist formula has a concentration of ingredients suitable for coating the glass by the whirler method:

Resist Formula W–2

Gilsonite	34 g.
White bleached beeswax	21 g.
Burgundy pitch	3 g.
Benzene	1,000 ml.

This resist was not attacked by the etching acid, could be engraved without sticking to the tool, did not become brittle on storage for several weeks, and bonded well to the glass.

A 6-inch disk of glass, coated at 600 rpm with resist W–1, had a coating about 0.00001 in. thick. Resist solution W–2 gave a coating thickness of about 0.00002 inch under the same conditions.

The thickness of coating may be adjusted to individual needs by changing the proportion of solvent to solids of the resist formula or by changing the speed of whirling. Dilution of solutions and higher whirling speeds produce thinner coatings. The size of the glass to be coated is another factor to be considered. Small glasses require higher speeds than large ones. For example, ½-inch-diameter reticules were coated at 1,800 rpm, whereas 6-inch glass was run at 400 to 600 rpm with the same resist.

A number of theodolite circles were coated with a modification of this resist by drawing them from the solution at a uniform rate, as explained in section 3.3. The formula for this modified resist is:

Resist Formula P–1[a]

Gilsonite	103 g.
White bleached beeswax	63 g.
Burgundy pitch	9 g.
Benzene	1,000 ml.

[a] The temperature of the solution at the time of use should be 45° to 46° C (see section 3.5).

The temperature at the time of ruling has a decided effect on the ruling characteristics of any wax resist. A temperature of 25° C or above will cause the P–1 resist to become soft and stick to the tool, and a poor engraving will result; at 18° to 20° C excellent results were obtained. Resist P–1 gave a coating about 0.00004 inch thick.

In reticule work the coating, ruling, and etching is usually completed in a short time, an hour or two at most; consequently, requirement (5) in section 3.1 can be relaxed. Additional formulas, which can be used for reticules, are given hereinafter. Formulas W–3 and W–4 are intended for the whirler method of coating.

Resist Formula W–3

Piccolyte S–85	27 g.
White bleached beeswax	7 g.
Burgundy pitch	27 g.
Benzene	1,000 ml.

[6] Any asphalt sold in the powdered form is almost certain to be gilsonite. The gilsonite (fusion point 125° to 135° C) was obtained in lump form.

Resist Formula W–4

Gilsonite	30 g.
Rosin	20 g.
Roofing asphalt	10 g.
Copper oleate	10 g.
Beeswax	2 g.
Benzene	1,000 ml.

The following formula is intended for the pull-out or dipping method of coating:

Resist Formula P–2 [a]

Gilsonite	87 g.
White bleached beeswax	58 g.
Burgundy pitch	7 g.
Copper stearate	7 g.
Benzene	1,000 ml.

[a] Coating temperature, 45° C.

A hot method has the advantage of ease of application of the resist with a minimum of equipment. On the other hand, it is more difficult to obtain uniform thickness. Further, the variety materials that can be used is somewhat restricted because some materials that are compatible when dissolved in a solvent may not be compatible when melted together. An example is the mixture of asphalt and wax prescribed for resist formula W–2. Glass to be coated by the hot method is heated over a Bunsen burner or, preferably, in an oven to a temperature somewhat higher than that required to melt the resist. A small amount of resist is applied and spread over the surface. Next, one end of a strip of paper somewhat wider than the glass is laid on the resist and slid off with a continuous motion in the direction of the plane of the glass. Following this, the glass should be reheated and placed on a level surface, so that any small streaks, created when the paper was pulled off, can level out. The following formula makes a satisfactory resist for hot application:

Resist Formula H–1

Gilsonite	12.5 parts by weight.
White bleached beeswax	4.5 parts by weight.
Ceresin	58 parts by weight.
Stearic acid	25 parts by weight.

4. Ruling and Engraving the Resist

4.1. Engraving Engines and Machines

Engraving wax resists is a step in the process used in the production of a wide variety of articles, ranging from metal nameplates for manufactured articles, on the one hand, to very precise scales read with high magnification, on the other. Obviously, graduations to be read with little or no magnification need not have the accuracy of placement or the quality of line necessary for scales intended for measurements of the highest possible precision. Jones and Hammond [9] have described machines and processes used for nameplates and scales for both general and precision use. Their article gives an interesting survey of the manufacture of etched metals, glass scales, etc., and includes a description of a number of instruments, including the circular dividing engine, the linear dividing engine, and the pantograph.

In the work described in this paper a circular dividing engine, which has been described in detail by Page [1], was used to graduate the master circles and theodolite limbs; a pantograph of original design was built and used for numbering graduations; and a specially designed engine was constructed and used for ruling reticles. The reticule-ruling machine is shown in figure 9. Figure 10 shows a typical pattern of the precisely ruled lines obtained with this machine; the accuracy of placement of the lines was ± 1 μ. As discussed above, the reticule-ruling machine was also used for testing experimental formulations of resists. When many reticules of the same design are to be made and the utmost in precision is not required, use of the pantograph is to be recommended. With a pantograph of suitable design and a large metal template engraved with the desired pattern, 10 or more recticules can be engraved at one time; curved lines, circles and characters can be made with equal facility.

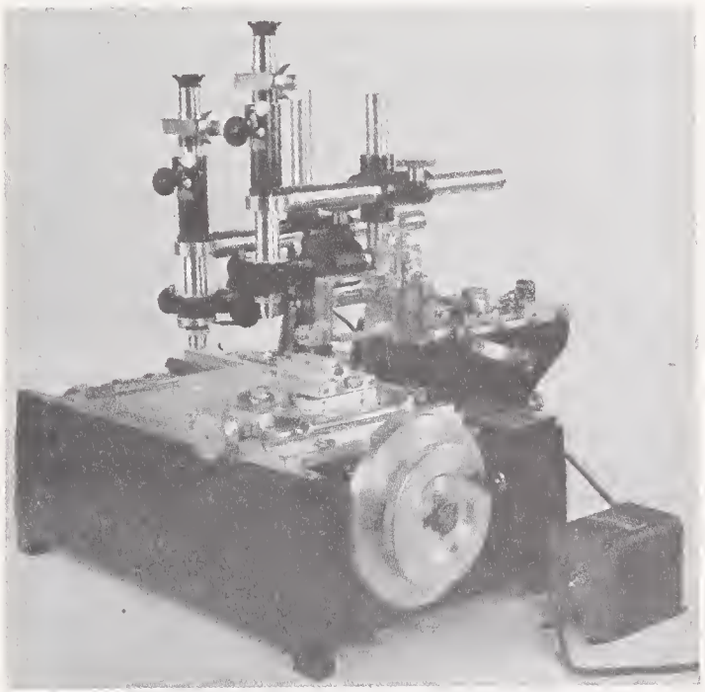
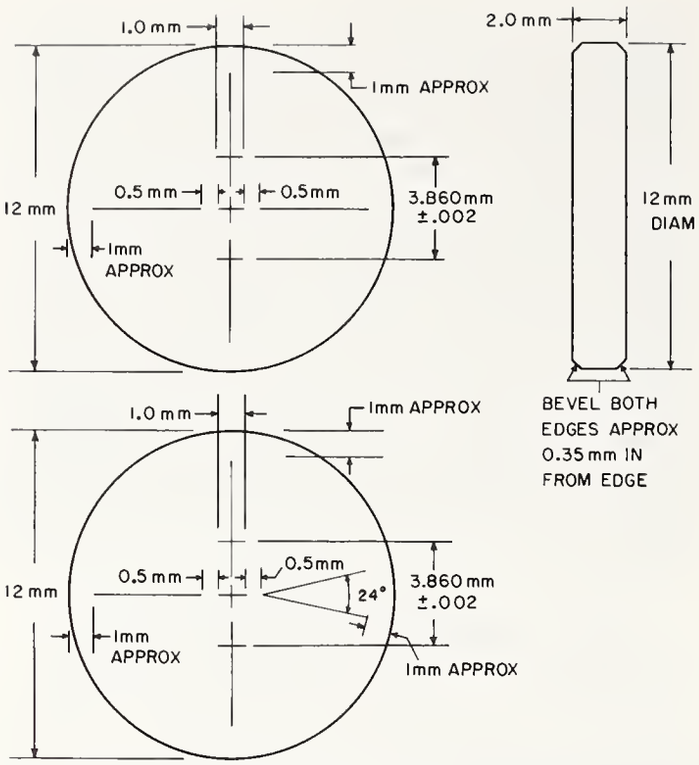


FIGURE 9. *Reticule-ruling and measuring machine.*

This apparatus, designed for laboratory use rather than for production of a large number of reticules, has an accurately lapped screw of 1-mm pitch. The graduated drum adjoining the hand wheel is fitted with a vernier and may be easily read to 1 μ. Equipped with microscope eyepieces and filar micrometers, as shown, it may also be used as a measuring instrument of high accuracy. The table attachment for holding reticules has longitudinal and crosswise adjustments, as well as a divided circle with vernier and tangent screws for angular positioning. The ruling point is supported on a light-weight frame, or yoke, with flexure plates instead of the usual pivots. A dash-pot controls the lowering of the ruling tool onto the work. The length of lines ruled with this instrument is controlled by a star wheel with adjustable stop screws.



ETCHED LINES 4 MICRONS WIDE, UNFILLED CENTRAL INTERSECTION WITHIN 12 MICRONS OF CENTER OF GLASS DISC.

FIGURE 10. *Two reticule patterns.*

The numbers for the graduations of a theodolite circle of 4-inch diameter are usually about 0.006-inch high. The numbers on the scribing template are 0.5 inch high and, by suitable adjustments, numbers in the range 0.004 to 0.007 inch can be produced. The numbers 4 and 8 were designed without "cross-overs," to avoid the trails so often shown at intersections of etched lines. The unique feature of the pantograph is that either coordinate of the motion of the tracing stylus in relation to that of the scribing needle can be reversed independently; by changing the position of two pins, the engravings can be made to be any one of the four possible mirror images of the scribing template.

4.2. Ruling Tools

One of the difficult problems in engraving fine lines and numbers in a resist is to obtain satisfactory tools to remove the resist cleanly and not wear out after a few rulings. Because the literature is particularly meager on the subject of engraving and scribing tools, the subject will be discussed in a general way, including a description of the method of making the scribing tools. It should be pointed out that much of this work was done under wartime pressure, and the most expeditious rather than the best way of making these steel tools had to be adopted.

a. Diamond Tools

The shaped diamond is, of course, the best tool because of its hardness and wear resistance, but it can easily be damaged by careless handling or improper use. Finding a source of supply is difficult, as most lapidaries are not prepared to furnish diamond tools with suitably sharp edges. The very small flats (3 to 6μ wide), necessary on the bearing surface to obtain a line of the desired width, are also an unusual requirement. The shapes of diamond tools, and their adjustment in the circular dividing engine, have been discussed by Page [1]. It is appropriate to mention here that the late Frederick Knoop, while a member of the Interferometry Section of the Bureau, worked out the techniques for the production of the diamond tools of exceptional quality used in this and other Bureau work.

It may be noted that the tools described by Page have a negative rake, which is necessary for ruling metal to prevent the tool from hogging (digging into the work), since the tool is not otherwise constrained in its motion to and from the work during the ruling. Although these tools gave satisfactory results in the ruling of resists on glass, it is believed that tools with a positive rake (leading face sloping backward) would produce cleaner engraving because the removed resist would be less likely to pile up on the face of the tool. To make such tools might create a problem for the lapidarist as, in order for the diamond to be lapped at the proper angle, it would have to be mounted so that a part of it would project beyond the side of the shank. It would be preferable for the face of the tool to be concave (that is, a section of a cone), with the axis inclined sufficiently to produce about a 10-degree positive rake. A tool of this shape would act somewhat like "chip-breakers" in cutting the edges of the engraving, and would also tend to funnel the carved wax out of the groove.

b. Steel Tools

Steel tools were used in engraving the resists on reticules, as well as in numbering the graduations of circles. They are much less expensive than diamond tools but wear out much faster. However, they are relatively simple to make and are easily resharpened. The engraving tools for reticule work were made from high-carbon steel (drill rod) 0.107 inch in diameter and 2 inches long. These were put on a surface grinder and 0.010 inch was ground off on one side, the full length of the tool, thus making a flat surface, which was used to clamp and position the tool. The tool blanks were heated to about 1,450°F (bright cherry red) and quenched in brine or mercury.

A relatively simple device was constructed to shape and sharpen the steel tools. This consisted of a small, shunt-wound, 120-volt d-c motor, mounted with the shaft vertical, as shown in figure 11. A steel hub with a sleeve was fastened to the motor shaft by a setscrew. A small abrasive wheel with a $\frac{3}{8}$-inch hole was clamped to the hub with a flange and nut as shown. The wheels used were ceramic-bonded alundum, about $\frac{1}{8}$ inch thick

by 4 inches in diameter; 120-grit wheels were used for roughing out, and 200-grit for finishing the rough grinding. At the right of the motor was a steel post about 1¼ inches in diameter and 5 inches high, having a ⅜-inch hole drilled through its longitudinal axis; the lower end of this hole was threaded for a cap screw to fasten the post to the base. A forked bracket on a ⅜-inch steel rod fitted into the post without any sidewise play. A clamping collar, with thumbscrew on one side and a 3/16-inch rod handle about 3 inches long on the other, served to position the height of the fork and permitted it to be turned with the handle as required. The tool holder consisted of a ⅜-inch steel rod 3 inches long, with a longitudinal hole about 1 inch deep in one end and a thumbscrew for clamping the tool. Near the other end were two steel pins, ⅛ inch in diameter by 1½ inches long, pushed halfway through the rod. One of the pins was parallel to the axis of the thumbscrew, and the other was perpendicular to it. A second tool holder, also shown in figure 11, differed from the one just described, in that it had pins on an adjustable collar, which permitted the adjustment of the grinding angle of the tool. These pins (in the form of a cross) supported the grinding fixture in the fork. The elevation of the fork controlled the included angle between the two sides of the tool.

After the tools had been hardened, they were ground on two sides (180 degrees apart). A low speed (about 600 rpm) of the wheel is desirable, but the wheel need not be wet, as at this low speed there is very little heating of the tool. During grinding, the tool is moved back and forth across the section of the wheel, which is rotating away from the tool. This prevents wearing a groove in the wheel and produces a smoother surface on the tool. The pressure of the tool on the wheel was of the order of 15 g (½ oz), and the angle of the tool above the plane of the grinding wheel was about 22 degrees. Grinding was continued, the tool and its holder being turned through 180 degrees occasionally, until a sharp edge was obtained.

For the final finishing of the edge of the tool, the abrasive wheel was replaced with a heat-treated, aluminum alloy disk, 5 inches in diameter by ⅛ inch thick, charged with 304 corundum. The aluminum lap was laid on a flat surface and sprinkled with No. 304 corundum powder before being mounted on the motor, then wetted with a small amount of water. A small piece of mild steel with a flat surface, such as a ¾-inch hex nut, was used for charging the lap. The whole surface was ground by rubbing the nut with a circular motion until the disk was a uniform gray. The charged lap was washed with soap and water, dried, and then mounted on the motor-shaft hub. The tool-holder fork was elevated about ¼ inch higher than in the rough grinding, so that the angle of the tool above the plane of the disk would be 28 to 30 degrees. The lapping was done with the same techniques as in grinding. The tool must be lowered gently on the lap to avoid damage to both

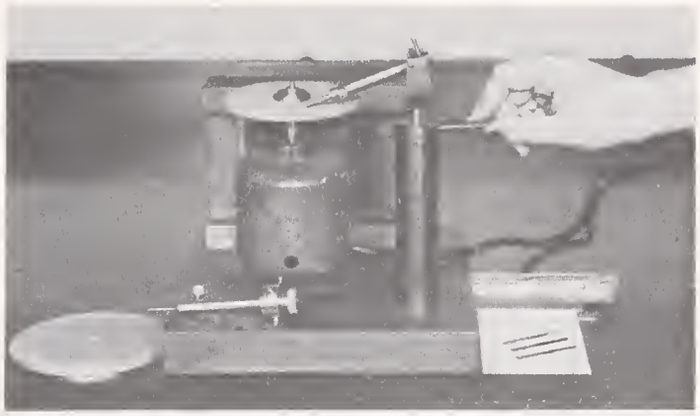


FIGURE 11. *Tool-grinding and lapping apparatus.*

Steel tools used for making reticules with the apparatus shown in figure 9 were shaped on this simple device. The disk on the bench to the left is an aluminum-alloy lap charged with very fine abrasive for finishing the edge of the tool.

the tool and the lap. A speed of about 300 rpm is recommended. The tool and holder were turned 180 degrees after 5 to 10 seconds of polishing. A 10-power jewelers loupe was used to follow the progress of the lapping. It is unnecessary to polish the tool more than ¼ mm back from the edge. When the edge appeared perfectly smooth, sharp, and bright, the tool-holder fork was lowered all the way, to obtain an angle of about 12 degrees between the tool and the plane of the lap. The tool and holder were then put in the fork so that the rounded side of the tool (90 degrees from previous polishing) was in contact with the lap The polishing procedure was continued until a face ½ to 1 mm in length was obtained.

Worn tools, if not damaged, may be reconditioned once or twice by relapping before regrinding is necessary. The lap should be recharged with the 304 abrasive when wear (bright areas) appears. This recharging of the lap by its grinding action helps to keep it free from ridges and grooves.

c. Scribing Tools for Numbering Graduations

For numbers or other irregular designs in wax resists, it is desirable that the tools have conical points, so that the width of lines will be the same, regardless of the direction of scribing. Because of the brittleness of the diamond, points made of this material must be in the shape of a blunt cone, with an included angle of 90 degrees. Perfectly conical diamond points are, however, exceedingly difficult to make. The tools must also be relatively thick, so as to be able to withstand stresses during lapping. Blunt, thick tools collect too much wax to be usable for the scribing of a resist, although they are satisfactory for scribing directly into metal surfaces. Further, the diamond crystal has certain planes that are so hard that it cannot be cut in these directions. Consequently, it is difficult to shape it to a symmetrical cone with a sharp apex.

Diamond chips or splinters have been used for microscopic writing directly on glass, but it is doubtful they would be satisfactory for scribing through a resist. Sometimes the effective point,

usually a corner, is so close to neighboring parts of the diamond fragment that much of it would rub on the resist adjacent to the scribed area. If this is the case, it would be likely to break off. McEwen [15] describes selecting a writing point and describes his pantograph. One of his writings of the Lord's Prayer (56 words) is in an area 0.0016 inch long by 0.0008 inch wide. Some of the problems of ruling microscopic scales and reticules are also discussed by Nichols [16].

As diamonds of a suitable size and shape are not usually available, a steel point is generally used. After some experience is acquired, a point with the required sharpness can be made from a steel phonograph needle in about half an hour. A satisfactory point should appear sharp when viewed at 100 diameters with a microscope (16-mm objective, 10-power eyepiece).

The steel needle was chucked in a collet of a small lathe. With the lathe running at about 200 to 400 rpm, a small, *natural* oilstone of fine texture is brought up under the needle with light but firm pressure and moved back and forth at a small angle to the needle so as to make a cone for about one-fourth the length of the needle and ending in a sharp point. Considerable care is necessary to avoid breaking the point during the final stages. The progress of this operation should be observed with a 10-power loupe.

The next operation is a slight reshaping and putting a high polish on the point. A microscope slide or similar piece of glass is wet with rouge and water for about three-fourths of its length on one side. The extreme end of the slide is grasped firmly between the thumb and forefinger and the glass brought up under the needle with a light pressure at an angle of about 30 degrees to the needle and moved, in a crosswise direction, back and forth a distance of about an inch until a sharp, highly polished point is obtained. The point is examined frequently with a 10-power loupe to determine when it is ready to be taken out of the lathe and examined with the microscope. To clean the point for this examination, push it carefully through a small piece of tablet paper or newsprint. If this examination shows that the point was broken off, start over again with the oilstone.

d. Positioning the Ruling and Scribing Tools in Ruling Machines

Preparatory to ruling or engraving, the ruling or scribing tool is mounted in the ruling machine, or pantograph, as nearly as possible in the correct orientation. Needle tools should be vertical. Ruling tools must be carefully adjusted so that the sharp V-edge is exactly parallel to the direction of ruling and also with 1 or 2 degrees of clearance. It is necessary to use at least a 10-power loupe or preferably a low-power ($\times 50$) microscope to properly make this adjustment for clearance. The tools should rest on a level glass surface, and the weight on the point should be at a minimum. The contact area of these tools is very small, and a load, or weight, of 1 g may exceed the elastic limit of the steel, and the point will be deformed unless precautions are taken to prevent it. Table 2 gives the pressure on the tool point produced by a 1-g load on specified contact areas.

After the tool has been properly positioned, a little water and rouge are put on the glass around the tool, and a small flat is lapped on the tool by moving the glass or, preferably, by moving the tool with its mechanism back and forth through the rouge a few strokes until the desired width of contact is obtained. This width is determined by ruling lines in the wax resist and measuring the line width. Lapping is also necessary on needles and the same procedure is used.

TABLE 2. *Contact pressure on tool points*

Contact area of tool	Contact pressure of 1-g load on tool point per—	
	Square centimeter	Square inch
μ^2	*kg*	*lb*
½	200,000	2,844,680
1	100,000	1,422,340
2	50,000	711,170
4	25,000	355,585
8	12,500	177,792
16	6,250	88,896
32	3,125	44,448
64	1,567	22,224

It is difficult to state specifically the load to be used on a tool for engraving the wax because the actual contact area is usually unknown. Further, tools from the same batch are not always uniformly hard, and the elastic limit will vary correspondingly. The stress in the part of these tools in contact with the glass is always higher than that considered good engineering practice. In fact it is remarkable that they function as well as they do. The conditions of use may be considered as a "special case," and no engineering data appear to be available on the behavior of metals of such small dimensions under such heavy loads. Of course, the support, or backing, of the contact area by much thicker sections contributes materially to their strength. The fact that the glass is not fractured is, no doubt, also due to the support or backing of the glass in the immediate vicinity, elastically absorbing the strain. If one could observe the contact area of the tool and glass with sufficiently high magnification, the expected elastic movements of both glass and steel during the ruling would make an interesting study.

High-carbon steel, which could be hardened to around 60 on the Rockwell C scale, was tried and found to work satisfactorily. Consequently, no systematic study was made to find the relative merits of the various steels for ruling tools. Tungsten carbide might be better than any of the steels because of its hardness and high compressive strength.

With steel tools, it was possible to rule lines in wax of 1 to 2μ in width and to etch these to a width of 4μ. The steel needles used to scribe

numbers were positioned in the numbering machine, and a small, flat surface, about 3μ in width, was lapped on the point, as previously described. Sometimes as many as six circles were numbered (a total of 2,160 graduation lines) with one needle before the point was worn too much for further use.

The useful life of a ruling tool, exclusive of accidents and wear, is also dependent upon the impact on the tool point by the ruling mechanism in setting the tool down on the work. The use of a rugged, heavy, rigid frame on pivots and a counterweight to reduce the load on the tool point to the desired amount is common practice. Many of these heavy fixtures were designed primarily for scribing, or engraving, relatively heavy lines in metal where the force required is much larger than is necessary for ruling in wax on glass. When a fixture of this type is used, the mass of the counterweight, as well as the pivoted frame and all other parts, which rotate about the pivots as a unit, contribute to the energy that is absorbed by the tool point at the instant of contact with the glass. Of course, the tool is lowered by a cam and levers so that the velocity is low. However, there are no springs between the tool and the arm and counterbalance framework to lessen the impact. It is fortunate that the wax resist helps to "soften the blow". However, for these fine lines, the mechanism should be designed to reduce the impact to a minimum.

5. Etching the Graduations

5.1. Examination of Engraving

Before etching the glass, the engraved lines and figures in the resist should be examined with a microscope. Note particularly the ends of lines for "pick-outs" caused by wax adhering to the tool; crossover points in numbers for wax trails; irregularities at edges of lines, small conchoidal fractures indicating brittleness; wax piled up on one side of line, indicating the tool was not set square with the direction of ruling, the heel of the tool was dragging on one side; small pieces of wax "chips" in lines or lying across lines; scratches or "digs" in resist; oil drops on resist, and other defects. If the defects are serious, the glass should not be etched because of its cost. A light repolishing is necessary before it can be recoated with resist because an invisible amount of glass is sometimes removed by the ruling tool. After the glass is etched, it should be reexamined, making use of the notes made of defects appearing in the ruled resist, because this is the only way that the importance of the various defects can be ascertained.

5.2. Methods for Etching

Glass etching is done with hydrofluoric acid because it attacks glass quite rapidly either in solution or in the gaseous state. It is also potent on human beings and will destroy living tissue [17]. If it is handled with the fingers, it will cause hard-to-heal sores and will work its way under the fingernails, where it is hard to get at to neutralize. If some should get on the skin, wash it off, and bathe immediately in lime water or borax solution or other deactivator. Keep a container of antidote conveniently at hand at all times. Protective gloves should be worn while working with it, and further, all operations should be carried out under a chemical hood with forced-draft ventilation to the outside. The gas, or vapors, are extremely irritating to the lungs. The glass to be etched should be held in tweezers, tongs, or special fixtures during the etching and washing operations. The apparatus and physical operations should be planned to reduce the hazard to a minimum.

For the class of work discussed in this paper, the glass may be etched by either of two methods: (1) hydrogen fluoride (gas), or (2) hydrofluoric acid (liquid). The gas form of etching can be done (a) with the presence of water vapor, or (b) with the dry gas without water vapor. With method 1–a the liquid acid is contained in a lead, wax, or other suitable container, and the glass being etched is held over the acid for a suitable time, usually from 10 to 60 seconds or more. In method 1–b, acid is mixed with sulfuric acid in a suitable container, which is kept closed at all times. In a closed container, the sulfuric acid takes up practically all water vapor in the space above the acid solution. The glass to be etched is subjected to the hydrofluoric acid gas (hydrogen fluoride) preferably by methods which will not require complete uncovering of the pot. Etching times are about the same as for method 1–a. These methods give a matte etch and are used principally for fine-line etching not intended for pigment filling. Vapor, or gas, etching has some advantages in that the etching times are longer than for most etching solutions, thus giving better control of the amount of etching. Chips of resist from rulings lying across lines (bridging) do not interfere with vapor or gas etching, and the method is clean and easy to operate.

Liquid etching solutions are more often used than the gas or vapor. These may be classified roughly into two groups: (a) solutions giving a clear etch, and (b) those giving a frosted etch. Clear etching is obtained with straight hydrofluoric acid (2–a). Dilute solutions up to a maximum strength of 48 to 60 percent are commonly used. Solutions giving a frosted (matte) etch (2–b) contain salts that promote the formation of insoluble fluorides of calcium, barium, or lead, depending upon the composition of the glass. The insoluble products formed during the etching

interfere with the etching, thus causing a rough, or matte, surface. With a given kind of glass, the degree of matting and the size of grain are controlled by the composition of the etching solution. A résumé of glass etching given by Schweig [18] should be consulted for further details.

Another method, known as "dry-etching" or "powder-etching", used principally for etching trade-marks and sometimes for etching graduations in volumetric glassware, was not studied because of a lack of sharpness at the edge of the etching. The active agent is a fluoride salt, such as ammonium bifluoride, in the form of a fine powder. Details of this process are given by Jones and Hammond [9] and elsewhere [19].

5.3. Vapor Etching of Masters and Reticules

After some study of etching methods, it was decided to etch the masters with the dry gas (hydrogen fluoride). This method was chosen because, with the large size of masters (6 inches in diameter), it was believed that a more uniform etching could be obtained than with the hydrofluoric acid. Further, the etching time required with the dry gas was about 60 seconds, compared with ½ to 1 second with 48-percent hydrofluoric acid, thus giving better control.

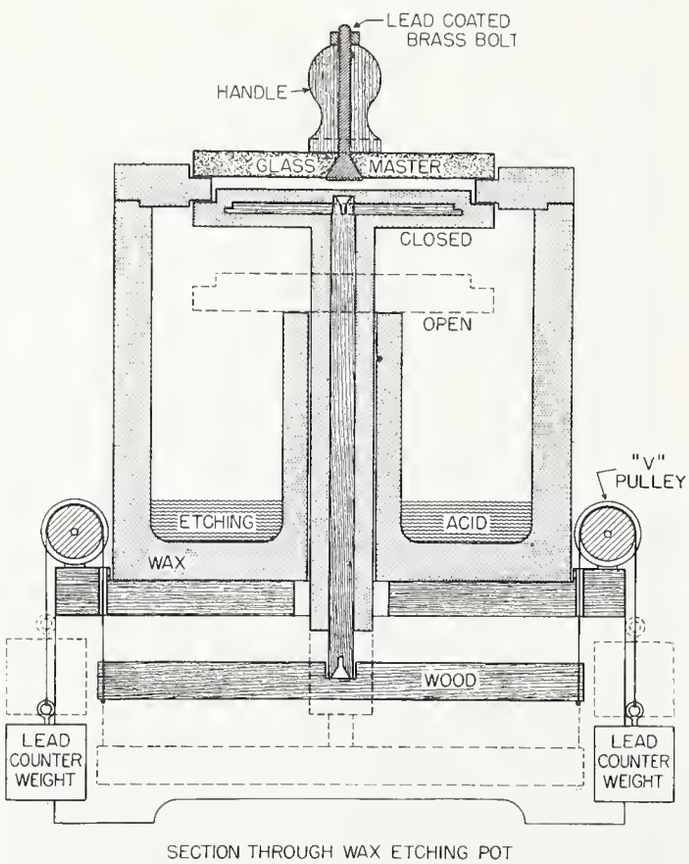


FIGURE 12. *Wax pot for vapor or gas etching of masters.*

The glass master, shown in position for etching at the top of the pot, is exposed to the etching vapors by lifting the two counterweights simultaneously. This causes the center unit to be lowered to the position shown by the dashed lines. Lowering the counterweights closes the pot.

Vapor etching of masters was done in a wax pot designed specifically for this purpose. Figure 12 shows a section through the pot, and its accessory fittings. The pot is cylindrical, 9⅛ inches in diameter by 8½ inches high. It was cast of ceresin, and the component parts were machined in a lathe [7] so that they fitted together accurately. The central piece has a round, wood stem with a plywood disk at one end for reinforcing the wax. The wood stem had centers drilled at the ends for turning this part to size on a lathe. This central unit rests on a small wooden beam; the beam is supported at the ends by cords that run over pulley wheels to lead counterweights. The counterweights are slightly heavier than the combined weights of the beam and the central or closure unit. The etching solution consisted of 500 ml of concentrated sulfuric acid to which was added (in the pot) from 10 to 20 g of calcium fluoride, depending upon the activity desired. This etching mixture should stand overnight to come to equilibrium. If desired, straight 48-percent hydrofluoric acid may be used. This will etch faster than the sulfuric acid mixture, and the etching will be nearly clear (only slightly matted).

In use, a wooden handle was attached to the master to be etched, and the master was placed face down on the pot in the recess provided for it, as shown in figure 12. The etching vapor had access to the master when the post closure, which opened inward, was lowered by lifting the two counterweights simultaneously. When the closure was lowered, a draft of vapor from the etching acid swept across the master from the edge toward the center. When the etching time had expired, the counterweights were lowered, thus closing the pot, and the master was removed and washed in a stream of water to stop the etching.

Excellent line quality was obtained by this method. The lines were of sufficient depth to be filled with lead sulfide but not deep enough to be filled satisfactorily with pigment by hand application. As previously mentioned (section 3.5), the resist used, principally Halowax and gilsonite, was somewhat permeable to hydrogen fluoride. The surface of the glass was faintly etched. However, the polishing required to remove the surface layer of lead sulfide also removed this surface etching. It may be of interest to mention that this pot, made 12 years ago, is much discolored but is still in good working order.

The same general etching procedure was used successfully in making about 500 precision reticules having line patterns, as shown in figure 10.

[7] The wax pot and the ring part of the cover were fastened to a wood disk attached to the face plate of the lathe. Holes were drilled in the bottom of the pot and also in the wax ring. Wood screws with iron washers were used for fastenings. After turning, the holes in the wax pot and cover ring were closed by filling with melted ceresin.

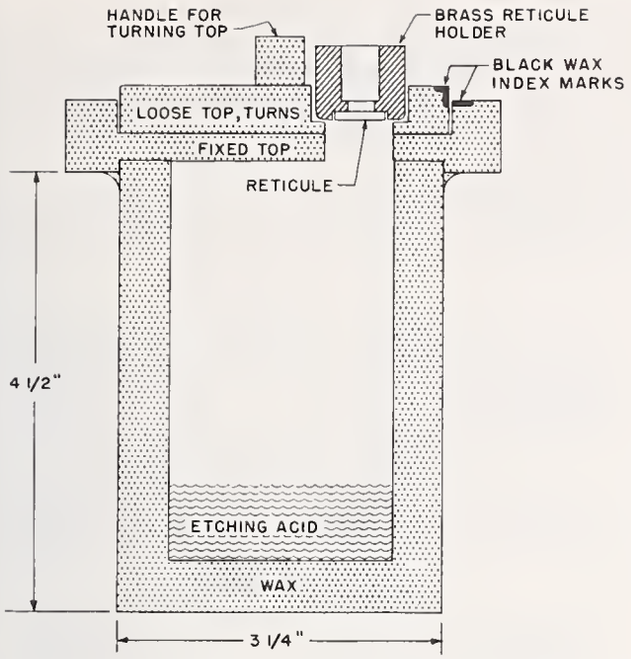


FIGURE 13. *Wax pot for vapor or gas etching of reticules.*

This sectional view of the pot shows a reticule in a brass holder in position for etching. Mating holes in the fixed top of the pot and in the rotatable top when in coincidence exposes the reticule to the etching vapors. Turning the movable top closes the pot.

The resist, mainly beeswax and gilsonite, was coated on the reticules with the whirler rotating at 1,800 rpm. The rulings were 1.5 μ in width and, when etched, were from 3 to not exceeding 4 microns in width. For the etching of reticules, a ceresin wax pot, approximately 4½ inches wide by 6¼ inches high, was constructed, as shown in figure 13. The main body of the pot, 3⅜ by 4½ inches, contains the etching acid. The top of the pot was permanently fixed on by sealing with hot ceresin. This top was recessed to provide a place for a disk, also of ceresin, which could be turned (rotated) in this recess. A hole, drilled off center through both the fixed cover and the disk, served as a gate valve and, when opened, provided access for the hydrogen fluoride to reach the reticule. The reticule, in a brass holder, rested on a shoulder in an enlarged part of the hole through the movable disk. Turning this disk about 90 degrees closed the pot.

The large etching pot (fig. 12) also can be used for etching a number of reticules at the same time by making an adapter consisting of a disk 6 inches in diameter of either ceresin or, preferably, of methyl methacrylate. This disk is drilled with holes of suitable size to support the reticules in their holders. This disk would be placed on the pot in the same position as the master.

These etching pots provide a clean, convenient, and safe method for vapor etching of glass. They could be made of lead or Monel metal. However, wax (ceresin) has some advantages in that the pots are relatively inexpensive, and their surfaces are always dry because the acid does not "crawl" on the wax. This is not the case with metal pots, particularly after they become old and the surface has become oxidized, corroded, or dirty.

5.4. Liquid Etching

Etching with aqueous hydrofluoric acid solutions is done by dipping the glass directly into the etching acid or by pouring the acid onto the engraved wax side of the glass. Of course the glass must be completely covered with resist, or otherwise protected, so that only the engraved areas of the glass are exposed to the action of the etching acid.

The resist does not limit the width of the etched line to that of the engraved line. In general, the width of the etched line exceeds the width of the engraved line by an amount equal to twice the depth of the etching. This approximate relation between the width of the engraved and etched line and its depth does not hold if the resist does not remain tightly bonded to the glass during the etching. A little reflection will show that an etched line that is narrow with respect to its depth can be filled more easily with a pigment than one that is wide with respect to its depth. As the width of an etched line cannot be less than twice its depth plus the width of the engraved line in the resist, it is obvious that the engraved line should be as narrow as other factors permit.

Experience indicates that the resist and etching techniques should be tailored to the particular job. A resist that works well on one kind of glass with a given etching solution may fail completely on a different kind of glass. Of course, one resist may be more versatile than another, but the requirements for etched glass scales and reticules are so varied that one combination (resist, glass, and etch) will not suffice.

As mentioned in section 3.1, materials used to compound resist coatings are, in general, nearly but not completely inert to either water or hydrofluoric acid or hydrogen fluoride. Some, like ceresin, paraffin, and beeswax, are better than others, such as coumar resin, rosin, and Halowax. Sometimes the water or acid is absorbed, and in other cases chemical reaction takes place or the resist is permeable. When these materials are coated on the glass in a thin layer, sometimes only a few hundred thousandths of an inch thick, it is to be expected that they will be troublesome at times. If the coating is made too thick, the etching acid may not wet the resist, and consequently, surface tension may prevent it from getting through a finely engraved groove to the glass, resulting in an etched line of irregular width and depth. Successful resists are usually a compromise of several factors.

Resists that are not inert to hydrofluoric acid or hydrogen fluoride (gas) are not suitable for deep etching. In some instances deep etchings can be obtained by repeated etching. For example, in one instance with a light-sensitive resist, the depth of etching was considerably increased by etching lightly, then washing off the etching acid, drying, and then reetching. This cycle could be repeated several times. The same result was obtained with a Halowax-gilsonite resist when etched with a liquid etching solution containing hydrofluoric and phosphoric acids (formula EP–1). In another instance, with the same Halowax-gilsonite resist and etching with hydrogen fluoride (gas), this procedure did not work; washing in water removed the resist.

Dilute etching solutions and some containing salts, such as those described by Schweig [18] for etching glassware, etch the glass more slowly than the concentrated 48-percent hydrofluoric acid. However, they are not always useful in the etching of fine lines because of a tendency of the etching acid to get between the resist and the glass and thereby produce an unusually wide etched line (abnormal undercutting). The same effect, only more pronounced, was observed with 48-percent hydrofluoric acid diluted with an equal volume of water. Concentrated 48-percent hydrofluoric acid etched rapidly and usually produced sharp-edged lines with no signs of abnormal undercutting. The principal difficulty with this strong acid was that the etching time for fine lines (10 μ or less) was too short to control accurately.

In section 3.2 mention is made of the use of Amino Silane for waterproofing the glass as a means of preventing abnormal (excessive) undercutting during the etching. Some resist formulations, when coated on glass, show abnormal undercutting during the etching; also some kinds of glass, particularly some of those having a fire polish, exhibit abnormal undercutting even with a resist that is satisfactory on other kinds of glass. Further, some etching solutions have a strong tendency for abnormal undercutting. The least satisfactory glass found is a particular lot of lantern-slide cover glass, and the etching solution with the greatest propensity for abnormal undercutting is etching formula ET–10 given below. This combination was used for testing resists and the effectiveness of waterproofing the glass with Amino Silane.

To illustrate the effectiveness of the Amino Silane treatment on the bonding of the resists to the glass and the effect of different etching solutions, the following experiment is described. Two lots of glass were chosen, (1) lanter-slide covers, and (2) plate glass known to be fairly good. After cleaning with chromic acid solution, one-half of each of the two lots of glass was treated with Amino Silane for 90 seconds, as described in section 3.2. All of the glasses were then coated with resist (resist formula W–2) on a whirler at 450 rpm. Test patterns were ruled in the resist on all of the glasses with the reticule-ruling machine. The pattern consisted of two sets of parallel lines about 1 cm long, one set ruled on top of the other at an approximate right angle, thus producing a small grid.

Six etching solutions were used in this test. The compositions of these and their identification is as follows:

Etching Solution EP–1

Phosphoric acid (85%)	100 ml.
Water	50 ml.
Hydrofluoric acid (48%)	25 ml.

Etching Solution EP–2

Phosphoric acid (85%)	150 ml.
Hydrofluoric acid (48%)	25 ml.

Etching Solution EP–3

Phosphoric acid (water removed by distillation)	100 ml.
Hydrofluoric acid (48%)	20 ml.

Etching Solution ET–10

Hydrofluoric acid (48%)	100 ml.
Water	100 ml.

Etching Solution EC–5

Hydrofluoric acid (48%)	qs

Etching Solution ES–7

Water	65 ml.
Ammonium bifluoride	100 g.
Sodium fluoride	11. 5 g.
Hydrofluoric acid (48%)	50 ml.

The first three formulas, EP–1, EP–2, and EP–3, are solutions of phosphoric and hydrofluoric acids, which differ principally in their water content. The solution containing the most water gives the least matte in the etching. These etching solutions were developed primarily for use with a photographic resist described in section 7. Their use with wax resists is recommended. Formula EC–5 is a 48-percent hydrofluoric acid solution. It etches rapidly and in most cases deeply etched lines of good quality can be obtained with it. Formula ES–7 is an etching solution containing salts to promote the formation of insoluble compounds during etching, such as the fluosilicates, which form particles that lie on the glass and interfere with the etching, thereby promoting an irregular etch or matte surface.

The results of these experiments are summarized in tables 3 and 4 and illustrated in figures 14 and 15. It may be noted in table 3 that the lantern-slide cover glass, untreated, gave good quality of lines without any signs of undercutting with the first three etching solutions and that the other three solutions produced unsatisfactory etches, all of which were severly undercut. The same glass treated with Amino Silane showed no under-

TABLE 3. *Effect of waterproofing on the quality of etching of lantern-slide glass*

Sample number in figure 14	Etching formula	Etching time	Quality of lines		Type of etch			Depth of etch			Edge of lines	
			Good	Poor	Clear	Matte	Coarse matte	Light	Medium	Heavy	Sharp	Undercut
Untreated												
		sec										
1	EP-1	20	×		×				×		×	
2	EP-2	30	×			×		×			×	
3	EP-3	30	×		×			×			×	
4	ET-10	10		×	×					×		×
5	EC-5	5	×		×					×		×
6	ES-7	20		×			×		×			×
Waterproofed with Amino Silane												
7	EP-1	20	×		×				×		×	
8	EP-2	30	×			×		×			×	
9	EP-3	30	×		×			×			×	
10	ET-10	10	×		×					×	×	
11	EC-5	5	×		×					×	×	
12	ES-7	20	×				×		×		×	

TABLE 4. *Effect of waterproofing on the quality of etching of plate glass*

Sample number in figure 15	Etching formula	Etching time	Quality of lines		Type of etch			Depth of etch			Edge of lines	
			Good	Poor	Clear	Matte	Coarse matte	Light	Medium	Heavy	Sharp	Undercut
Untreated												
		sec										
13	EP-1	20	×		×				×		×	
14	EP-2	30	×			×		×			×	
15	EP-3	30	×			×		×			×	
16	ET-10	10		×	×					×		×
17	EC-5	5	×		×					×	×	
18	ES-7	30		×			×			×		×
Waterproofed with Amino Silane												
19	EP-1	20	×		×				×		×	
20	EP-2	30	×			×		×			×	
21	EP-3	30	×		×			×			×	
22	ET-10	10	×		×				×		×	
23	EC-5	5	×		×					×	×	
24	ES-7	30	× [a]				×			×	×	

[a] Etching solution is the type used for frosting glass. Consequently rough edges are expected.

cutting and the etchings were all satisfactory. In table 4, the results obtained with untreated glass show that all of the etching solutions except ET–10 gave good results. The same glass treated with Amino Silane again gave good results with all of the etching solutions. Not shown in these tables is the effect of the Amino Silane on the quality of the etching. In all cases the quality of the etch is definitely improved, even for those solutions that gave good line quality without its use. Also it should be mentioned that the amount of etching of the treated glasses is slightly less (possibly 10 percent) than of the untreated glass. That the improvement noted is not due to slightly less etching has been verified by other tests. Other experiments have shown that the waterproofing treatment also gives a similar improvement in vapor etching.

There are several products on the market for waterproofing by fuming or bathing with silane and silicone compounds. These waterproofing compounds were not tried, but it is believed that they would be effective. Some of these would require the use of a special cabinet.

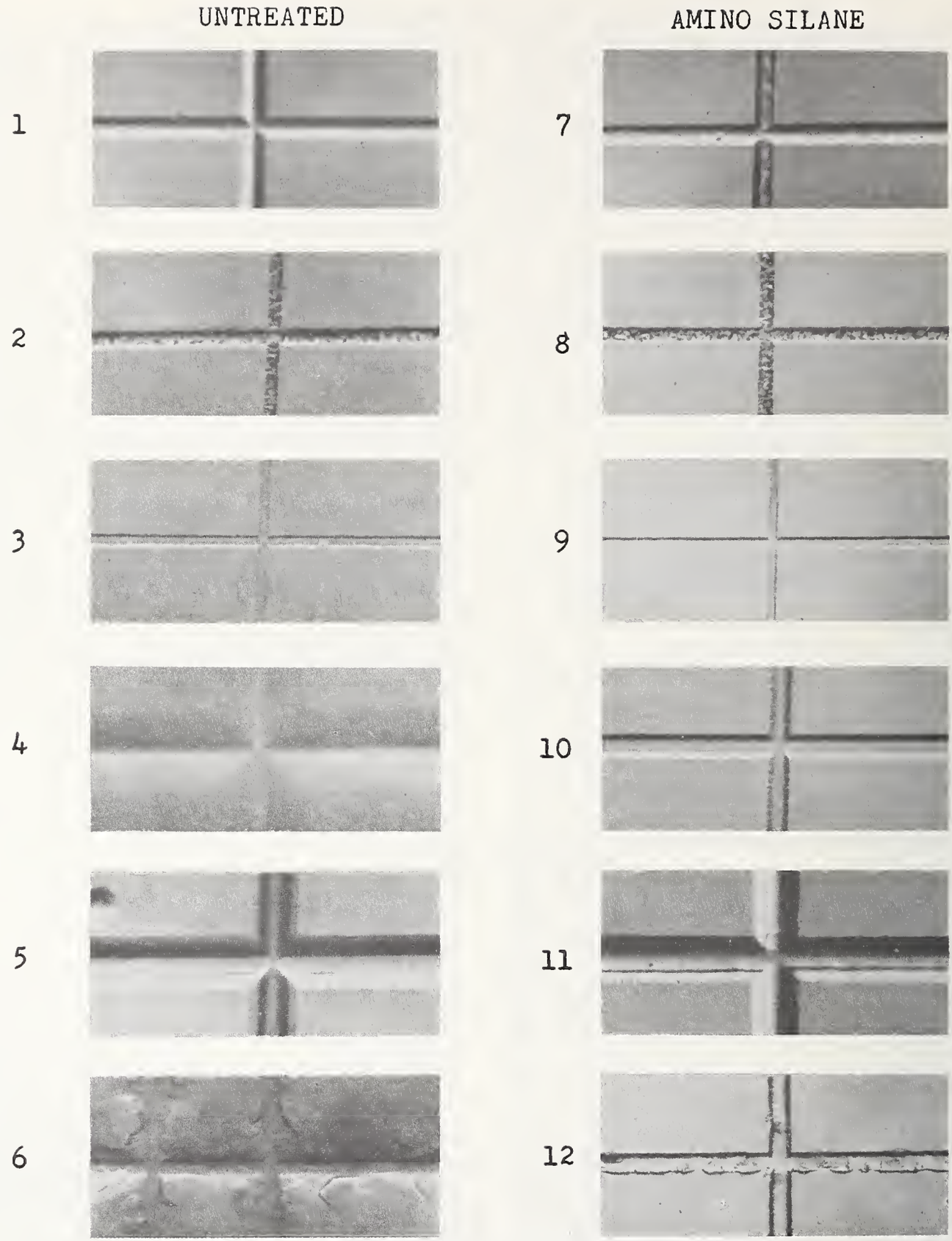


FIGURE 14. *Photomicrographs of etched glasses listed in table 3.*

The magnification is such that 5 μ equals 1 mm in the illustrations.

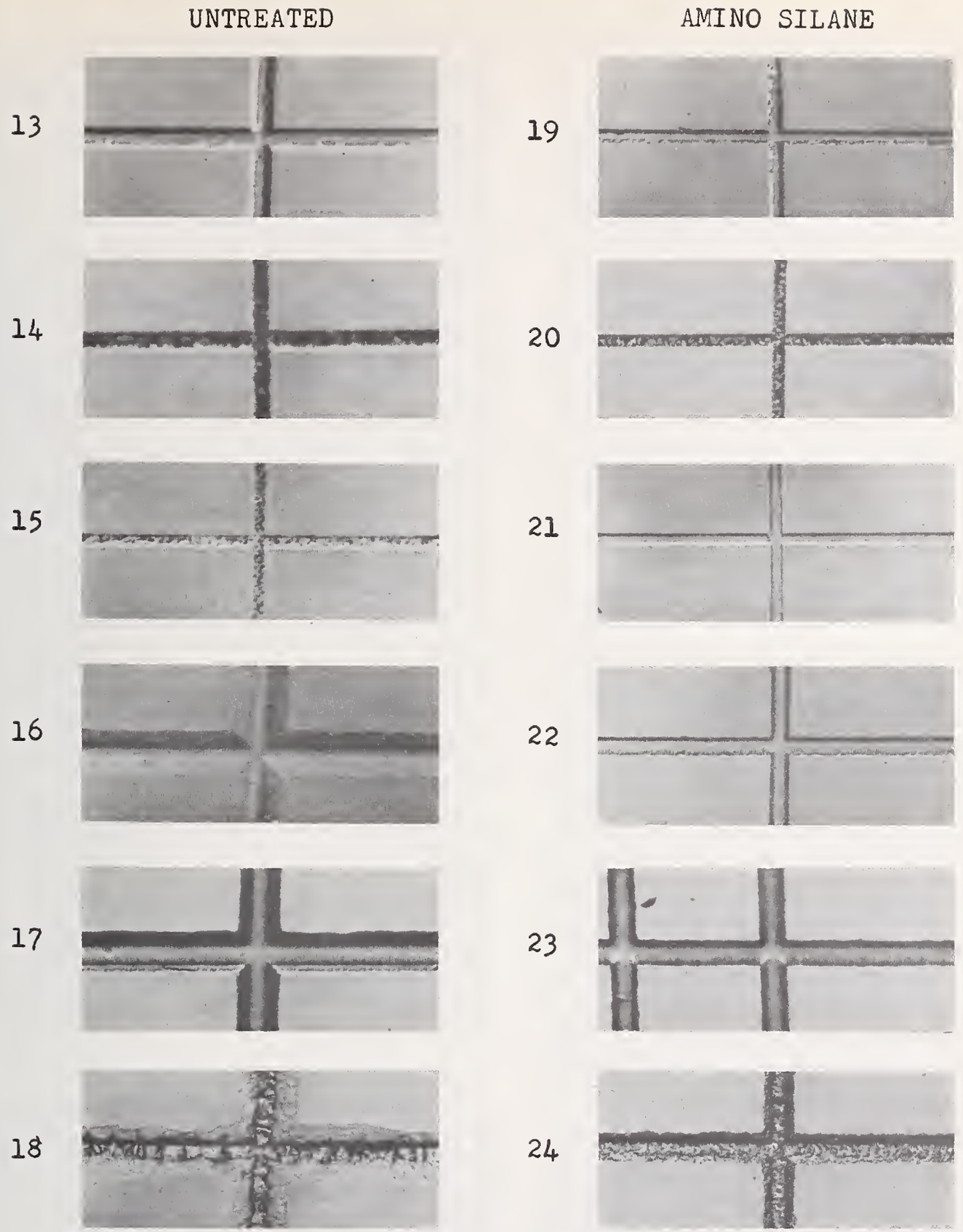


FIGURE 15. *Photomicrographs of etched glasses listed in table 4.*
The magnification is such that 5 μ equals 1 mm in the illustrations.

518-743 O - 73 - 54

6. Filling of Etched Graduations

6.1. Filling With Lead Sulfide

Shallow etching, which produced narrow lines, was used on many masters and circles, and as the etched lines were not of sufficient depth to be filled with pigment, lead sulfide filling was used. The lead sulfide was deposited chemically on the surface of the glass in much the same way as in mirror silvering and then it was polished off, thus leaving it only in the etched areas.

The formulas found in the literature for depositing lead sulfide on glass proved difficult to use as a precipitate always formed on mixing the solutions and heating was usually necessary. The following formula was developed which gave satisfactory results:

Solution A:

Lead nitrate	10 g.
Distilled water to make	1 liter.

Solution B:

Thiourea	20 g.
Distilled water to make	1 liter.

Solution C:

Potassium hydroxide, 85% assay	23.5 g.
Distilled water to make	1 liter.

The sulfiding solution is prepared by taking equal volumes of each solution. Combine solutions A and B and add solution C to the mixture. When first mixed the solution should be clear. If a slight whitish precipitate persists, add just enough of solution C to make it disappear. Best results are obtained with stock solutions made up just before use, but they are usable for about 3 days. The time required for the coating to start depositing on the glass gradually increases with the age of the solutions. Experiments showed that the presence of sodium silicate inhibits the reaction and and prevents the deposition of the lead sulfide. Thus, the increase in time required for old solutions to work is caused by the dissolving of silicate from the glass bottles in which solutions are stored. As the concentration of the silicate increases, the time for deposition also increases and eventually heat must be applied to start and continue the reaction. The use of polyethylene bottles is suggested for storing these solutions.

The glass to be coated should be cleaned with chromic acid solution (see section 3.2) and not allowed to dry between cleaning and coating with lead sulfide. Pour the sulfiding solution into a dish containing the glass to be coated. When the stock solutions are new, the sulfiding solution is clear at first and in from 1 to 5 minutes it begins to deposit lead sulfide on the glass at room temperatures of from 25° to 30° C. With fresh solutions, the precipitation is continued for about 10 minutes. Three coats are applied, rinsing the glass and dish with distilled water between coats. During the deposition period, the glass surface should be swabbed gently with a tuft of absorbent cotton to prevent particles that form in the solution from adhering to the surface. When the coating is complete, the glass is washed in distilled water, dried on a whirler, and then placed in an oven at 50° C (120° F) and left until it is approximately at oven temperature. After this, it is removed and the graduated area is painted with some of the beeswax-gilsonite resist (formula W–2) to protect it during the following operation. The glass is next immersed in a 20-percent solution of ferric chloride to remove the lead sulfide except where protected by the wax resist. This is done to simplify the polishing off of the sulfide from the surface of the glass with a lap. The glass was dried on a whirler and the wax masking removed by warm carbon tetrachloride (60° C) under a well-ventilated hood. The lead sulfide was polished off the surface of the glass, leaving a coating of lead sulfide in the lines and numbers.

Polishing was done on a beeswax lap charged with USP magnesium oxide. The customary pitch lap could not be used as it seemed to seize to the lead sulfide. The lap consisted of a 6-inch disk of cast iron about ½ inch thick, machined flat on one side. The other side had a boss in the center of the disk. The boss had a threaded hole to take a steel rod that was used to clamp the disk to a bench or vise. The flat side of the lap was covered with a sheet of honeycomb foundation (beeswax) by coating the metal surface with some wax masking solution and pressing the wax foundation into contact, using a piece of plate glass, and left until dry.

The lap was fastened to the bench, face up, and charged by sprinkling it with a small amount of powdered magnesium oxide, followed by a little distilled water. The lap was conditioned, or broken in, by taking a glass similar to a master and moving it in the usual manner until the lap was charged and the top edges of the honeycomb cells were brought into a plane. When this was completed, the master or circle was put on, additional water and magnesium oxide applied, if needed, and the lead sulfide polished off the surface.

6.2. Filling With Pigments

It is customary to fill deeply etched lines with a pigment. Sometimes a white pigment, such as titanium dioxide or zinc oxide, is used. These pigments are sufficiently opaque to use with transmitted light but will appear white when illuminated for night use. More often the pigment used is lampblack or rouge.

To fill the lines, the glass is clamped by the edges, face up, on a rigid support. A dilute sodium silicate solution is prepared by diluting 1 part of water glass (sodium silicate solution, 40 to 42° Baumé) with 7 parts of water. A few drops of this dilute solution is placed on a piece of glass, and a little pile of the dry pigment is placed nearby. A lint-free cotton or silk cloth is

wrapped around the index finger. With the end of this finger, first touch the dilute sodium silicate solution and then the dry pigment. The silicate and pigment adhering to the finger are rubbed on the etched glass with a back and forth and sidewise motion to work the pigment into the etched lines. The excess is wiped off with a rotary motion of the side of the cloth-covered finger as it is moved across the glass surface, thus a clean surface is continually being used in the wiping. Considerable experience is required to fill even a small reticule. The chief difficulty is not in getting the lines filled, but in getting a clean glass surface free from smear. If the sodium silicate drys on the surface of the glass, it is very difficult to remove without scratching the glass.

7. Replicates of Precise Scales by Photoetching Techniques

7.1. Photographic Methods

As mentioned in section 1, a part of this work was the development of a photographic method for the production of etched glass replicates from an original ruled, etched, and filled master circle. Photographic methods employing lenses could not be considered because image distortion resulting from lens aberrations [20, 21, 22] would exceed the small tolerance, ± 0.00001 inch (0.25μ), allowed for the position of graduations.

Contact printing with a light-sensitive coating on the glass circle blanks from an optically flat master offers the best promise of freedom from distortion. This light-sensitive coating, after exposure and development, must result in a mask or stencil similar to that of an engraved wax resist and likewise be impermeable and resistant to hydrofluoric acid so that the image of the graduations of the master can be etched into the glass. Unfortunately, as judged by the literature, no light-sensitive resist meeting these requirements was known.

A number of photographic processes have been used for producing reticules and scales of high quality. Rheinberg [23, 24], Geiser [25, 27], and Leistner [26] have reviewed and described a number of these processes.

It may be helpful to point out briefly the principal kinds of light-sensitive coatings used in making glass reticules and scales and follow this with a simple classification of types of reticule processes, including some references to the literature.

The light-sensitive coatings used in reticule work are (1) silver halide emulsions in gelatin, albumen, or collodion having very fine grain, high resolution, and high contrast; (2) a dichromate sensitized photoengravers glue, albumen, shellac (cold top), etc., and (3) iodoform sensitized resins, asphalt, etc. Photographic emulsions (1) are sometimes used for making master negatives or positives, but in recent years they are being used less for making reticules because the clear parts of reticules (where there is no image) are covered with a layer of gelatin, collodion, or other vehicle for the silver halide. The surface between parts of the image is not always perfectly clear and is difficult to clean. The surface is not an optical plane and, therefore, in an optical system it scatters some light, thus, degrading image contrast. Cementing a cover glass on emulsion side will reduce the scattered light and also protect the image. Coatings (2) and (3) do not have this characteristic. These dichromate or iodoform sensitized coatings become insoluble on exposure to light. Development of the image consists of dissolving out the soluble (unexposed) areas with warm water, alcohol, kerosine, turpentine, or other suitable solvent, depending upon the nature of the colloid. Exposed through a negative of a line pattern, the lines are reproduced on the reticule as an insoluble colloid on clear glass. If exposed through a positive, the lines are reproduced as clear glass, thus the resulting image resembles a stencil, or mask. The final image is deposited on the glass through this stencil or the stencil is used as a resist to etch an undercoat or, if it is a resist for hydrofluoric acid, the glass is etched. Following this the stencil layer is removed with suitable solvents.

Dichromate sensitized glue or gelatin containing silver halide or colloidal silver is often used to obtain an opaque image on clear glass. After exposure in contact with a negative, the soluble unexposed areas are removed by developing in warm water, leaving the image as isolated areas, or "islands", on the glass. The silver halide in the image area is blackened with a photographic developer, then intensified, if necessary, with a mercury or silver intensifier as in regular photography. In the case of colloidal silver, physical development and intensification are necessary to make the image black and opaque. Glue images not containing silver or silver halide may also be blackened by alternate treatments with silver salts and reducing agents.

The lack of a light-sensitive coating of the stencil layer type that can withstand the action of hydrofluoric acid necessary to etch the reticule pattern into glass has lead to multiple layer coatings. For example, the glass may be coated with a layer that is not light sensitive but which is inert to hydrofluoric acid. Metallic silver, Bakelite, and asphalt have been used as such an undercoat. The stencil layer may be dichromate sensitized glue or shellac. After developing the stencil layer, the uncovered subcoat is then dissolved away, leaving the glass exposed. When the coating is dry the back and other parts of the glass that are not to be etched are protected by painting them with asphalt or wax. The glass is then etched by exposing it to hydrogen fluoride vapor or dipping it in the etching acid.

For convenience and clarity, the processes available for making reticules and scales may be classified into three main groups as follows: Type I, reticule image on the surface of the glass; type II, image fused in or on the glass, and type III, image etched in the glass. Each type is further classified by the nature or kind of image (continuous, discontinuous, or stencil, etc.) and finally by a process number. Thus each kind of reticule and the process used for its manufacture is approximately identified by three characters, thus type I–A–1 in the table below would represent a reticule made with a silver halide emulsion in which the vehicle layer (gelatin collodion, etc.) covers the face of the reticule. A specific process could be identified by adding another digit, but this is unnecessary for the present purpose. At the beginning of each process description the letter (N) or (P) indicates that a negative or positive is used in exposing the sensitized coating.

TYPE I. IMAGE ON SURFACE OF GLASS

A. Single Continuous Layer Processes

1. (N) Silver halide emulsion in gelatin albumen or collodion. Very fine grain and high contrast required for reticule work, such as obtained with spectroscopic plates type 649 or with albumen and collodion emulsions [26, 27, 28, 30].

2. (P) Diazo image in gelatin layer. Transparent grainless image, many colors available.[8] Several patterns of different colors may be put on same reticule by applying same or different sensitizer to previous image layer and develop with appropriate developer to obtain desired colors. (Experiment with fixed out and washed lantern slide plates. Apply sensitizer with swab of absorbent cotton [36, 37, 38].

3. (N) Blue print image in gelatin, layer gives transparent grainless images [35, 38].

B. Single Discontinuous (Filmless) Layer

(Image as "islands" on glass surface)

1. (N) Dichromate sensitized colloid such as glue or gelatin containing silver halide or colloidal silver. Silver in the colloid image developed and intensified to make it black and opaque [26, 29, 31].

2. (N) Dichromate sensitized colloid, such as glue or gelatin. Image made black and opaque with alternate treatments of silver salts and reducing agents [27]. May also be blackened with lead sulfide [31].

3. (N) Cold top enamel. Dye in developer gives color and increases visual contrast [32, 33].

C. Single Stencil Layer (Metal Image Deposited on Glass Through Stencil)

1. (P) Dichromate sensitized colloid, glue cold top, etc., used to make a stencil layer. The glass and stencil is coated with a metal, such as silver, aluminum, chromium, etc., by vacuum evaporation. Removing stencil layer with solvents; also removes metal overcoat except in image areas where the metal is on the glass [34].

D. Two-Layer Process (Metal Layer Under Image)

1. (P) Glass coated with silver by chemical deposition or by evaporation. Dichromate sensitized colloid coated on top. After exposure and development, the uncovered undercoat is dissolved off with Farmer's reducer. Glue or gelatin image on top of silver may be intensified as in process 1B2 [25].

2. (P) Glass coated with lead sulfide undercoat. Dichromate sensitized glue coated on top of lead sulfide after exposure and development of colloid layer, after baking the glue remove uncovered lead sulfide with hydrochloric acid-bichromate bath [31].

TYPE II. PROCESS, SCALE OR RETICULE PATTERN FIRED OR FUSED IN OR ON GLASS SURFACE

A. Single Discontinuous Image Layer

1. Ceramic transparent [24].
2. Ceramic opaque [24]. (No details available.)

TYPE III. PROCESS SCALE OR RETICULE PATTERN BELOW SURFACE OF GLASS (ETCHED)

A. Single-Layer Process (Ideal Resist)

1. (P) Stencil layer resistant to hydrofluoric acid or to hydrogen fluoride [47]. This paper, section 8.

B. Two-Layer Processes; Resist Undercoat Not Light Sensitive

1. (P) Metal-film undercoat (usually silver). Overcoated with dichromate sensitized colloid. Metal uncovered after developing stencil layer is removed with Farmer's reducer or other suitable solvent. The exposed glass is etched with hydrofluoric acid or hydrogen fluoride [39, 25].

2. (P) Undercoat of asphalt, Bakelite, varnish, etc., overcoated with dichromate sensitized colloid. Undercoat exposed by developing stencil layer, is dissolved out with suitable solvent. The exposed glass is etched with hydrofluoric acid or hydrogen fluoride [40, 41, 42].

C. Two-Layer Process. Resist Overcoat on Stencil Layer

1. (P) Light sensitive undercoat of dichromate sensitized albumen. After exposure, albumen is rolled up with developing ink and developed as usual and dried. Inked surface dusted with powdered asphaltum and heated to fuse the powder with the ink. It is then etched with hydrofluoric acid [42].

D. Two-Layer Process. Fine-Grain Frosted Etch Continuous Permeable Overcoat on Resist Stencil Layer

1. (P) After developing stencil layers having complete or partial resistance to hydrofluoric acid, it is overcoated with dichromate sensitized glue, gelatin, gum arabic, etc., and exposed to harden this layer. This layer is not developed. After protecting back of glass with wax it is dipped in frosted type of etching solution for a suitable time (see section 7.3).

7.2. Survey of Resists for Photoetching

A number of the light-sensitive coatings used in photolithography and photoengraving were tried and found to be unsatisfactory for use as resists for etching the glass directly with hydrofluoric acid. The coatings would invariably crack, undercut, decompose, or become permeable to the etching solution before the etching was complete. Substances coated from water solutions such as glue, albumen, gelatin, gum arabic, polyvinyl alcohol, and cold-top enamel were permeable to the etching solution. With the exception of albumen and cold-top enamel, these materials did not give the sharp, crisp image edges necessary for very fine line (5 to 10 μ) work. Wood [43] and

[8] Not found in literature, but could be used where colored lines are wanted. The resolution of both diazo and blue print in a thin layer of gelatin is very high.

Smethurst [44] have reviewed many of the coatings used in lithography and photoengraving. Most of the hydrophobic light-sensitive resist coatings were found difficult to develop with the organic developer solutions; clean lines were not obtained without the exposed portion of the coatings being attacked by the developer. For example, methyl ethyl ketone-furfural resin was found difficult to develop as there was the tendency for the exposed part of the coating to be attacked by the developer when the development was sufficient to remove the unexposed portions of the coatings in the fine lines. The amount of exposure and the time and temperature of the developer were so critical that it was difficult, if not impossible, to fully evaluate their usefulness as a light-sensitive resist for etching fine line patterns. Sensitized asphalt [45, 46, 47] has been used as a resist, principally for etching of metals, but its low sensitivity to light and unsatisfactory development makes it difficult to use in production work. Some of these coatings require rubbing with a tuft of cotton during development because the unexposed areas are not sufficiently soluble to dissolve out and it must be removed by friction. This procedure cannot be recommended in the production of fine lines required in theodolite circles unless the adhesion of the coating to the glass is very strong and the resist tough enough so that the edges of the lines are not likely to be damaged.

Processes using an insensitive resist undercoat that is dissolved out through a light-sensitive overcoated stencil layer (type III–B) were considered. In these processes the glass is first coated with some material that is resistant to hydrofluoric acid, such as silver or other metal, Bakelite, asphalt, varnish, etc. For best results this layer should be coated as thin as possible but not so thin that it will lose its resistance to hydrofluoric acid. Further, it must be very uniform in thickness. On top of this subcoat dichromate sensitized glue, albumen, or cold-top enamel is coated with a whirler, exposed in contact with a positive and developed, thus forming a stencil on top of the resist undercoat. Suitable solvents are then applied to dissolve the resist undercoat through the stencil openings. The solvent must not affect the stencil layer. The resist undercoat must dissolve uniformly and completely. If some areas require more time for dissolving than other areas because of variation of thickness, the resist layer will be dissolved out from under the edge of the stencil, thus altering the dimensions of the pattern in the thin areas.

7.3. Photoetched Reflection-Type Circles

When this project was initiated there was an immediate demand for a few reflection-type graduated glass circles for theodolites. Solution of the problem appeared to be the use of a light-sensitive resist in a photographic process whereby the image of the lines and numbers would be printed on the circle blank in contact with a master negative. With a stencil type of resist the clear unexposed developed out areas could then be matte-etched so that the lines and numbers would appear black when viewed with vertical illumination, as explained in section 2.1 for type 2 circles. Several light-sensitive coatings were tested, but all failed as resists.

Further work led to the discovery that if the glass was coated with sensitized and exposed glue during the etching, the glue controlled the grain size of the etching. Used with the matte-etching solution ES–7 the grain size of the matte etch was the finest and most uniform observed in the present work. With mirror silver deposited on top of the etching it appears very white under ordinary white light illumination, but when viewed in an optical system with a vertical illuminator (illumination through the objective) it appears black, and of course the unetched parts (also mirror-silvered) appear white (see fig. 16). Gelatin and gum arabic were also effective, but the grain was appreciably larger than that obtained with photoengraving glue.

It was also found that by overlaying a developed image of an organic stencil layer coating with the glue, that the glue protected the stencil layer coating long enough to give the fine matte etch of the graduations in the glass surface. Cold top enamel was selected as the light-sensitive coating for this work because it develops readily and with smooth, sharp-edged lines. The light-sensitive glue was coated over the developed image of cold top enamel and exposed just long enough to actinic light before etching to make it insoluble in water.

The cold top enamel was sensitized according to the manufacturer's directions and coated by running a small stream of it from a separatory funnel on the glass circle on a whirler at 600 rpm. After drying the coated circle for about 10 minutes it was exposed in contact with the master negative, using an EH–4 mercury flood lamp with a ground-glass diffuser to make the illumination uniform. The exposure time was about 15 minutes at a distance of about 22 inches. The exposed coating

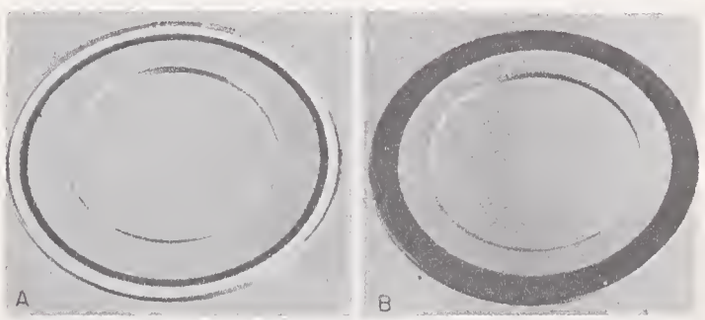


FIGURE 16. *Front and back views of the reflection-type circle.*

The circle (A) on the left shows a white etched band about one-eighth inch wide, the graduations are in the silvered band (dark) adjacent to it. The right-hand figure (B) is the back of a similar circle of slightly different dimensions, showing the copper plating to protect the mirror-silvered area.

was then developed and dyed in the cold top enamel developer (alcohol plus dye), rinsed in ethyl alcohol, and finally in acetone, which dried the coating. The cold top enamel is sensitive to moisture in the air and will not give a good image if exposed to a damp atmosphere too long. The coated glass circle should be exposed and developed as soon as possible after coating and drying. See section 8.7.b, for the effect of relative humidity on cold-top enamel. The developed cold-top image was then coated with sensitized glue on a whirler. With the whirler running at 100 to 200 rpm a small quantity (3 to 4 ml) of sensitized glue was poured over the circle. Then the speed of the whirler was immediately increased to 600 rpm and allowed to whirl until the glue was dry, usually about 15 minutes. The sensitized glue solution had the following composition:

Photoengraving glue	33 g.
Water	120 ml.
Ammonium dichromate	2 g.

This solution was made up as follows: A solution of 90 ml of water and 33 g of glue was heated in the steam bath in a sealed pressure bottle from 6 to 24 hours, or until the solution became clear. Then 2 g of ammonium dichromate dissolved in 30 ml of water was added to the cooled glue solution and filtered.

The sensitized glue was exposed about 8 minutes to a mercury EH–4 flood lamp at a distance of 18 inches. A $\frac{1}{16}$-inch-thick piece of ground glass over this lamp was necessary to obtain uniform illumination.

Before etching the glass, the back, edges, and the areas on the top surface, where no etching was desired, were painted with a wax solution. The circle was then etched by dipping in the following solution:

Etching Solution ES–7

Water	65 ml.
Ammonium bifluoride	100 g.
Sodium fluoride	11.5 g.
Hydrofluoric acid (48%)	50 ml.

After etching the circle about 15 to 20 seconds it was removed and quickly immersed in a 10-percent sodium hydroxide solution to stop the etching action and also to remove the cold top and glue.

After removing the wax with solvents it was chemically cleaned and silvered by the Rochelle salts process. A forebath, freshly prepared, of dilute stannous chloride (about 1 to 1,000) followed by rinsing in distilled water prior to immersion in the silvering solution will speed up the mirror silvering. When the silvering solution was spent it was replaced with a fresh one in order to get a heavy coat of silver. After silvering the circle, it was rinsed off, mounted on a fixture, and placed in a copper-plating bath (100 g of copper sulfate ($CuSO_4 \cdot 5H_2O$) per liter), and a layer of copper was electroplated over the silver for mechanical protection, using the procedure given on pages 10 and 11 of reference [48].

The fixture for holding the glass circle during the electroplating was constructed as follows: The circle ring was clamped between two Bakelite disks. These disks cover about half of the width of the ring. The silvered surface on the upper (unetched) side was contacted with spring fingers for the electric connection. It was then immersed in the plating bath with the top surface of the glass just above the solution. The copper was deposited on the edge and on the outside half of the etched side of the circle. After copper plating the circle it was washed and dried. A thin layer of shellac was painted over the copper for protection against corrosion. The appearance of these circles is shown in figures 16 and 17.

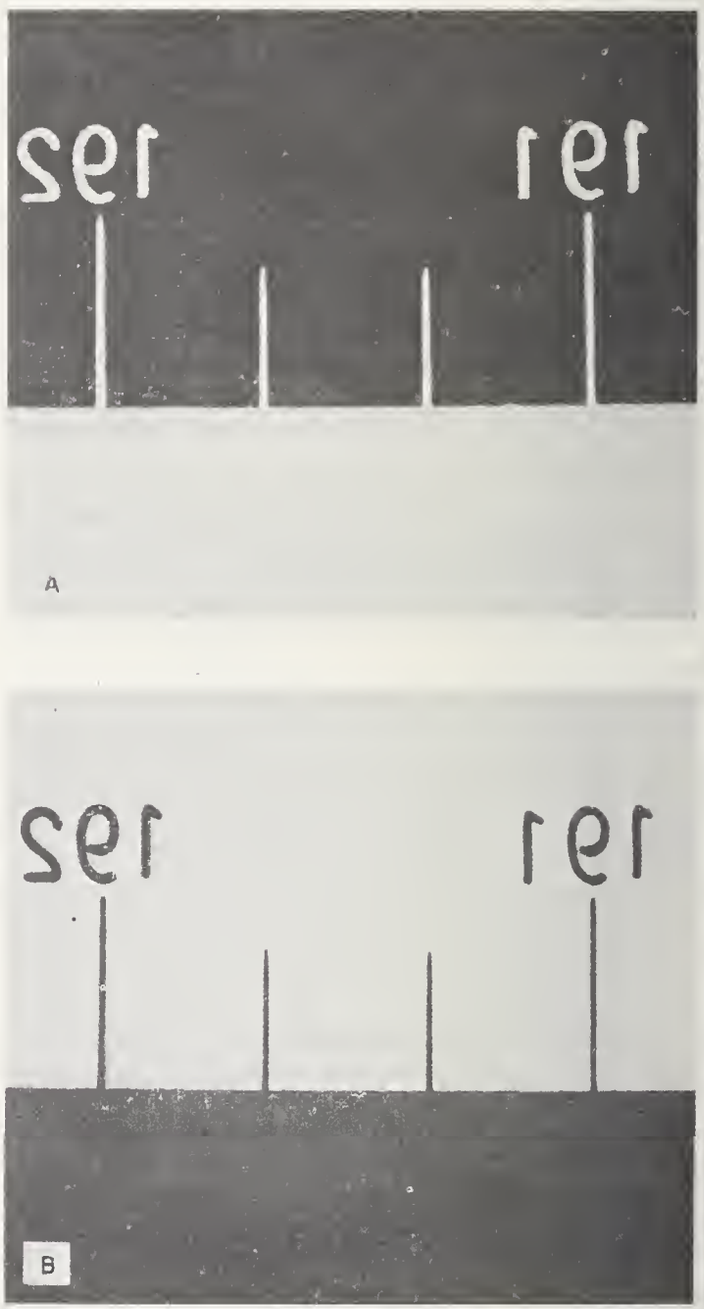


FIGURE 17. *Photomicrographs (×125) showing the appearance of the etched graduations and band of the same part of the circle under different conditions of illumination.*

The top figure (A) was made with diffuse illumination. The lower figure (B) was made by changing the illumination to specular light (illumination through the objective).

Extreme care is necessary in exposing the glue overcoat and an excessive exposure will cause the glue to crack. Too much heat at the time of exposure will also promote cracking by reducing the moisture content of the glue. These cracks may be too fine to see with the eye and also may be confined to a small area. When cracks occur, the etching acid reaches the cold top directly instead of by diffusion through the glue, and breaks through the cold-top enamel and etches the glass, thus spoiling the work. Trash in the glue will also promote break-throughs.

The phenol-formaldehyde light-sensitive resin developed after these circles were produced may be used in place of cold top enamel to produce a similar etching. It is much more resistant to hydrofluoric acid, and consequently the effects described above are less likely to occur.

8. Photoetching with the Phenol-Formaldehyde Resist

8.1. Phenol-Formaldehyde Resin

Phenol-formaldehyde resin was known to be resistant to the action of hydrofluoric acid, and it seemed possible that a thin layer coated on glass might serve as a resist for hydrofluoric acid etching after development of the image. Doelker [49] patented a process whereby phenol-formaldehyde resin was recommended as an etching resist for zinc plates. The resin was dissolved in alcohol and, immediately before coating, was sensitized with ammonium bichromate. The exposed plate was developed with a dilute solution of caustic potash. Beebe, Murray, and Herlinger [50] patented the use of light-sensitive phenol-formaldehyde resin for photoengraving processes. They sensitized the resin with iodine, iodoform, etc., and, after exposure, developed the coating in a solution of about equal parts of alcohol and water. The resin made according to their directions was soft, remained tacky too long after coating, and when sensitized, was low in light sensitivity. Any light-sensitive resin used in contact printing must become tack-free shortly after coating. In the course of the investigation it was found that the softness of the resin was due in part to the presence of methyl alcohol in the formaldehyde solution. The addition of methyl alcohol to trioxymethylene (solid) and phenol in a concentration corresponding to that generally found in 40-percent formaldehyde solutions yielded a soft, tacky resin. Therefore, trioxymethylene was used instead of the 40-percent aqueous solution of formaldehyde for production of the phenol-formaldehyde resin in this work. Also the alcohol-water developer used by Beebe et al [50] was not a satisfactory developer for the resin on glass.

Many organic substances are sensitive to or hardened by actinic light, but it is generally a difficult problem to find a developer solution that will dissolve out the unexposed portions of the resin and not attack the exposed portions. This difficult hurdle was overcome by developing an organic solution that dissolved the unexposed light-sensitive resin quickly and yet did not attack the exposed portion when it had been properly exposed to light even if the development was extended several times beyond the normal developing time. Iodoform was found to be a very good sensitizer. A resorcinol-formaldehyde resin was also made for study. It was found useful as an additive in the phenol-formaldehyde resin. A concentration of 5 percent of the resorcinol-formaldehyde resin improved the sensitivity and resist properties of the phenol-formaldehyde resin, although its presence is not necessary to make the phenol-formaldehyde resin function as a resist for etching glass.

Light-sensitive resists made with phenol-formaldehyde resins will not withstand the action of concentrated 48-percent hydrofluoric acid. However, an etching solution containing hydrofluoric acid and phosphoric acid (formula EP–1, section 5.4) was found that can be used to produce excellent etchings with etching time up to at least 15 seconds.

In the production of the theodolite circles, a 5-second etch was used to keep the lines as narrow as possible, and the etched portions were filled with lead sulfide chemically deposited as described for filling the etched graduations of masters in section 6.1. Even though the resin coating will not withstand a long period of etching, deep lines may be obtained with the special etching solution by etching the circle several times for 3 to 4 seconds and washing and drying it between each etch.

The details for the synthesis of these resins are given in section 9.

8.2. Sensitization of the Resin with Iodoform

Iodine sensitizes the resin to actinic light, but the sensitized solution is unstable and must be used as soon as sensitized. Iodoform proved to be a very satisfactory sensitizer, and the sensitized resin solution was very stable in methyl ethyl ketone when it did not contain water and was free from acid. Iodoform is unique as a sensitizer because the iodine is released by the actinic light during the exposure when the catalytic action of the sensitizer is needed. Both sensitized and unsensitized solutions of the resin should be stored in brown bottles in a dark cool room or preferably in a refrigerator. Sensitized solutions stored in brown bottles in the dark at room temperature over anhydrous sodium sulfate were still usable after 1 year, and when stored in the refrigerator were good after 3 years. Acetone could not be used as a solvent for the resin because the iodoform underwent slow decomposition.

The effect of the concentration of the iodoform on the sensitivity of the resin was determined in the following manner: Four 100-ml solutions of the phenol-formaldehyde resin containing 20 g of resin each were prepared from the same resin batch. To each solution was added different amounts of iodoform. Glass test plates 1 by 3 inches were coated by drawing them out of the sensitized resin solution at the rate of 1 foot per 80 seconds, using a synchronous motor-driven apparatus similar to that described in section 3.3. After storing the coated glass plates in the dark at room temperature and approximately 50-percent relative humidity for 4 hours they were given progressively increasing exposures in contact with a glass master ⅜ inch thick to an EH–4 mercury flood lamp placed at a distance of 11 inches from the glass-plate grid. After development, the minimum exposure time necessary to give a good image was determined by inspection with a microscope at ×80. The exposure times giving the best images were plotted against the concentration of the iodoform in percent, expressed as the percentage of iodoform based on the concentration of the resin. These results are shown in figure 18. Crystals appeared in the dried resin coating when the concentration of iodoform reached about 20 percent, indicating that the iodoform concentration had exceeded the saturation point. A concentration of 10 percent of iodoform as sensitizer based on the resin concentration was selected because it gave good sensitivity to actinic light without getting too close to the saturation point.

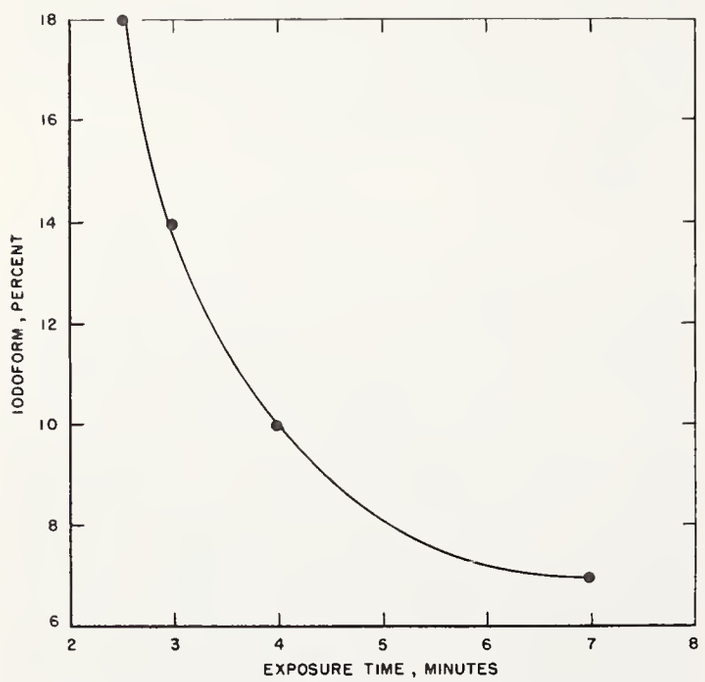


FIGURE 18. *Effect of the iodoform concentration on the exposure time.*

The following coating formula was used for the production of the theodolite circles:

Phenol-formaldehyde resin, 25% solution. 800ml.[a]
Resorcinol-formaldehyde resin ..... 10 g.
Iodoform [b] ..... 21 g.
Methyl ethyl ketone to make ..... 1 liter.

Then add about 100 g. of anhydrous sodium sulfate to the solution, shake occasionally, and allow to stand overnight. Filter the solution through a coarse fritted-disk filtering funnel without pressure.

[a] The concentration of the phenol-formaldehyde solution, which varied from batch to batch, ranged from 22.5 to 30 g of resin per 100 ml of soluton. The volume of solution used was adjusted so that 200 gof resin is used in the above formula.

[b] The iodoform used was Eastman Kodak Cat. No. P-341.

8.3. Coating the Circles

The phenol-formaldehyde coating is somewhat less sensitive to light than cold top enamel. Either of these materials can be used in a room lighted with incandescent lamps if care is taken to protect the dry coating from direct strong light either by shielding or keeping it covered. The circles were coated in a closed cabinet to protect them from drafts, dust, and light during the coating operation. The glass circle was coated by the "pull-out" method described in section 3.3 and shown schematically in figure 7. The sensitizing solution was contained in a glass battery jar 2⅞ inches by 5¼ inches and 7 inches deep (inside dimensions). It was filled to about ¼ inch of the top of the jar with the resist solution. The top edges of the jar were ground flat and when not in use, was covered with a glass plate to prevent evaporation.

The circles were cleaned and dried by the same procedure as that described in section 3.2 for coating the engraving resist, except they were not treated with amino silane (waterproofing agent). The dried circle was placed on a glass tripod (fig. 6) and washed in filtered methyl ethyl ketone by swabbing with cotton, rinsed with filtered methyl ethyl ketone, and quickly placed on a wire hook and immersed in the sensitized resin solution. The wire holder is designed so that the inside of the circle rests on the wire without touching the front or back side of the circle. After immersion, the circle was moved up and down a few times within the solution to remove any bubbles of air that might adhere to the glass. When the surface of the solution became quiet the circle was drawn out with the synchronous motor at the rate of 1 inch per 6¾ seconds. The door of the cabinet was kept closed after the immersion of the circle and opened after it had been drawn from the solution. The circle was clasped with an aluminum-band holder and placed face down over the neck of a 300-ml Erlenmeyer flask. The resin coating on the circle was allowed to dry at least 1 hour before exposure, but 4 hours was found to be better. If this precaution is not observed, there is danger that some of the resist coating may stick to the master.

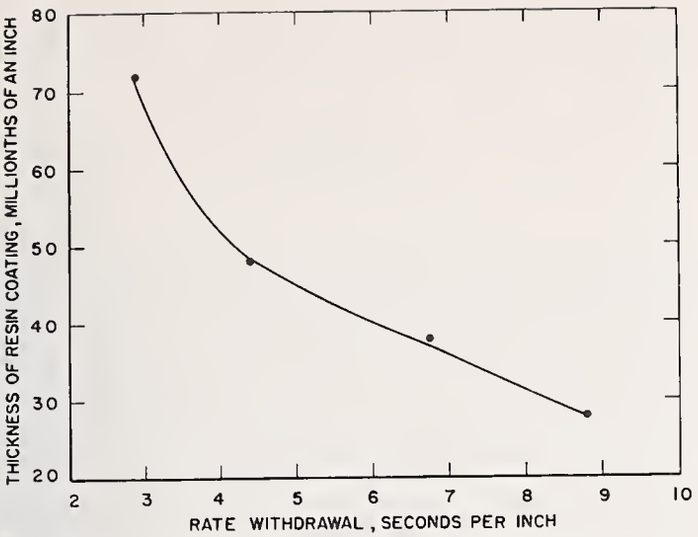


FIGURE 19. *Thickness of the resin coating as a function of the rate of withdrawal of the glass from the coating solution.*

The resin concentration of 21 g per 100 ml of solution, and the rate of withdrawal of 1 in. per 6¾ seconds yielded a thin and very uniform coating on the glass circle, which, after development, measured 38 millionths of an inch thick. The solvent, methyl ethyl ketone, evaporates fast enough at room temperature to leave the desired coating on the glass as it is drawn from the solution. The thickness of this resin coating must be very uniform as a slight change in thickness affected the time of exposure, which decreased with an increase in thickness. During the coating operation the circle must not twist, turn, or touch the sides of the container because variations in the thickness of the coating will result. The thickness of the coating may be changed by varying the concentration of the resin or the rate of withdrawal. The thickness of the developed coating was determined for four different rates of withdrawal. The thickness of the developed samples was measured by the interference microscope, and the results are given in figure 19. A. G. Strang of the NBS Engineering Metrology Section made the measurements of the thickness of the resist images. The withdrawal rate of 1-inch in 8⅘ seconds gave a coating that was too thin. The curve shows that for the solvent used the thickest coating obtainable would be about 50 millionths of an inch at the rate of withdrawal of 1 inch in 4⅕ seconds because beyond this rate the evaporation of the solvent is not sufficiently complete, so that the solution drains off, causing a nonuniform coating. Of course the thickness of the coating is also determined by the concentration of the resist and the temperature.

8.4. Contact Printing and Exposure Effects

The sensitized resin coating is sensitive to the violet and ultraviolet parts of the spectrum, and consequently, light sources having high energy in this spectral region, such as sunlight, carbon arc, and the mercury vapor lamp, are necessary. The light source used for the experimental work was an EH–4 mercury flood lamp. A 3/8-inch-thick glass plate containing filled etched lines 10 μ wide was mounted 11 inches from the face of the lamp. The coated glass plates were 1 by 3 inches, and were exposed in contact with the line image. The different batches of resin synthesized in this work had exposure times varying from 2 to 5 minutes with the above described exposing apparatus. The experimental glass plates were coated by drawing them from the sensitized resin solution with a synchronous motor. An AH–6 high-pressure mercury-quartz lamp, cooled by running water, was used as the light source for the production of the theodolite glass circles. This water-cooled lamp develops a mercury-vapor pressure of 110 atmospheres and operates at 840 volts and 1.4 amperes. The lamp, transformer, magnetic switch, and water solenoid valve were mounted in a cabinet. The master negative ½ inch thick was mounted 19 inches from the light source. A penthouse fitted with a door was constructed on the top of the cabinet. It served to protect the glass master and to reduce the hazard from dust particles. A 25-watt lamp located below and to one side of the glass master was turned on to observe any dust particles that might be present on the master. If present, they were removed or picked off with a camel's hair brush. The circle was removed from the Erlenmeyer flask by means of the aluminum clamp and placed carefully in position on the master around the centering plug and washer, which fits into the hole in the circle glass ring. The circle was pressed down lightly with the fingers and covered with a black felt cloth. The 25-watt lamp was turned off, and the resin coating was exposed to the AH–6 lamp by opening a small door type of shutter located above the AH–6 lamp. After exposure the circle was removed from the master negative with the aluminum clamp and placed on the glass tripod. The exposure time was about 12 minutes.

It was found that the polymerization of the resin continued after the action of the light was stopped. Thus the image of lines may be narrowed slightly (similar to the effects of a more full exposure) by storing the exposed coating in the dark for 15 minutes or longer, but if continued too long, an overexposure effect takes place. When the sensitized resin was given a one-half normal exposure and stored 1 to 3 hours in the dark it generally developed a good image, but a fully exposed plate when stored in the dark too long will become overexposed. While the above effects may be useful in some instances, it is recommended that the exposed resin coating be developed as soon as the exposure is complete. It is also interesting to note that the sensitized resin coating becomes insoluble and cannot be developed if kept in the dark at 50° C for 4 hours. This is due to the thermosetting properties of the resin.

8.5. Development of the Exposed Resin Coating

The unexposed areas of the resin coating were dissolved out (developed), and the exposed areas were dyed simultaneously by means of an organic solution developed for use on this resin coating. The developer solution was made up as follows:

Ethyl alcohol 95 %	1000 ml.
Normal butyl alcohol, bp 117° to 117.8° C	1000 ml.
Methyl ethyl ketone, bp 79° to 80.5° C	200 ml.
Methyl violet	2 g.
Basic fuchsin (aniline red, magenta)	2 g.

The developing solutions were filtered through a coarse fritted disk filtering funnel.

The ethyl alcohol, normal butyl alcohol, and methyl ethyl ketone were of reagent grade and were used without distillation or drying. The developer solution contained about 3 percent of water. The mixture of methyl violet and basic fuchsin dyed the resin coating a heavy blue-purple and was found to be the best combination of dyes tested. Methyl violet dyed the resin coating blue-green; basic fuchsin, red; and malachite green, green. Another developer solution was also made up without the dye for use in washing off the excess dye on the surface that accumulates during the developing procedure.

The exposed resin coating was developed as follows: The circle mounted on the glass tripod was twirled gently for 30 to 40 seconds in the developer-dye solution; then twirled gently for about 3 seconds successively in three developer solutions that contained no dye; and placed at once on the whirler to dry. All solutions were at room temperature. The resin coating becomes soft during development but quickly hardens on drying in about 2 minutes. Although the development took place in 30 to 40 seconds, there were no adverse effects on the resin coating if the developing time was extended for as long as 2 minutes. No rubbing is required because the unexposed portion of the resin dissolves very quickly. This development procedure is very practical because the development of most organic solutions require close control of time and temperature.

The activity of the developer may be increased by increasing the concentration of the methyl ethyl ketone, and this might be necessary in instances where the resin has condensed very close to the critical point due to age and also to temperature being too high during storage. It was also found that if the sensitized resin had condensed so far that it did not give complete development by the normal procedure or by increasing the concentration of the methyl ethyl ketone, that the addition of 25 to 50 percent of a fresh sensitized resin batch made it capable of proper development. Sensitized resin batches were found to be readily developable after storage for 1 year at room temperature. Sensitized solutions of the resin increased slightly in sensitivity on storage at room temperature. One sample of sensitized resin that was stored in a refrigerator developed a good image after 3 years.

8.6. Etching and Filling the Graduations in the Circles

After the development of the resin coating it was baked to increase its resistance to the hydrofluoric acid. The coated circle supported on an Erlenmeyer flask was placed in an oven at room temperature, and the temperature of the oven was raised to 100° to 135° C and maintained at this temperature for 3 hours. The oven was turned off, and the sample was allowed to cool nearly to room temperature before removal. This precaution of heating and cooling slowly was taken to avoid the possibility of breaking the glass due to thermal shock. Concentrated hydrofluoric acid solutions (48 percent) attack the phenol-formaldehyde resin coating. The problem was to find a solvent for the hydrofluoric acid that would delay its attack on the resin coating and prevent the formation of insoluble salts during the etching. Experimentally, the answer was found in using phosphoric acid in aqueous solution with hydrofluoric acid. Also phosphoric acid assists the hydrofluoric acid in etching the glass [18]. The phosphoric acid retards the attack of the hydrofluoric acid on the developed resin and gives a clear etch when sufficient water is added. The following formula was found to be satisfactory:

Etching Solution EP–1

Phosphoric acid (85%)	100 ml.
Water	50 ml.
Hydrofluoric acid (48%)	25 ml.

The phosphoric and hydrofluoric acids were of reagent quality, meeting ACS specifications. This etching solution is the same one used in etching the ruled wax-coated circles described in section 5.4. The etching solution was simple to make up as no noticeable heating took place when the chemicals were mixed. The etching solution may be used as soon as it is made up, and, if stored in a wax pot fitted with a tight cover, it will keep for several months. Decreasing the water in the above formula gave a coarse matte etch, whereas increasing the water and hydrofluoric acid yields a more rapid etching solution, giving a clear etch. Although the theodolite circles were made from plate glass, borosilicate glass etched equally well.

The glass circle was etched by masking the areas not covered by the resin with a mixture of 20 g of gilsonite and 5 g of burgundy pitch in 100 ml of benzene. The fingers must not at any time touch the portion to be etched. The circle mounted on the glass tripod was immersed in the etching solution for 5 seconds with a gentle twirling movement and then immediately immersed in a 20-percent solution of sodium hydroxide solution to neutralize the hydrofluoric acid. Then the

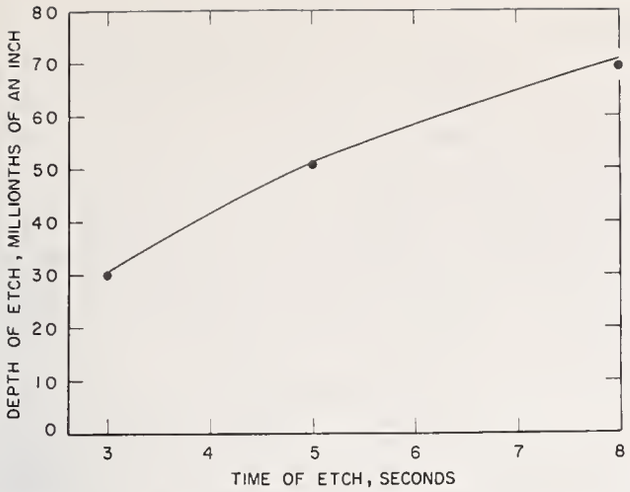


FIGURE 20. *Depth of etch versus time of etch, using etching solution EP–1.*

circle was washed, dried, and the wax masking coating removed with carbon tetrachloride in a crystallizing dish under the hood while mounted on the glass tripod. The remainder of the resin coating was removed by first treating with a 20-percent solution of sodium hydroxide and then with a cleaning solution (potassium dichromate and concentrated sulfuric acid). The etched lines and numbers were filled with lead sulfide, using the same procedure described in section 6.1.

The resin coating withstood the etching solution about 15 seconds at room temperature. This etch is not quite deep enough to fill the 10-μ, lines with pigments as described in section 6.2. However, a deeper etch on the circle was not desireable as it widened the lines. Unlike reticules, the cost of the glass circle was too high to risk damage by scratching. The etching used on the circle was deep enough to fill with lead sulfide by chemical deposition. The etched lines in the circles after filling averaged 1 to 2 μ wider than those in the resist layer. With full exposure the line width in the developed resist layer may be made slightly narrower than those on the master negative. Tests were made on the glass circles to determine the depth of the etch for different etching times. The depth of the etch was measured with an interference microscope, and the results are shown in figure 20. As might be expected, the depth of etch is nearly proportional to the time of etch. However, it was found that a deep etch may be obtained with the phenol-formaldehyde resin coating by etching 3 to 4 seconds several times with the above etching solution. After each etch the etching solution is removed immediately with copious amounts of distilled water and the glass dried at 100° C for about 30 minutes, cooled, and the etching repeated. Three to four such etches were found to give a deep etch that could be filled with pigments.

The appearance of the graduations of a master and a circle made by contact printing with the phenol-formaldehyde resist is shown by the photomicrographs in figure 21. It may be noted that the graduations lines on the circle (B) are slightly wider than those on the master (A) by about 1 μ.

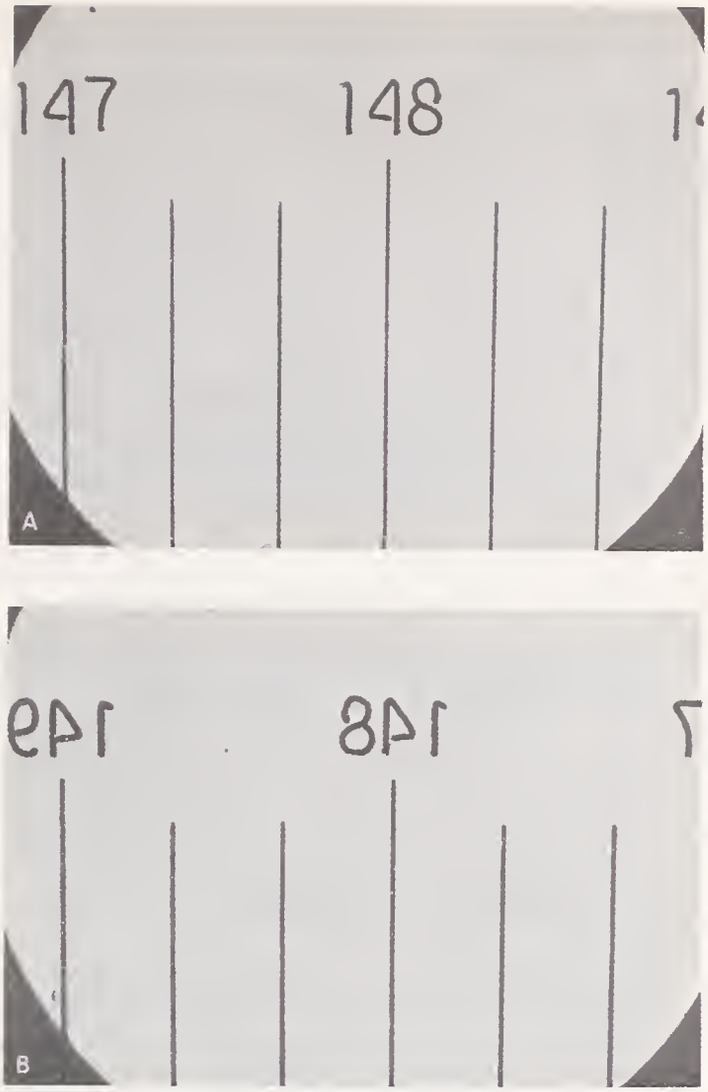


FIGURE 21. *Photomicrographs of master and circle made from it by contact printing, using the light-sensitive phenol-formaldehyde resist.*

A, Master; B, circle. Both have been etched and filled with lead sulfide by chemical deposition.

8.7. Effect of Relative Humidity on the Phenol-Formaldehyde Resin and Cold Top Enamel

a. Phenol-Formaldehyde Resin

It is well known in the photomechanical industry that glue, albumen, cold top enamel, and other light-sensitive coatings are affected by high humidity, and the effects are peculiar to each particular coating. The effect of atmospheric moisture on the sensitivity and developability of the phenol-formaldehyde resin and cold top enamel used in this work was studied. A humidity cabinet having a volume of 7 cubic feet and designed for determining dimensional changes in photographic materials as a function of relative humidity was used to subject the sample coatings to the desired relative humidity. This cabinet is

well insulated and is completely moisture tight. A fan on a long shaft (motor outside) drives a stream of air down onto the surface of a tray of saturated salt solution, from which it flows into the partitioned space containing the test samples. The air then passes over the top of the partition and back to the fan, giving a continuous stream of conditioned air. The relative humidity is controlled by using saturated salt solutions, which give the desired relative humidity. This apparatus is efficient and will condition photographic film and paper in about 2 hours.

Tests of the effect of the relative humidity on the sensitivity and developability of the phenol-formaldehyde resin were made at 14-, 51-, 65-, 76-, and 88-percent relative humidity. For each relative humidity, coated glass plates were exposed in contact with a glass master having filled etched lines about 10 μ wide.

At different time intervals samples were removed from the humidity cabinet and given the same exposure and development as that needed for control samples run at the same time. Both the control and tested samples were subjected to room relative humidity (about 50 percent) and room temperature (25° to 30° C) for about 15 minutes after coating. The control samples were exposed and developed 15 minutes after coating. The evaluation of the tests was made with a microscope by noting the quality of the developed image.

The results of the tests are summarized in table 5 and show that at a relative humidity of 50 percent or less the relative humidity has little effect on the sensitized coating and that the effect of relative humidity up to 75 percent is not serious if the light-sensitive resin coating is exposed soon after coating. High relative humidity in the humidity cabinet affected the sensitivity and developability very rapidly, and this was reflected in a weak image with loss of gloss and a partially dissolved or decomposed coating. However, under practical working conditions in still air the deterioration of the sensitivity and developability is not nearly as rapid as that indicated by the tests in the humidity cabinet. Coated glass plates stored in a dark box at room temperature and humidity for 5 days showed only a slight loss in sensitivity and developability even though the weather was damp and it had rained once during the period and the room was not artificially heated.

In the production of theodolite circles no loss of sensitivity and developability was experienced when the coated samples were stored 4 hours before exposure and development. The tests in the humidity cabinet accentuate the difference in the effect on sensitivity and developability on the phenol-formaldehyde resin coating when it is stored in a stream of rapidly moving air as compared to its storage in still air of the same relative humidity.

TABLE 5. *Effect of relative humidity on the sensitivity and developability of the sensitized phenol-formaldehyde resist coating in the relative-humidity cabinet*

Relative humidity	Time to cause loss of sensitivity and developability as image takes on a dull, weak appearance
%	
4 [a]	Not affected after 55 days.
14 [b]	Not affected after 4 days.
51	48 to 72 hours.
65	About 6 hours.
76	30 to 60 minutes.
88	About 5 minutes.

[a] Sample was stored in a desiccator over anhydrous calcium chloride.
[b] Fan not running.

b. Cold Top Enamel

Experiments similar to those described in the section 8.7, a, were also made with sensitized cold top enamel. Cold top enamel when freshly sensitized is "soft" working and becomes more contrasty with age. Solutions that are about a week old give best results for line work on glass. During periods of high relative humidity it was difficult to get satisfactory development of the exposed cold top enamel. Apparently the presence of moisture in the cold top enamel after coating promotes a chemical reaction that causes a hardening of the unexposed areas with time, rendering them less soluble.

Increasing the relative humidity increases this chemical action, and the image becomes partly or completely undevelopable even if the exposure time is shortened. About 1 hour was about as long as a glass plate coated with cold top enamel could be stored at 60-percent relative humidity and 23° C. The results of the tests made on the cold top enamel are summarized in table 6 and show that increasing the relative humidity causes the coating to become insoluble (undevelopable) and to take on a dull, weak appearance.

TABLE 6. *Effect of relative humidity on the sensitivity and developability of sensitized cold top enamel in the relative-humidity cabinet*

Relative humidity	Time in cabinet	Effects noted
%	*Min*	
51	30	Lines developed.
	60	Lines incompletely developed.
76	15	Lines developed.
	30	Lines incompletely developed.
95	5	Lines developed.
	10	Lines incompletely developed, coating weak and dull in appearance.

The tests further indicate that cold top enamel when used as a sensitized coating on glass should be handled in an atmosphere of 50-percent relative humidity or less and that the exposure should be made as soon as the coating is dry and developed immediately. To overcome the effects of high humidity, one manufacturer of cold top enamel [33] recommends heating the plate on a whirler to remove moisture from the coating and to prevent sticking during contact printing. It was not found necessary to use heat with relative humidity at 50 percent or lower.

9. Synthesis of Phenol-Formaldehyde and Resorcinol-Formaldehyde Resins

9.1. Phenol-formaldehyde (Stage-A) Resin

Baekeland [51, 52, 53] divided the phenol-formaldehyde reaction during resinification into three stages, A, B, and C, which are characterized by the appearance and solubility of the resin. The A-stage, or initial product, is a low molecular weight condensation product soluble in acetone; the B-stage, or immediate condensation product, is swelled by acetone but not soluble in it; and the C-stage, or final condensation product, is infusible and completely insoluble in acetone. Sometimes the A-stage resin is called a resol; the B-stage, a resitol; and the C-stage, a resite. The resin used in this work is that formed in the A-stage (resol) and condensed to the state where it is just soluble in 95-percent ethyl alcohol. The resin is a thermosetting phenolic that slowly passes into B-stage (resitol) on drying at room temperature. After the condensation reaction was completed the resin was pressed free of as much water as possible, dissolved immediately in methyl ethyl ketone, and dried over anhydrous sodium sulfate. When the resin solution was made up to about 20 g of resin per 100 ml of solution it kept in the dark at room temperature for at least 3 years, provided the solution was dry and free from acid.

Sodium acetate, a weak alkaline salt, proved to be the best catalyst for the condensation of the phenol and formaldehyde. It yielded a resin that had the greatest light sensitivity and that was readily developable by an organic solution. Other salts tested as catalysts were potassium acetate, lithium acetate, calcium acetate, barium acetate, and sodium formate. These resins were low in light sensitivity and could not be as easily developed as the resin formed with sodium acetate. Acids such as hydrochloric acid act as strong catalysts and produce a resin that is not as light sensitive as that produced by the weak alkaline catalysts mentioned above and further the condensation reaction is difficult to control because the resin passes rapidly from the A-stage to the B-stage.

Investigation of the condensation reaction showed that 1 mole of phenol, 3 moles of trioxymethylene, 6 moles of water, 0.3 mole of sodium acetate ($CH_3COONa \cdot 3H_2O$) produced a satisfactory light-sensitive resin that was not too soft. Decreasing the sodium acetate gave a softer resin that was less light sensitive.

The following formula and procedure were worked out for synthesizing the light-sensitive resin used in printing theodolite circles from a glass master negative:

Phenol	282 g.
Water	324 g.
Sodium acetate ($CH_3COONa \cdot 3H_2O$)	122 g.
Trioxymethylene (paraformaldehyde)	270 g.

The phenol and trioxymethylene were USP grade, and the sodium acetate was reagent quality, meeting ACS specifications. Distilled water was used in the synthesis and tap water in washing the resin.

The apparatus consisted of a 3,000-ml round-bottomed flask fitted with a reflux condenser by means of a ground glass joint and mounted on a ring stand. The apparatus should be located in a chemical hood. The chemicals were added to the flask in the order listed in the formula, followed by the addition of boiling chips made from an unglazed porcelain plate. The mixture was heated through an asbestos pad with a Bunsen burner to a gentle boil. The ring stand was rocked gently and frequently with the hand to prevent superheating until the boiling began. The boiling was continued throughout the period of the reaction, and the time required was 2½ to 3½ hours. The temperature of the boiling solution during the reaction before precipitation was 105° C. Toward the end of the condensation reaction the solution became cloudy, and finally opaque to transmitted light about 20 minutes, before two definite phases developed, accompanied by foaming. After the two phases appeared, gentle and frequent rocking of the ring stand was helpful to prevent the resin, which slowly precipitates from the boiling mixture, from sticking to the flask. After the two phases appeared, the boiling was continued about 5 minutes, and the hot reaction mixture was immediately poured into cool water with rapid stirring, thus completing the precipitation of the resin from the hot reaction mixture. The resin was washed and kneaded in about 5 changes of water and tested for degree of condensation. This test was made by taking about 10 g of the resin batch and heating it on the steam bath in a small porcelain dish while stirring with a glass rod and the time noted for the resin to pass into the B-stage. This was observed visually when the resin became rubbery, stringy, or lumpy and is the point at which all or most of the resin becomes insoluble in 95-percent ethyl alcohol. If the resin has condensed close to the B-stage the small sample of resin should become insoluble in 95-percent ethyl alcohol after heating for 10 minutes on the steam bath, and if this is the case, the large resin batch is sufficiently condensed. If the time was greater than 10 minutes, the precipitated resin batch was heated on the steam bath in a large porcelain dish while stirring constantly with a glass rod, and the time of heating was estimated from the time required for the small sample to become insoluble in 95-percent ethyl alcohol. With experience, the operator will find no difficulty in determining the end-point of the condensation. As a safeguard against carrying the condensation too far on the steam bath the resin batch may be

cooled with water at intervals and a small sample tested for degree of condensation as described above. The resin passes rapidly from A-stage to B-stage at a temperature of 100° C, and if the resin shows any signs of stringiness or a tendency to be rubbery and lumpy, it should be cooled immediately by stirring in cool water. If a few insoluble particles appear in the alcohol solution, the resin is usable, but if it is essentially insoluble in the alcohol, it should be discarded. The highest degree of light sensitivity along with developability was obtained by condensing the resin as close to the B-stage as possible, without the resin becoming insoluble in 95-percent ethyl alcohol.

After the resin had been condensed to the desired state and cooled, as much of the water as possible was pressed out and the resin dissolved immediately in 700 ml of methyl ethyl ketone (2-butanone) of a good commercial grade, free from acid or any substance that would react with iodoform. Then 500 g of anhydrous sodium sulfate was added and the resin solution allowed to stand for several days, with occasional shaking. After filtering the resin solution, it was ready for sensitizing and use. The amount of resin determined from the dried resin solutions of different batches gave yields of 220 to 300 g. The dried resin solutions contained from 22.5 to 30 g of resin per 100 ml of solution. The concentration of the resin in the filtered solution was determined by noting the volume, evaporating 10 ml of the solution to dryness in a tared 50-ml beaker, and weighing.

Another formula containing ethylene glycol as part of the solvent was developed. This resin was not used in the production of the theodolite circles as the tests on it had not been completed. The resin has the same sensitivity as that produced by the above formula. The resin is easier to produce by using this formula, but a trace of ethylene glycol remains in the resin. Ethylene glycol, because of its hydroscopic nature, causes a slight decrease in the light sensitivity of the resin coating when working at high relative humidities. This effect is nil when working with the resin at a relative humidity of 50 percent or less. The chief advantage in using this formula is that the resin is generally sufficiently condensed when precipitated so that it needs no further treatment on the steam bath. However, it should be tested for degree of condensation as described above before dissolving it in the methly ethyl ketone.

The following alternative formula was worked out for synthesizing the phenol-formaldehyde resin:

Phenol	282 g.
Ethylene glycol	186 g.
Water	270 g.
Sodium acetate ($CH_3COONa \cdot 3H_2O$)	122 g.
Trioxymethylene (paraformaldehyde)	270 g.

The ethylene glycol was of reagent grade. The apparatus and procedure for producing the resin were the same as described for the previous formula. The time required for the reaction was 2 to 3 hours. The temperature of the boiling solution during the reaction before precipitation was 107.5° C. The reaction mixture remained clear until very close to the end of the reaction. The solution became cloudy and the boiling was continued about 3 minutes after it had become opaque to transmitted light. The resin was immediately precipitated from the hot reaction mixture by pouring into cool water with rapid stirring. The resin was treated in the same manner as that obtained from the first formula.

9.2. Resorcinol-Formaldehyde Alcohol-Soluble Resin

The phenol-formaldehyde resin works well as a resist, and this resin was used in most of the experimental work. However, it was thought that the incorporation of other substances might improve the sensitivity and resist properties of the resin coating. Small amounts of shellac in the phenol-formaldehyde resin caused the sensitized solution to become unstable. The resorcinol-formaldehyde resin, because of its reactivity and chemical similarity to the phenol-formaldehyde resin, should improve the sensitivity and resist properties of the latter, and this was found to be true. A resin, soluble in 95-percent ethyl alcohol, was produced from the condensation of resorcinol and formaldehyde. While this resin was light sensitive, it was not possible to develop the image satisfactorily. Like many of the organic resins, no differential solvent was found that would dissolve the unexposed areas and leave the exposed image on the glass. It is somewhat more resistant to hydrofluoric acid than the phenol-formaldehyde resin. The addition of 5-percent of the resin, based on the weight of the phenol-formaldehyde resin, increased the sensitivity about 25 percent, increased the intensity of the dyeing of the image during development, and improved its resist properties slightly.

The following formula and procedure were used in synthesizing the alcohol-soluble resorcinol-formaldehyde resin:

Resorcinol	55 g.
Water	75 ml.
Sodium sulfate, anhydrous	12.5 g.
40-percent formaldehyde solution	41 ml.

The resorcinol was Eastman Kodak Cat. No. P–222, and the sodium sulfate was reagent quality, meeting ACS specifications. The formaldehyde was USP grade. Distilled water was used both in the synthesis and in washing the precipitate. The chemicals were added to a 250-ml Erlenmeyer flask and brought into solution. The reaction mixture was stored without agitation in a constant temperature room at 22° ±1° C. The solution

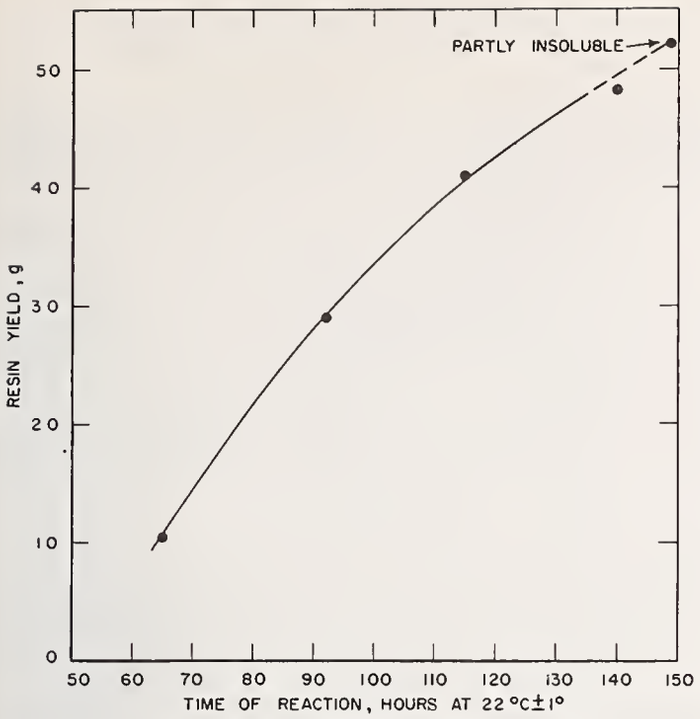


FIGURE 22. *Yield of resorcinol-formaldehyde resin versus time of reaction.*

was clear after 24 hours, but after 40 hours a precipitate began to settle out. At the end of 140 hours the reaction mixture was diluted to 1,500 ml, and then 100 g of anhydrous sodium sulfate was added. The sodium sulfate salts out the resin during the reaction and also on dilution before filtering. The precipitated resin was removed with suction by means of a Buchner filtering funnel. The precipitate was thoroughly washed, air dried overnight, and finally dried in the dark in a desiccator over anhydrous calcium chloride. The yield of the ethyl alcohol-soluble resin was determined for different times of reaction, and the results are shown graphically in figure 22. The resin became insoluble in 95-percent ethyl alcohol when the reaction time was extended much beyond 140 hours.

The rate of the condensation reaction increases with increase in temperature accompanied by an increase in the yield of an alcohol insoluble resin. At 100° C the reaction is very rapid, and the resin is completely insoluble in ethyl alcohol. It was difficult to obtain a good yield of the alcohol-soluble resorcinol-formaldehyde resin when the reaction was carried out at a temperature much above 25° C.

The dried resin when stored in a dry brown bottle kept in the dark at room temperature was still soluble in 95-percent ethyl alcohol after 2 years. The resin when dissolved in methyl ethyl ketone over anhydrous sodium sulfate at a concentration of about 20 g per 100 ml of solution was still usable after 3 years of storage at room temperature.

The authors welcome this opportunity to express their indebtedness to Benjamin L. Page for the numbering and ruling of the graduations in the wax resists of the master circles and for the calibration of the circles produced in this work, as well as for the many helpful suggestions during the progress of this work. The authors also express their thanks to Emory J. Stovall, Jr. and Henry G. Dorsett, Jr. for their assistance in the production of the reflection-type circles, using cold top enamel.

10. References

[1] Benjamin L. Page, The graduation of precise circles, Surveying and Mapping **13** No. 2, 149–161 (1953).
[2] Heinrich Wild, U. S. Patent 1,508,585 (1924).
[3] Thomas Y. Baker and Ralph W. Cheshire, U. S. Patent 1,754,872 (1930).
[4] Willard L. Egy, U. S. Patent 1,864,896 (1932).
[5] Knowlton H. Rich, U. S. Patent 2,188,014 (1940).
[6] Willard L. Egy, U. S. Patent 2,188,038 (1940).
[7] Heinrich Wild, U. S. Patent 2,221,317 (1940).
[8] Wild catalogue Geo. 20e, Geodetic and topographic instruments, Wild astronomical theodolite T4 (Heerbrugg, Switzerland).
[9] Franklin D. Jones and Edward K. Hammond, Graduating, engraving and etching, Machinery 1–19 (Sept. 1917).
[10] Hans Schulz, Glasätzung, Die Glashütte-Dresden A24, 828 (Nr. 48/1938).
[11] H. Bennett, Commercial waxes (Chemical Publishing Co., Inc., Brooklyn, N. Y., 1944).
[12] Albin H. Worth, The chemistry and technology of waxes (Reinhold Publishing Corp., New York, N. Y., 1947).
[13] Herbert Abraham, Asphalts and allied substances, **1**, (D. Van Nostrand Co., Inc., New York, N. Y., 1945).
[14] Donald Hubbard, Given W. Cleek and Gerald F. Rynders, Electrode function (pH response), hygroscopicity and chemical durability of Na_2O-CaO-SiO_2 glasses, J. Research NBS **44**, 247 (1950) RP2076.
[15] Alfred McEwen, Microscopic test plates, Watson's Microscope Record, Nos. 8 and 9 (W. Watson & Sons, Limited, London WC1, 1926).
[16] Lyman Nichols, Microscopic ruling, The Instrument Maker, p. 20 (March-April 1947).
[17] N. Irving Sax, Handbook of dangerous materials, p. 198 (Reinhold Publishing Corp., New York, N. Y., 1951).
[18] Bruno Schweig, Principles and methods of etching and embossing, Glass, p. 143 (April 1938); p. 184 (May 1938).
[19] Dry etching of glass (unpublished).
[20] W. Merte, R. Richter, and M. v. Rohr, Handbuch der wissenschaftlichen und angewandten photographie; Band I, das photographische objectiv (Julius Springer, Vienna, Austria, 1932).
[21] A. H. Bennett, Aberrations of long focus anastigmatic photographic objectives, BS Sci. Pap. **19**, 587 (1924) S494.
[22] I. C. Gardner and F. A. Case, The lateral chromatic aberration of apochromatic microscope systems, BS J. Research **6** (1931) RP316.

[23] Richard Glazebrook, A dictionary of applied physics, **IV**, p. 120 (Macmillan and Co., Ltd., London, 1923).
[24] Julius Rheinberg, Graticules, Transactions of the Optical Society **20**, No. 8, p. 17 (May 1919).
[25] Ralph D. Geiser, Reticle production; photographic, mechanical, and high-vacuum techniques, The Instrument Maker p. 4 (July-August 1948).
[26] Karl Leistner, Photographic methods for producing reticles, Photographic Engineering **1**, No. 1, p. 7 (January 1950).
[27] Ralph D. Geiser, Modern techniques of producing precision scales and reticles, Photographic Engineering **4**, No. 1, p. 1 (1953).
[28] A. Miethe, Fine grain photographic plates, Brit. J. Photography **59**, 707 (1912).
[29] F. H. Smith, The production of graticules–II, The Photographic Journal, Section B, **88B**, 18 (1948).
[30] F. Burmistrov, The photgraphic production of large precision scales and graticules, Technical Physics of the U. S. S. R. **5**, No. 6, 431 (1938).
[31] A. J. Bull and H. M. Cartwright, The production of graticules, The Photographic Journal, Section B, **87B**, 43 (1947).
[32] P. C. Smethurst, The cold enamel process, Process Engraver's Monthly **49**, No. 1, 5 (1942).
[33] Cold top enamel handbook (Mallinckrodt Chemical Works, St. Louis, Mo., 1948).
[34] Allan R. A. Beeber, David D. Jacobus and Carol W. Keuffel, U. S. Patents 2,447,836 (1948) and 2,559,389 (1951).
[35] L. P. Clerc, Photography theory and practice, p. 444 (Pitman Publishing Corp., New York, N. Y., 1937).
[36] Joseph S. Friedman, History of color photography, pp. 487–493 (The American Photographic Publishing Co., Boston, Mass., 1944).
[37] K. H. Sanders, The aromatic diazo compounds (Edward Arnold & Co., London, 1949).
[38] C. B. Neblette, Photography its principles and practice (D. Van Nostrand Co., Inc., New York, N. Y., 1942).
[39] John V. Sigford and Waldo H. Kliever, U. S. Patent 2,357,913 (1944).
40] R. S. Cox and C. D. Hallam, The production of graticules–III, The Photographic Journal, Section B, **88B**, 70 (1948).
[41] E. E. Loening, A process for making acid resists for graticules, Process Engraver's Monthly **57**, 266, 269–270, 297–298, 301 (1950).
[42] J. S. Mertle, Glass etching, Process Engraver's Monthly **49**, 231 (1942).
[43] William H. Wood, Light-sensitive coatings for lithography, The National Lithographer **45**, 30, 66 (1938).
[44] P. C. Smethurst, Light-sensitive plastics in photoengraving, Plastics **6**, 108–111 (1942).
[45] Alex Brooking Davis, U. S. Patent 1,751,908 (1930).
[46] Murray C. Beebe and Alexander Murray, U. S. Patent 1,587,271 (1926).
[47] Alexander Murray, U. S. Patent 2,091,715 (1937).
[48] The making of mirrors by the deposition of metal on glass, NBS Circular 389 (1931).
[49] Ernst Doelker, British Patent 183,817 (1923).
[50] Murray C. Beebe, Alexander Murray and Harold V. Herlinger, U. S. Patent 1,587,270 (1926).
[51] L. H. Baekeland, Ind. Eng. Chem. **1**, 149 (1909).
[52] R. L. Wakeman, The chemistry of commerical plastics, pp. 115–116 (Reinhold Publishing Corp., New York, N. Y., 1947).
[53] Carlton Ellis, The chemistry of synthetic resins **I**, pp. 295, 296, 342 (Reinhold Publishing Corp., New York, N. Y., 1935).

WASHINGTON, May 3, 1955.

Journal of Research of the National Bureau of Standards Vol. 55, No. 3, September 1955 Research Paper 2612

Development of a Photoresist for Etching Designs in Glass

Chester I. Pope and Raymond Davis

The preparation of a light-sensitive coating from phenol-formaldehyde resin is described. The light-sensitive coating may be used as a resist or mask in etching scales or reticule patterns in glass. A new etching solution was developed for use with this resist, and it is also recommended for fine line etching with wax resists.

1. Introduction

Etched glass scales and reticules are commonly produced by using a wax resist as a mask. A thin layer of wax coated on the glass is engraved or ruled in such a manner that the design or pattern is accurately positioned and the glass is uncovered in the engraved areas. The design is etched into the glass with hydrofluoric acid. After removing the resist, the etched areas are usually filled with an opaque pigment to increase the contrast and legibility. The workers in this field have a need for a light-sensitive coating, or resist, so that duplicate etchings may be made by contact printing, as in photography.

The problem of finding a suitable substance for making a light-sensitive resist is difficult because the glass must be etched with hydrofluoric acid, which attacks all of the known light-sensitive coatings. The coatings invariably crack, undercut, decompose, or become permeable to the etching solution before the etching is complete. Procedures using an undercoat, such as silver or a varnish, under the light-sensitive resist did not prove satisfactory.

Phenol-formaldehyde resin was known to be quite resistant to the action of hydrofluoric acid, and it seemed possible that a thin layer coated on glass might serve as a resist for hydrofluoric-acid etching after development of the image. Doelker [1][1] patented a process in which a phenol-formaldehyde resin, sensitized with ammonium dichromate, was recommended as an etching resist for zinc plates. Beebe, Murray, and Herlinger [2] patented the use of light-sensitive phenol-formaldehyde resin for photographic processes. They sensitized the resin with iodine, iodoform, etc., and, after exposure, developed the coating in a solution of about equal parts of alcohol and water. The resin made according to their directions was quite soft, remained tacky too long after coating for contact printing, and was low in light sensitivity. In the course of the present investigation it was found that the softness and tackiness of the resin was due in part to the presence of methyl alcohol in the formaldehyde solution used in the synthesis of the resin. The addition of methyl alcohol to trioxymethylene and phenol, in a concentration corresponding to that generally found in 40-percent formaldehyde solutions, yielded a soft, tacky resin. Therefore, trioxymethylene was used instead of the 40-percent aqueous solution of formaldehyde for the production of the phenol-formaldehyde resin in this work. Also the alcohol-water developer used by Beebe, Murray, and Herlinger [2] was not a satisfactory developer for the resin on glass. A resorcinol-formaldehyde resin was also made and was found useful as an additive in the phenol-formaldehyde resin.

[1] Figures in brackets indicate the literature references at the end of this paper.

2. Synthesis of the Phenol-formaldehyde Resin and the Resorcinol-formaldehyde Resin

2.1. Synthesis of the Phenol-formaldehyde (State–A) Resin

Baekeland [3, 4, 5] divided the phenol-formaldehyde reaction during resinification into three stages, A, B, and C, which are characterized by the appearance and solubility of the resin. The A-stage, or initial product, is a low molecular weight condensation product soluble in acetone; the B-stage, or intermediate condensation product, is swelled by acetone but not soluble in it; and the C-stage, or final condensation product, is infusible and completely insoluble in acetone. The resin used for the photoresist is that formed in the A-stage and is condensed to the state where it is just soluble in 95-percent ethyl alcohol. The resin is a thermosetting phenolic, which slowly passes into the B-stage on drying at room temperature.

Sodium acetate, a weak alkaline salt, proved to be the best catalyst for the condensation of the phenol and formaldehyde. It yielded a resin that had the greatest light sensitivity and which was readily developable by an organic solution. Other salts tested as catalysts were potassium acetate, lithium acetate, calcium acetate, barium acetate, and sodium formate. These salts produced resins that were low in light sensitivity and could not be as easily developed as the resin formed with sodium acetate. Acids such as hydrochloric acid act as strong catalysts and produce a resin that is not as light sensitive as that produced by the weak alkaline catalyst, sodium acetate, and the condensation

518-743 O - 73 - 55

reaction is difficult to control because the resin passes rapidly from the A-stage to the B-stage.

Investigation of the condensation reaction showed that the ratio of 1 mole of phenol, 3 moles of trioxymethylene, 6 moles of water, and 0.3 mole of sodium acetate ($CH_3COONa.3H_2O$), produced a satisfactory light-sensitive resin that was not too soft. Decreasing the sodium acetate gave a softer resin, which is less light sensitive.

The following formula and procedure were worked out for synthesizing the light-sensitive resin:

Phenol	282 g
Water	324 ml
Sodium acetate ($CH_3COONa.3H_2O$)	122 g
Trioxymethylene (paraformaldehyde)	270 g

The phenol and trioxymethylene were USP grade, and the sodium acetate was reagent quality, meeting ACS specifications. Distilled water was used in the synthesis and tap water in washing the resin.

The apparatus consisted of a 3,000-ml round-bottomed flask fitted with a reflux condenser by means of a ground-glass joint and mounted on a ring stand. The apparatus should be located in a chemical hood. The chemicals were added to the flask in the order listed in the formula followed by the addition of boiling ships made from an unglazed porcelain plate. The mixture was heated through an asbestos pad with a Bunsen burner to a gentle boil. The ring stand was rocked gently and frequently to prevent superheating until the boiling began. The boiling was continued throughout the period of the reaction, and the time required was 2½ to 3½ hr. The temperature of the boiling solution during the reaction and before precipitation was 105° C. Toward the end of the condensation reaction the solution became cloudy. It became opaque to transmitted light about 20 min before two definite phases developed, accompanied by foaming. After the two phases appeared, a gentle and frequent rocking of the ring stand was helpful in preventing the resin, which slowly precipitates from the boiling mixture, from sticking to the flask. After the two phases appeared, the boiling was continued about 5 min, and the hot reaction mixture was immediately poured into cool water while being rapidly stirred, thus completing the precipitation of the resin from the hot reaction mixture. The resin was washed and kneaded in about five changes of water and tested for degree of condensation. This test was made by taking about 10 g of the resin batch, heating it on the steam bath in a small porcelain dish while stirring with a glass rod, and noting the time for the resin to pass into the B-stage. This time was observed visually when the resin became rubbery, stringy, or lumpy and was the point at which all or most of the resin became insoluble in 95-percent ethyl alcohol. If the resin has condensed close enough to the B-stage, the small sample of resin should become insoluble in 95-percent ethyl alcohol after being heated for 10 min on the steam bath. If the time was much greater than 10 min, the precipitated resin batch was heated on the steam bath in a large porcelain dish and stirred constantly with a glass rod, and the time of the heating was estimated from the time required for the small sample to become insoluble in 95-percent ethyl alcohol. With experience, the operator will find no difficulty in determining the end-point of the condensation. As a safeguard against carrying the condensation of the precipitated resin too far on the steam bath, the resin batch may be cooled with water at intervals and a small sample tested for degree of condensation as described above. The resin passes rapidly from A-stage to B-stage at a temperature of 100° C., and if the resin shows any signs of stringiness, or a tendency to be rubbery and lumpy, it should be cooled immediately by stirring in cool water. The resin is usable even if a few insoluble particles appear in the alcohol solution, but it should be discarded if essentially insoluble in alcohol. The highest degree of light sensitivity and developability were obtained by condensing the resin as close to the B-stage as possible, without the resin becoming insoluble in 95-percent ethyl alcohol.

After the resin had been condensed to the desired state and cooled, as much of the water as possible was pressed out and the resin dissolved immediately in 700 ml of methyl ethyl ketone (2-butanone) of a good commercial grade, free from acid or any substance that would react with iodoform. Then 500 g of anhydrous sodium sulfate was added and the resin solution allowed to stand for several days, with occasional shaking. After filtering the resin solution it was ready for sensitizing and use. The amount of resin, determined from the water-free solutions of different batches, gave yields of 220 to 300 g. The resin solutions contained from 22.5 to 30 g of resin per 100 ml of solution. The concentration of the resin in the filtered solution was determined by noting the volume, evaporating 10 ml of the solution to dryness in a tared 50-ml beaker, and weighing.

For use as a light-sensitive resist coating, the concentration of the resin was adjusted to 20 g of resin per 100 ml of solution and sensitized by adding 1 g of iodoform for each 10 g of resin. The iodoform used was Eastman Kodak Cat. No. P–341. Just before use, the sensitized resin solution was filtered through a coarse fritted-disk filtering funnel without pressure. Sensitized resin solutions have been usable after storage for 2 years in a refrigerator. Unsensitized resin solutions stored in the dark at room temperature were still usable after 3 years. The phenol-formaldehyde resin required about 2 to 3 times more exposure to actinic light than the well-known cold top enamel to render the coating insoluble in the developer solution.

A second formula containing ethylene glycol as part of the solvent was developed for the synthesis of the resin. The resin made by this last formula was generally sufficiently condensed when precipitated so that no further treatment was necessary on the steam bath. However, it should be tested for degree of condensation, as described above, before dissolving it in the methyl ethyl ketone. The resin has the same sensitivity as that made by the first formula but contains a trace of ethylene glycol.

Ethylene glycol, because of its hydroscopic nature, caused a slight decrease in the light sensitivity of the resin coating at high relative humidities. The effect was nil at a relative humidity of 50 percent or less.

The following alternative formula was worked out for synthesizing the phenol-formaldehyde resin:

Phenol	282 g
Ethylene glycol	186 g
Water	270 ml
Sodium acetate ($CH_3COONa \cdot 3H_2O$)	122 g
Trioxymethylene (paraformaldehyde)	270 g

The ethylene glycol was of reagent grade. The apparatus and procedure for producing the resin were the same as described for the previous formula. The time required for the reaction was 2 to 3 hr. The temperature of the boiling solution during the reaction and before precipitation was 107.5° C. The reaction mixture remained clear until very close to the end of the reaction. The solution became cloudy, and the boiling was continued about 3 min after it had become opaque to transmitted light. The resin was immediately precipitated from the hot reaction mixture by pouring into cool water with rapid stirring. The resin was treated in the same manner as that obtained from the first formula.

2.2. Synthesis of the Resorcinol-formaldehyde Resin

The resorcinol-formaldehyde resin, because of its reactivity and chemical similarity to the phenol-formaldehyde resin, should improve the sensitivity and resist properties of the phenol-formaldehyde resin, and this was found to be true. A resin, soluble in 95-percent ethyl alcohol, was produced from the condensation of resorcinol and formaldehyde. Although this resin was light sensitive, it was not possible to develop the image satisfactorily. It is somewhat more resistant to hydrofluoric acid than the phenol-formaldehyde resin. The addition of 5-percent of the resin, based on the weight of the phenol-formaldehyde resin, increased the sensitivity about 25 percent, increased the intensity of the dyeing of the image during development, and improved its resist properties slightly.

The following formula and procedure were used in synthesizing the alcohol-soluble resorcinol-formaldehyde resin:

Resorcinol	55 g
Water	75 ml
Sodium sulfate, anhydrous	12.5 g
Formaldehyde solution, 40%	41 ml

The resorcinol was Eastman Kodak Cat. No. P–222, and the sodium sulfate was reagent quality, meeting ACS specifications. The formaldehyde was USP grade. Distilled water was used both in the synthesis and in washing the precipitate. The chemicals were added to a 250-ml Erlenmeyer flask and brought into solution. The reaction mixture was stored without agitation in a constant-temperature room at 22 $\pm 1°$ C. The solution was clear after 24 hr, but after 40 hr a precipitate began to settle out. At the end of 140 hr the reaction mixture was diluted to 1,500 ml, and then 100 g of anhydrous sodium sulfate was added. The sodium sulfate salts out the resin during the reaction and also on dilution before filtering. The precipitated resin was removed with suction by means of a Buchner filtering funnel. The precipitate was thoroughly washed, air-dried overnight, and finally dried in the dark in a desiccator over anhydrous calcium chloride. When the reaction time was extended much beyond 140 hr the resin became insoluble in 95-percent ethyl alcohol.

The rate of the condensation reaction increased with increase in temperature, accompanied by an increase in the yield of an alcohol-insoluble resin. At 100° C the reaction was very rapid, and the resin was completely insoluble in ethyl alcohol. It was difficult to obtain a good yield of the alcohol-soluble resorcinol-formaldehyde resin when the reaction was carried out at a temperature much above 25° C.

The dried resin, when stored in a dry brown bottle in the dark at room temperature, was still soluble in 95-percent ethyl alcohol after 2 years. The resin, when dissolved in methyl ethyl ketone over anhydrous sodium sulfate at a concentration of about 20 g per 100 ml of solution, was still usable after 3 years of storage at room temperature.

3. Development of the Exposed Resin Coating

After suitable exposure in contact with a master positive, the unexposed areas of the resin coating were dissolved out (developed), and the exposed areas were dyed simultaneously by means of an organic solution developed for use on this resin coating. The developer solution was made up as follows:

Ethyl alcohol, 95%	1,000 ml
Normal butyl alcohol, bp 117 to 117.8° C	1,000 ml
Methyl ethyl ketone, bp 79 to 80.5° C	200 ml
Methyl violet	2 g
Basic fuchsin (aniline red, magenta)	2 g

Before use the solution was filtered through a coarse frittered-disk filtering funnel.

The ethyl alcohol, normal butyl alcohol, and methyl ethyl ketone were of reagent grade and were used without distillation or drying. The developer solution contained about 3 percent of water. The mixture of methyl violet and basic fuchsin dyed the resin coating a heavy blue-purple and was found to be the best combination of dyes tested. Methyl violet dyed the resin coating blue-green; basic fuchsin, red; and malachite green, green. Another developer solution was also made up without the dye for use in removing the excess dye that collected on the surface during the developing procedure. The exposed resin coating was developed by immersing the coated specimen in the developer-dye solution for 30 to 40 sec with gentle agitation and rinsing it 2 or 3 times in developer solutions that contained no dye. Although the development took place in 30 to 40 sec at room temperature, there were no adverse effects on the resin coating if the developing time was extended for as long as 2 min.

4. Etching Solution

After developing and drying, the resistance of the resin coating to hydrofluoric acid was increased by baking it in an oven between 100 to 135° C for 1 to 3 hr. Concentrated hydrofluoric acid (48%) attacked the phenol-formaldehyde resin coating. The problem was to find a solvent for the hydrofluoric acid that would delay its attack on the resin coating and prevent the formation of insoluble salts during the etching. Experimentally the answer was found in using hydrofluoric acid in aqueous solution with phosphoric acid. Also, phosphoric acid assisted the hydrofluoric acid in etching the glass [6]. The phosphoric acid retarded the attack of the hydrofluoric acid on the developed resin and gave a clear etch when sufficient water was added. The following formula was developed:

Ortho phosphoric acid, 85%	100 ml
Water	50 ml
Hydrofluoric acid, 48%	25 ml

The phosphoric and hydrofluoric acids were of reagent quality, meeting ACS specifications. The etching solution was easy to make up as no noticeable heating took place when the chemicals were mixed. The etching solution may be used as soon as it is made up, and, if stored in a wax pot fitted with a tight cover, it will keep for several months. If the developed phenol-formaldehyde resin coating is not too thin (not less than 0.00004 in. thick), and has been baked, it will withstand the etching solution for periods up to 15 sec at room temperature. A 5-sec etch is recommended for fine line work where the etch is to be filled chemically. A deep etch may be obtained by etching several times for 3 to 4 sec, washing and drying at 100° C for about 30 min between each etch. At the end of the etch, the glass specimen was immersed immediately in a 20-percent solution of sodium hydroxide solution to neutralize the hydrofluoric acid and to remove the resin coating.

5. Conclusion

Many organic substances are sensitive to or hardened by actinic light, but it is generally a difficult problem to find a developer solution that will dissolve out the unexposed areas of the resin and not attack the exposed areas. The developer solution dissolved the unexposed light-sensitive resin quickly and yet did not attack the exposed resin, even if the development was extended several times beyond the normal developing time. Iodoform was found to be a very good sensitizer. The addition of 5 percent of the resorcinol-formaldehyde resin, based on the weight of the phenol-formaldehyde resin, improved the sensitivity and resist properties of the phenol-formaldehyde resin, although its presence was not necessary to make the phenol-formaldehyde resin function as a resist for etching glass.

The light-sensitive resist made with the phenol-formaldehyde resin will not withstand the action of concentrated hydrofluoric acid (48%). However, the etching solution containing hydrofluoric acid and phosphoric acid may be used for etching times up to 15 sec. An etching time of 5 sec was used to produce fine lines (10 microns wide). Even though the light-sensitive resist will not withstand a long etching time, deep lines may be obtained with the hydrofluoric acid-phosphoric acid etching solution by etching the glass several times for 3 or 4 sec, washing, and drying between each etch.

For the details on the use of the phenol-formaldehyde light-sensitive resist, see National Bureau of Standards Circular 565 [7] Techniques for Ruling and Etching Precise Scales in Glass and Their Reproduction by Photoetching with a New Light-sensitive Resist.

6. References

[1] E. Doelker, British Patent 183,817 (1923).

[2] M. C. Becbe, A. Murray, and H. V. Herlinger, U. S. Patent 1,587,270 (1926).

[3] L. H. Baekeland, Ind. Eng. Chem. **1**, 149 (1909).

[4] R. L. Wakeman, The chemistry of commercial plastics, p. 115 (Reinhold Publishing Corp., New York, N. Y., 1947).

[5] C. Ellis, The chemistry of synthetic resins, vol. 1, pp. 295, 342 (Reinhold Publishing Corp., New York, N. Y., 1935).

[6] B. Schweig, Principles and methods of etching and embossing, Glass, 143 (April 1938).

[7] Raymond Davis and Chester I. Pope, NBS Circular 565 (in press).

Washington, April 5, 1955.

PHOTOGRAPHIC IMAGE STRUCTURE EVALUATION

C. S. McCamy
National Bureau of Standards
Washington, D. C. 20234

Introduction

The first optical instruments were aids to vision and photography has mainly been used to make pictures for the benefit of human observers. Quite naturally, the quality of such instruments and processes has been judged by their ability to satisfy the needs of human vision. The testing and evaluation of such systems required methods which were partly physical and partly psychological. We call them "psychophysical" methods.

When lenses were to be used to expose photographic plates rather than as aids to direct vision, new lens designs were required. The lenses had to meet new criteria of quality. In a similar manner, photographic materials and processes designed for the production of portraits are not those we use for the processes of photochemical fabrication. The end product in the latter case is not just supposed to look good, it is supposed to work. The criteria of quality for pictorial processes are not necessarily the proper criteria of quality for fabrication processes, but full advantage should be taken of the lessons of pictorial image quality analysis and the concepts and techniques should be generalized if possible, to serve the needs of fabrication. It is important to understand clearly the psychophysical basis of several of the measures of image quality so that we may avoid the hazards of applying them directly and that we may benefit from the appropriate physical generalizations which the analysis of the psychophysics may suggest. Today, photographic image structure evaluation must be thought of as encompassing these generalizations beyond pictorial evaluation. We must hasten to admit that this branch of science has yet to be adequately explored, but we may here chart some part of the course.

Optical Density

By virtue of the general tendency of the human senses to equate logarithmic increments of stimuli, equal steps of optical density appear equally different in lightness, but there is a direct physical implication of optical density--which is where it gets its name.[1] The transmission density of a layer of partly absorbing material is proportional to the concentration of the material in terms of mass per unit of area. In the case of homogeneous materials, it is proportional to thickness. Thus transmission density is not only well suited as a measure of how the layer looks, but as a measure of its physical nature. One must bear in mind that absorption is not the only modulating process in optics and that refraction, reflection, and diffraction effects can greatly alter appearances and influence measured optical densities, particularly in the microscopic world. Transmission density is the negative logarithm to base 10 of the transmittance factor. Similarly, reflection density is the negative logarithm to base 10 of the

[1] F. Hurter and V. C. Driffield, J. Soc. Chem. Ind., 9, 445 (1890).

reflectance factor. The measurement of optical densities on very small areas, known as microdensitometry, is the basis of almost all photographic image structure evaluation. To make such measurements, one uses a microdensitometer. The sample is placed on a stage, where it is illuminated by means of microscope optics operated in reverse. The transmitted or reflected flux is collected by a second microscope system and the collected flux is directed to a photosensitive receiver. The stage is usually moved relative to the photometric system and the beam is confined by appropriate slits so that the sample is scanned. The resulting electrical signal corresponding to optical density changes is usually automatically recorded on a strip chart or a suitable medium for input to a digital computer. A computer is nearly indispensible if image structure evaluation is to be done routinely.

The measured value of optical density depends on the geometric distribution of the flux incident on the sample, the geometric distribution of the sensitivity of the optical system which receives flux from the sample, the spectral distribution of the incident flux, and the spectral sensitivity of the receiver system. Terminology and notation for describing such measuring systems has been published and is being considered for standardization.[2] The proposed notation for the case of microdensitometry is illustrated by this example for transmission density:

$$D_T\ (0.25;\ 2850K\ :\ 0.40;\ S\text{-}4)_n$$

where D_T means transmission density, 0.25 is the numerical aperture of the influx optics, 2850K is the temperature of the Planckian distribution representing the influx spectrum, 0.40 is the numerical aperture of the receiver optics, S-4 is the standard designation of the photocell sensitivity which describes the whole receiver in this case, and the subscript n means that the geometric conditions are described by numerical apertures.

Ideally, all of these conditions should be matched when a density standard is calibrated for the purpose of calibrating the working microdensitometer. The National Bureau of Standards has a microdensitometer which may be adjusted to a wide variety of conditions and is, itself, calibrated by application of the inverse-square-law of illumination. The working standard should be a step tablet made of the same kind of material as is to be measured.

In many instances in image evaluation, the absolute density need not be known. In such cases, a step tablet calibrated for diffuse transmission density on a "macro" densitometer is commonly used.

As a general rule, if a measurement of optical density is to be a significant measurement, the measuring conditions must simulate the conditions of the application. Most of the problems in measuring optical density are traceable to a lack of this simulation.

Optical Limitations

In all photographic work involving fine detail or sharp edges we must be mindful of the limitations imposed by the nature of light. It is sufficient to note that the maximum linear resolving power, in cycles per millimeter, is about 2000 divided by the f-number of the lens or about 4000 times its numerical aperture. Thus an f/1 lens cannot resolve over 2000

---

[2] C. S. McCamy, Photo. Sci. Eng., 10, 314 (1966).

cycles per millimeter and an f/64 lens can't be expected to resolve over 31 cycles per millimeter. Bear in mind, these are limitations on what one could see through ideal optical instruments if the original object were well-formed high contrast lines of adequate length, but are not useful limits for practical optical instruments, much less for fabrication processes. We do often observe more than 80% of this limit at the center of the field of fine modern camera objectives and we do observe a continual increase in that visual resolving power as the apertures of such objectives are opened from smallest to largest aperture. At the widest aperture settings, there is a considerable reduction in contrast at the lower frequencies, which in many applications are the frequencies of interest. This illustrates the important point that the maximum visual resolving power of a system may not be a useful measure of performance. The resolution rules-of-thumb are most useful in avoiding attempts to do the impossible.

The failure of lenses to resolve beyond the limits given results from the fact that in image formation, even with ideal lenses, light spreads beyond the geometrical boundaries of the ideal images. Thus a point source object is imaged as a bright spot surrounded by a series of faint rings, growing ever fainter as their diameters increase. The image of a bright line is made up of the sum of all such point images lying in a line. The image of the sharp edge between a light and a dark field reveals a gradual reduction of illuminance, from the light to dark parts of the image. This spreading of points, lines, and edges persists whether or not the limiting optical resolving power is of any interest. This fact is of more significance for photochemical fabrication systems than the fact that resolution is limited. The image spread in real optical systems depends not only on the influence of the nature of light but on the peculiar aberrations of the particular system. As a rule-of-thumb, the illuminance will always be a few percent of the maximum value at a distance (measured in micrometers) from the geometric edge equal to half the f-number or one fourth the reciprocal of the numerical aperture and the illuminance rises sharply from that point to the maximum value at the geometric edge.

Photographic Resolving Power

Early photographers naturally adopted the methods of measuring resolving power that had come into use in the evaluation of optical instruments. The usual test object for many years has been a field of alternately light and dark bars of equal width. The customary practice in the testing of photographic materials today is to use an optical system capable of resolving 1600 cycles/mm or more on one kind of emulsion, project a specified bar pattern onto the test emulsion under controlled conditions, make a series of different exposures (varying illuminance or time), make a series of focal settings, process the material in a specified manner, and examine the photographic image visually under specified conditions of microscopy to determine the highest spatial frequency for which the bars can be counted with reasonable assurance. The requirement that the bars be countable accords with the initial concept of optical resolving power, the ability to detect the number of stars in a cluster that might look like one star to the unaided eye.

The measured optical or photographic resolving power depends on the wavelength of the flux employed, the contrast between bars and spaces, ratio of bar length to width, ratio of bar width to space width, the number of bars in the pattern, and the extent of the surrounding unpatterned field. In some systems the value depends also on the orientation of the bars with respect to the remainder of the system and, in the photographic case, it

depends on exposure, processing, and optical image motion with respect to the emulsion.

In measuring resolving power, one should beware of spurious resolution. If an optical system is slightly out of focus it can produce a fairly distinct image of parallel bars of a certain spatial frequency even though patterns of lower frequency are not resolved. In such cases the apparent number of bars in the image is not the number in the real object. In judging resolution, in addition to the requirement of countability, it is well to require an actual count as a safeguard against this type of resolution, known as spurious resolution.

From time to time, pattern forms other than parallel bars have been proposed. In some instances the patterns were designed to test the ability of the system to produce discrete images and these were properly regarded as resolution targets suitable for resolution measurements. However, some patterns have been designed to test other characteristics, such as the ability to just indicate the direction of a set of parallel lines. Since the fragmentary delineation of a single one of a set of bars suffices to indicate direction, this can hardly be considered a true criterion of resolution. Needless to say, such a criterion would utterly fail to cope with the problem of spurious resolution, where pattern direction may be clearly evident when the system is intolerably out of focus.

Long after the concept of resolving power was quite well established, it became apparent that, by means of interference phenomena, one could expose an emulsion in such a way that numerous layers or lines of optical density were produced on development. Though the layers or lines might be too finely spaced or indistinct to be counted or even observed individually, the periodic nature of the density could be detected by the diffraction or interference effects produced. The lines or layers were said to be resolved and, in a peculiar sense, they were. It is useful to differentiate this kind of resolving power from the kind originally defined. Discrete resolving power is the ability of a system to produce the individual bar images with sufficient distinction to permit them to be counted with reasonable assurance. Statistical resolving power is the ability of a system to produce an image pattern with a degree of periodicity which is detectable only by the utilization of the combined optical effect of a number of layers or bar images. Although one might make some measurements and compute the approximate number of bars in a statistically resolved pattern, the bars are not directly countable. It may be noted that large numbers of discretely resolved layers or bars can interact very well to produce interference effects, but, by definition, statistically resolved bars are not discretely resolved. Incidentally, spurious resolution must be regarded as statistical rather than discrete, because the optical image of "one bar" is a confusion of light from two or more bars in the object.

Discrete resolving power is of two kinds, visual resolving power which is determined by visual observation with or without optical magnification and instrumental resolving power which involves the determination of discreteness by instrumental means other than vision. Instrumental resolving power is purely physical and objective while visual resolving power is psychophysical and subjective, as it should be for its intended purpose. The simplest instance of instrumental resolving power would be the use of a microdensitometer to determine whether or not the optical density of an image varies periodically and with measurable amplitude. In photochemical fabrication of electronic components, one often places a number of conductors or other components close together. The objective is to produce components

which act independently to some desired degree in an electronic sense rather than in a visual or other optical sense. The criterion of resolution should be an electronic one. This is a legitimate and useful generalization of the concept of resolving power. It would be one case of what I would call nonvisual resolving power. The great preponderance of visual applications in photography suggests the subdivision into visual and nonvisual even though, admittedly, the visual case is but one of many possible cases. We might call the present case electronic resolving power, a subclass of nonvisual resolving power.

Performance and Indicative Tests

It is useful to distinguish between performance tests and indicative tests. A performance test is a trial of a material or process under actual or closely simulated conditions of use. An indicative test is a test which one has reason to believe, on theoretical or empirical grounds, will correlate well with a performance test but has some advantage over the performance test, such as convenience, rapidity, or low cost. If one wants to be able to count parallel lines on a photograph, such as pipelines or logs on an aerial photograph, a visual resolving power test is a simulation so close as to be regarded as a performance test. If, as in document microphotography, one desires legibility of greatly reduced print, a visual resolution test is an indicative test having a basis in theory and experience. The resolution test is much simpler, much quicker, and requires fewer observers than a performance test of legibility by measuring literal response, the percentage of correctly identified letters in garbled test printed in a range of sizes[3]. Thus a performance test in one case may be an indicative text in another.

It is not unlikely that desirable end results may be correlated with visual resolving power of materials or processes at intermediate or final stages in production. If visual inspection is found useful in this way, it should be fully utilized as an indicative test, even though the quality of the final product may not be directly related to appearances. Similarly, one might measure instrumental resolving power and correlate it with electronic resolving power.

Since inductive and capacitive effects are known functions of the distance between conductors, it should be possible to compute the expected electronic interactions between nearby components and attribute to the fabrication process only the interactions which exceed those otherwise expected. The ultimate interest is in the total effect, of course, so it may not be necessary to differentiate effects, but doing so should provide clearer appraisal of the effects attributable to the fabrication process and help avoid efforts to fabricate impossible systems or to improve the fabrication of systems which could not be materially improved even if the fabrication were perfect. One is reminded of the futility of further polishing a diffraction-limited lens and the value of knowing when that quality has been achieved.

Sharpness and Acutance

It is well known that pictures differ in sharpness, but the establishment of a quantitative description of this phenomenon required considerable psychophysical investigation.[4] It was found that a large number of observers

[3] C. S. McCamy, Appl. Opt. 4, 405 (1965).

[4] G. C. Higgins and L. A. Jones, J. Soc. Mot. Pict. Telev. Engrs., 58, 277 (1952).

could sort pictures in piles according to sharpness and do so with reasonable consistency. This established the existence of a subjective impression which could be evaluated on a scale, but such psychological testing is laborious and requires elaborate control. The edges of images on the graded pictures were scanned with a microdensitometer and the edge traces were analyzed to determine what characteristic of the edges was associated with the appearance of sharpness. A purely objective quantity called acutance, derived from the edge trace by computation, was found to correlate well with the subjective impression of sharpness, for the given conditions of testing. This finding very much simplified laboratory evaluation of the sharpness of pictorial materials, providing a rather direct indicative test. Acutance is not necessarily associated with photography. One can measure the acutance of edges on a printed page, an oil painting, or a butterfly's wing. But what significance might such measurements have in the technology of photochemical fabrication?

The psychophysical approach is not appropriate except to the extent that it applies to visual inspection of parts. There is, of course, the possibility that something in the nature of the fabrication process or the end use happens to respond favorably to the same edge contour characteristic as the human eye. In the interest of unifying science and simplifying communication, it is desirable to try acutance and, if it is a useful indicative test, use it.

If not, the search for a quantity that does correlate with some measure of desirable performance of the general nature of sharpness may be warranted. In our laboratory, such a quantity has been called generalized acutance. Perhaps nonvisual acutance would be more exact. This is not to imply that we have a great deal of experience in applications of nonvisual acutance, but that it has come up enough to be recognized as a useful concept. For the final electronic product, the proper measures of desirable performance are electronic characteristics and it is not obvious that any such characteristics correspond to sharpness, but for masters and intermediates, the measure of performance is the ability to produce successive generations of high quality, a characteristic which might be called nonvisual sharpness. Since this is largely an optical characteristic, it should not differ in kind very greatly from visual sharpness.

Graininess and Granularity

However uniformly exposed and processed photographic emulsions may be, they usually exhibit a nonuniformity of density when adequately enlarged. The subjective impression of this nonuniformity is called graininess. The objective correlate, called granularity, is related to the standard deviation of the measured optical density when uniformly exposed and processed film is scanned in a microdensitometer with a circular scanning aperture of specified size, usually 20 to 30 μm in diameter. The measure of granularity proposed by Selwyn is generally used. Selwyn granularity is the product of the standard deviation of density and the square root of twice the area of the scanning aperture [5]. Granularity limits the ability of photographic materials to image fine detail.

Information Capacity

Just as noise in electronic communication channels limits the information capacity of the channel measured in bits per second, so granularity

5 C. E. K. Mees and T. H. James, "The Theory of the Photographic Process", Macmillan, N. Y., 1966. Chapter 23 of this book is the best available review of image structure and its evaluation, including 326 references.

limits the information capacity of photographs, measured in bits per square millimeter. For high-contrast visual applications the information capacity is about equal to the square of the visual resolving power. [3,6] The main aspects of the derivation of that relationship should hold equally well for nonvisual resolving power in nonvisual applications.

Image Spread in Photographic Materials

Photographic materials scatter light as it passes through them. This characteristic is known as turbidity. However well an optical imaging system may focus light to a point, line, or edge, the photographic image resulting from exposing the material to such an optical image exhibits a spreading of the point to a circle, a broadening of lines, and a decrease in the density gradient at edges. The study of this phenomenon is made difficult and imprecise by the difficulty of locating the extremities of the image for measurement purposes. The apparent position of the edge depends on the edge gradient function, total density difference, retinal illuminance, and the state of adaptation of the observers eye [7]. When a microdensitometer is used, a fractional density point (1/3, 1/2, or 2/3 the density range) is often taken as the location of the edge.

The diameter of the image of a point is approximately a linear function of the logarithm of exposure.

When lines are being imaged, very thin lines may not print as dense as wider lines. Negative lines, that is spaces between exposed areas, are exposed to scattered light and may have greater density as the space width decreases. Both of these effects can be described for high-contrast high-resolution materials by a linear relationship between optical density and the logarithm of line width, for the widths between that which just makes an image and that which gives the same density as a wide line (say 50 μm) of the same exposure. The slopes of these two lines for a given material are opposite and the lines cross.

For line widths less than that at which they cross, the densities of spaces exceed those for lines. One must bear in mind the ever-present optical limitation of line width, the minimum width in micrometers being about equal to the f-number of the exposing lens.

Spatial Effects in Photographic Processing

The optical density is a function of exposure and the relationship of density to the logarithm of exposure over broad areas is the well-known characteristic curve. As was just noted, in the consideration of fine lines, the effective exposure depends somewhat on the geometric distribution of the optical image. The chemistry of conventional developing, fixing, and washing are also affected by the optical image structure and, as a result, the optical image is not truly reproduced in the photographic image.

As a general rule, photographic grains respond independently, each becoming fully developable if it receives a given minimum exposure. However, in some emulsions, the grains tend to clump together and act as single grains. As a general rule, chemical reaction products of the development process tend to retard development but, in a few developers, they actually accelerate it. These reaction products affect not only the development of

[6] J. Riesenfeld, Phot. Sci. Eng., 11, 415 (1967).

[7] W. N. Charman, Opt. Acta, 10, 129 (1963).

exposed grains but also the nature of the gelatin and its subsequent drying.

During development of the edge of an exposed area, reaction products diffuse into the less exposed area, retarding development and producing less density just outside the geometric edge than is produced by the same exposure farther away. Conversely, fresh developer diffusing into the exposed area increases the density just inside the edge. These edge effects tend to sharpen the image, compensating somewhat for the spread of images, even overcompensating in some cases. For example, lines of the order of 100 micrometers in width are often found to have a higher density than much wider lines of the same exposure, an effect known as the Ebenhard effect. Retarded development between two exposed points or lines, tending to make their images farther apart, is known as the Kostinsky effect. The gelatin effect causes shifts of the image position during drying. We refer to systematic changes in the relative positions of fine details due to the gelatin effect as the Ross effect. There are random as well as systematic gelatin effects and some similar effects can be attributed to the base on which the emulsion is coated. Cellulose acetate produces rather large effects compared to glass and the polyester bases (polyethylene terephthalate) are intermediate.

Only a few of the more important such effects have been cited here. Microdensitometry reveals these effects and it is of paramount importance to anyone measuring and analyzing image structure and utilizing such analysis in the control of complex systems of fabrication to know these effects and their basis in physics and chemistry.

Modulation Transfer Function

The electrical characteristics of circuits often limit their usefulness to a certain range of frequencies. The useful output for a constant input, as a function of frequency, is the frequency response. The exactly corresponding measurement in optics and photography is the spectral response. There is, however, an analogous concept which is used more nearly in the same way as the frequency response in treating photographic systems as communication channels. It is the spatial response, the ratio of the modulation contrast in the output to that in the input, as a function of spatial frequency; spatial frequency being the reciprocal of the period of a sinusoidal variation in luminance of the input object or transmittance factor of the photograph. The spatial response of a lens alone is called the optical transfer function. It is complex, the modulus being called the modulation transfer function and the phase component being called the phase transfer function. The scattering in the emulsion of a photographic material can be evaluated by measuring the modulation transfer function of the emulsion, using the emulsion itself as the photometric means to determine the effective exposure by assuming that the characteristic curve is constant at all frequencies. The spatial response of these systems indicates not only the range of useful spatial frequencies, which is the resolving power, but indicates how the modulation contrast varies at all frequencies within this range. Problems arise in the application of transfer functions because nonlinearities make optical transfer functions dependent on the state of coherence of the object and the emulsion, as we have noted, is hardly a "well behaved" linear component.

The National Bureau of Standards provides patterns which are sinusoidal in transmittance factor, with a range of spatial frequencies. When these patterns are diffusely transilluminated, they provide a range of sinusoidal luminance distributions for optical and photographic testing.

Statistical Optical Approach

Statistical optics, the rigorous treatment of optical systems, including the quantitative treatment of the state of coherence at each stage, has developed over the past fifteen years. Mathematically, it is at least as involved as rather complex circuit theory and is, as one might expect, not as yet widely known and practiced in engineering. Statistical optics is the wave of the future, and it is crashing upon our shores. Laboratories are beginning to measure the quantities requisite to rigorous analysis of systems and to apply them to optical production.

Standardization

Standard methods of measuring some of the quantities discussed here are described in documentary standards issued by the United States of America Standard Institute, 10 East 40th Street, New York, N. Y. 10016. The existing standardson transmission and reflection density are being revised. A standard method of measuring the resolving powers of photographic materials has been in preparation for some time and should be published within a year. Microdensitometry, the modulation transfer functions of lenses, and the spatial response of emulsions are being acted on by committees.

The National Bureau of Standards provides some physical standards, including two types of paper resolution charts, the sinusoidal patterns mentioned earlier, and photographic sharp edges for edge exposures. Several firms supply special test charts, either ready-made or to special specification. Some of the test films used in the motion picture industry may be useful. These are obtained from the Society of Motion Picture and Television Engineers, 9 East 41st Street, New York, N. Y. 10016.

Conclusions

Measurements must be related to applications and the nature of the relationship should be well understood. Many characteristics of optical and photographic systems contribute to image quality. Just as the pathologist must dread the infected tissues and cultures in his laboratory, so you must beware of mononumerosis, the unhealthy insistence on characterizing the quality of a complex system by a single number.

Professional Biography

C. S. McCamy earned his B. Ch. E. in 1945 and his M.S. in 1950 at the University of Minnesota where he was an instructor in Mathematics and Mechanics from 1947 to 1950. After a teaching post at Clemson College in physics from 1950 to 1952, he joined the National Bureau of Standards as a physicist in photographic research where he is currently engaged. He is a member of the National Research Council, the Royal Photographic Society of Optics, a Fellow of the Optical Society of America, and a Senior Member of the S.P.S.E.

THE NBS MICROCOPY RESOLUTION TEST CHART

Bernard H. Fouquet*

ABSTRACT

The National Bureau of Standards supplies resolution charts for measuring the resolving power of microfilming systems. This chart is made up of 21 patterns ranging in spatial frequency from 1.0 to 10 lines per millimeter. Recently, the master negatives used in the production of the charts were completely remade. This presented the opportunity to make certain desirable changes in the interest of standardization and efficiency of production, but the principal design features have been retained.

Measurements for the negatives and the resulting charts indicate that an accuracy of two percent is obtained in pattern line width and frequency.

The use of the chart is reviewed and the instructions issued with the charts are appended.

Introduction

The National Bureau of Standards Microcopy Resolution Test Chart was first produced in the early 1940's. It was designed to aid in determining the performance of microfilming systems. In such systems, the performance of the optics, the characteristics of the photographic material, and the effect of motion and vibration combine to limit the quality of the microfilm image. The chart is intended to provide a convenient method of measuring the system resolution under the conditions encountered during microfilming. The chart is placed on the camera copy board and photographed. The resulting image is then examined under a microscope to determine the resolving power.

The chart is produced photographically by contact printing. Recently the master negatives were completely remade and it appeared desirable to make some minor changes in the design of the chart at that time. The design changes were directed toward increasing production efficiency, improving the accuracy of the chart,

*National Bureau of Standards, Washington, D.C.

conforming to standardized preferred values of spatial frequency, and improving the appearance of the chart. The old chart is shown in figure 1 and the new chart in figure 2.

Resolving Power

The image quality produced by a microfilming system is usually evaluated in terms of resolving power. This measurement indicates the capability of the system to reproduce fine detail and is expressed in lines (or cycles) per millimeter. (In photography, a "line" is taken to mean a dark line and the associated light line, whereas in television practice these are considered two lines.) If the mechanical camera factors are controlled adequately, the following equation[1] approximates the relation between the separate resolving powers of the lens and film and the resolving power of the lens-film combination:

$$1/r_{f+l} = 1/r_f + 1/r_l \ .$$

For example, a well corrected lens with a resolving power of 300 lines per millimeter and a microfilm resolving 500 lines per millimeter were found to resolve 180 lines per millimeter when used in combination, a value in close agreement with the predicted value of 187 lines per millimeter.

After the chart is photographed, the resolving power is determined by visual examination. Variability of interpretation may lead to inconsistencies in the results obtained. To minimize such effects and assure consistent results, the resolution criteria must be clearly defined, clearly understood by the observers, and consistently applied. The instructions for the use of the National Bureau of Standards Microcopy Resolution Test Chart provide the definition of resolution criteria for use with this chart (See Appendix I). The copy of the chart is examined with the aid of a microscope, with a magnification from 1/3 to 1 times the resolving power (in cycles per millimeter) to be observed. For example, to view 100 lines per millimeter, the magnification should be between 30 and 100 times. The numerical aperture of the microscope objective should not be less than 1/1000 of the resolving power to be measured. In this way the observer is reasonably sure that the resolution being measured is that of the microcopying system and not that of the microscope. The use of proper magnification is necessary if the visual system is to be utilized effectively to perform this task.

---

[1] A. H. Katz, J. Opt. Soc. Am. 38, 604 (1948).

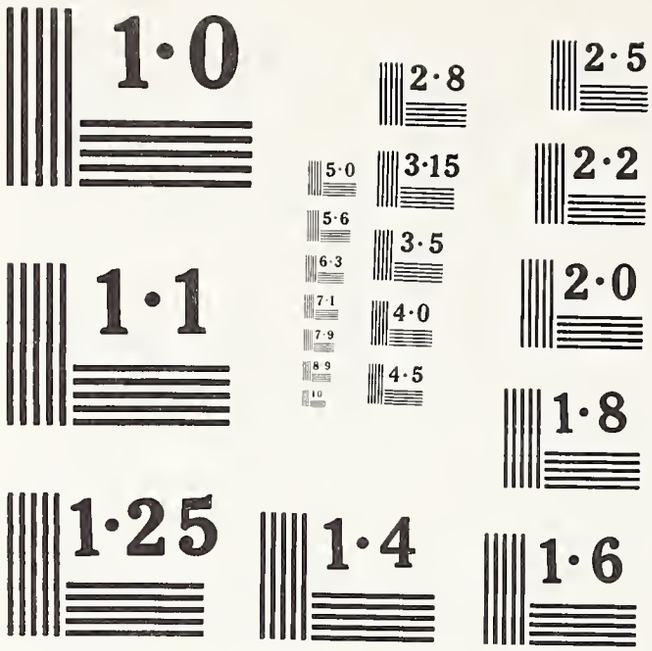


Figure 1. National Bureau of Standards
Microcopy Resolution Test Chart lo 1940.

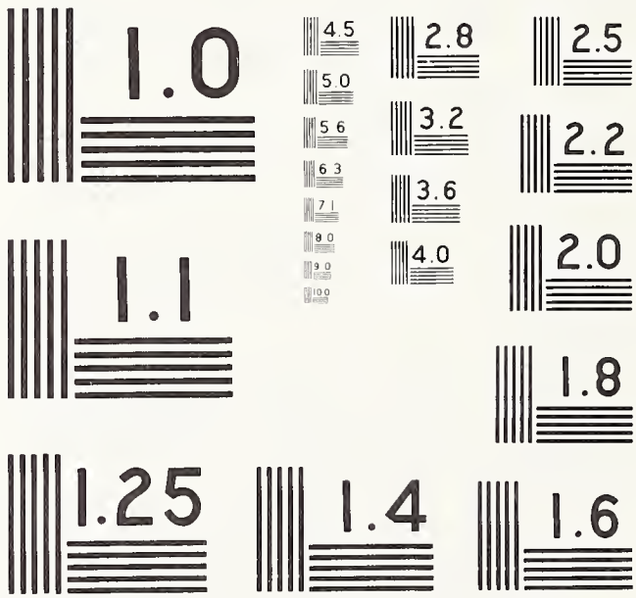


Figure 2. National Bureau of Standards
Microcopy Resolution Test Chart of 1963.

Note: These figures are for illustration only and should not be used for test purposes.

Among the phenomena to be considered is the retinal mosaic effect discussed in Helmholtz's celebrated treatise.[2] A pattern is considered resolved if the lines can be counted with certainty and if there are 5 lines in the pattern. If the camera were slightly out of focus, the number of lines on the copy may be other than 5. This is known as "spurious resolution" and is further indicated by failure to resolve at one spatial frequency when apparent resolution occurs at a higher frequency. The number of lines per millimeter, on the smallest chart pattern in which the lines can be counted with certainty, multiplied by the reduction ratio, gives the resolving power of the system. The ability to count the lines requires that there be a clear distinction between the lines and the background. A minor defect at one point in the pattern need not disqualify the pattern if it is clearly resolved over the remainder of the pattern.

It should be pointed out in this connection that the design of the resolution pattern is intimately related to the criterion of resolution to be employed. In some optical experiments involving visual processes, it has been useful to adopt a criterion of identification which merely involves identifying the direction of the lines in a pattern. The fundamental concept of resolution is the ability of a system to produce separate images of closely-spaced objects. Thus, an essential requirement of a criterion of resolution is that the lines be separate. When an observer is merely required to ascertain the direction of lines, the pattern must be symmetrical, either square or round, and the patterns must be randomly oriented so that the direction of the lines cannot be deduced from the shape of the chart. When the array is longer than it is wide or has a regular geometrical arrangement, as in the NBS Resolution Test Chart, the direction of the lines is very well known even before one looks at the pattern. In that case, any general impression of grain alignment, however vague, is sufficiently suggestive to make a person think that he sees lines where he already knows they ought to be. It is possible to design a chart for this kind of use, but the NBS Resolution Test Chart is not so designed. To require nothing more than to note the direction of the lines on this chart is a very weak requirement, which ignores both the fundamental definition of resolution and the lessons of psychophysics. As has been pointed out, the lines must be counted with certainty if one is to have any safeguard against spurious resolution.

---

[2] H. von Helmholtz. "Helmholtz's treatise on Physiological Optics," 3rd Ed. (1909), Vol II, p. 34 (translated), Dover Publications, Inc., New York, New York, (1962).

518-743 O - 73 - 56

Design of the Chart

In designing a resolution chart, many factors that influence resolving power must be considered. Measured resolving power is a function of the length of the lines, the number of lines, the line-to-space ratio, and the contrast of the lines of the test-target patterns. In the design of the NBS chart, these design variables have been selected so that small variations in them will give the least variation in measured resolving power. The length of the lines is 24 times their width. Although the measured resolving power is known to increase with line length, the lines of the chart are of such length that any increase in length would only produce a slight increase in measured values of resolving power. The reflectance ratio is approximately 100 to 1, which is the maximum obtainable on the photographic paper. Although there might be some argument for the use of a resolution chart of lower contrast, the production quality control problems would be increased many-fold if an attempt were made to print lines over this wide frequency range and maintain a reflectance ratio very much less than the maximum obtainable on the paper. The chart has 21 patterns, ranging in spatial frequency from 1 to 10 lines per millimeter. One line per millimeter is a pattern so coarse that it is unlikely that systems having a resolution less than this would be of any interest. The upper frequency limit, 10 lines per millimeter, is approximately the limit of resolving power of the human visual system at the normal viewing distance, without optical aids. Thus, this upper limit is a practical limit for testing the resolving power of microcopying systems intended to copy materials which are legible in the original. Successive patterns increase in spatial frequency by approximately the 20th root of 10. A coarser spacing would not provide a scale sufficiently finely divided for quality control in modern microfilming systems. Any finer subdivision of this scale would considerably increase the variability of evaluation from observer to observer. Measured resolving power increases with the number of lines in the test target, but there is not much increase beyond the values obtained with 5-line targets. In the NBS chart, each pattern is made up of two sets of 5-line targets at right angles to each other, permitting the measurement of radial and tangential resolving power simultaneously.

In redesigning the chart the values of spatial frequencies were assigned on the basis of the International Organization for Standardization Recommendation R3, a recommendation adopted at the 1953 meeting of the ISO in Geneva. The series used for the chart is based on Basic Series R20, which is given in Table I. With the exception of the 1.25 value, all the numbers were rounded off to two

Table I

Values of Spatial Frequency

Previous Chart	ISO R20	1963 Chart
1.0	1.00	1.0
1.1	1.12	1.1
1.25	1.25	1.25
1.4	1.40	1.4
1.6	1.60	1.6
1.8	1.80	1.8
2.0	2.00	2.0
2.2	2.24	2.2
2.5	2.50	2.5
2.8	2.80	2.8
3.15	3.15	3.2
3.5	3.55	3.6
4.0	4.00	4.0
4.5	4.50	4.5
5.0	5.00	5.0
5.6	5.60	5.6
6.3	6.30	6.3
7.1	7.10	7.1
7.9	8.00	8.0
8.9	9.00	9.0
10.0	10.00	10.0

figures. The previously used values 3.15, 3.5, 7.9, and 8.9 were given the values 3.2, 3.6, 8.0, and 9.0 respectively. The greatest percentage change is the 2.86 percent change from 3.5 to 3.6. Since observers often disagree on the reading of patterns differing in spatial frequency by 12.5 percent, this change may be regarded as practically imperceptible.

To permit greater economy and efficiency in production, it was decided to print 6 charts instead of 4 on an 8 x 10-inch sheet of photographic paper. The old charts are 4 x 5 inches; the new ones are 3-1/3 x 4 inches. Even though the new chart is on a smaller sheet of paper, the smallest margin is wider than the smallest margin on the old design. The center of the array of patterns on the new chart is in the horizontal lines of the finest pattern.

The patterns were slightly rearranged to leave space in the center of the array in which to place a 10-times reduction of the chart when it is desired to extend the range to 100 lines per

millimeter. The range could be extended again to 1000 lines per millimeter if desired. The new arrangement also permits abstracting a nearly-square array of patterns with frequencies from 2 to 10 lines per millimeter. In most applications the frequencies from 1 to 1.8 lines per millimeter are not often used.

The most important effect of the changes will be in increased efficiency of production. Production will be increased by 50 percent without a corresponding increase in operating cost with the exception of quality control and cutting operations.

Military Specification MIL-M-9868A, which gives requirements for microfilming of engineering documents on 35 mm film, requires that the corner charts be positioned "so that the number 4.0 is located on the test target diagonal line" at a given distance from the center of the test frame. The distance from the center of the chart to the decimal point in "4.0" differs by only 2 millimeters between the old chart and the new one. Since the symbol "4.0" is more than 2 millimeters in length, the variation from old chart to new chart is less than the uncertainty in the military specification. It is hardly to be expected that a microfilming system which would ordinarily pass a resolution test would fail because the chart were displaced 2 millimeters. Thus, it may be concluded that for all practical purposes, changing from the old chart to the new one would have no more than a trivial effect on the measurements.

Preparation of the Master Negatives

A very large scale model of the 5-line pattern was made by cementing precisely-cut strips of black paper on a plexiglass plate. This model was photographed repeatedly with the appropriate numbers placed on it. The negatives were separated into four groups: group I consisted of patterns 1.0 to 2.5 inclusive; group II, 2.8 to 4.0; group III, 4.5 to 6.3; and group IV, 7.1 to 10. Group I negatives were then contact printed and the resulting prints were reduced in proportion to their assigned spatial frequencies. Groups II, III, and IV were reduced and the resulting positives were arranged in the required order within their groups. These three arrays were then reduced once more to dimensions corresponding to those of group I negatives. The patterns were assembled in the final arrangement and reduced to 2.25 times the intended size. From this positive the master negatives were made and arranged to form the 6-chart configuration for printing on 8 x 10-inch sheets of photographic paper.

Variables Affecting Accuracy

The two major problems in producing charts by photographic methods are the dimensional instability of the photographic paper

on which the charts are printed and the dependence of the line-to-space ratio on exposure and processing. On the master negatives the line-to-space ratio is 0.99 ± 0.01. In a series of prints made recently for which the exposure time was controlled to ±0.1 second and the development time was controlled to ±1 second, the line-to-space ratio of the smallest pattern varied by less than 1 percent.

In the ferrotyping process there are dimensional changes in the photographic paper. These dimensional changes are ordinarily different for the directions corresponding to the length and width of the original rolls from which the sheets were cut. This effect may cause the widths of the lines and spaces in corresponding vertical and horizontal patterns to vary as much as 1-1/4 percent from the mean. Although this effect is well known, no attempt was made to compensate for it in making the master negatives because some variability is expected in the stretch of photographic papers from batch to batch. This effect could be compensated by printing the charts with the lines at 45 degrees to the horizontal and vertical, but it seems likely that this would lead to confusion in designating the sets of lines now known as "horizontal" and "vertical" when reporting resolving power. In a series of prints made recently, using the new master negatives, the spatial frequencies of the patterns, the line-to-space ratios, and the variability of the spacing in the horizontal and vertical lines were all maintained within 2 percent of the mean or nominal values. The dimensional instability in the paper causes greater variations than any other factor.

Conclusion

The new charts conform more accurately to the nominal values of frequency than the charts produced previously. The values of spatial frequency were assigned on the basis of a preferred number series contained in an ISO Recommendation to which the National Bureau of Standards subscribes. To permit greater economy and efficiency of production, 6 charts instead of 4 are now printed on an 8 x 10-inch sheet of photographic paper. On the new chart the patterns have been arranged to facilitate the extension or abridgement of the range of spatial frequency and to place the finest pattern in the center of the array.

THE PRODUCTION OF PHOTOGRAPHIC EDGES OF EXTREME SHARPNESS

Calvin S. McCamy and Myron A. Berkovitz
National Bureau of Standards
Washington, D. C.

Edge test objects of high sharpness are useful for the measurement of acutance and the edge spread function of photographic materials, and the edge response of camera systems and microdensitometers.

A number of different methods of producing these edges were evaluated before one was developed which produced edges of superior quality. Several different metal knife edges were made and various methods were used to hold them in contact with the film during exposure. To obtain a sharp photographic edge the metal knife edge had to be sharp and it was necessary to use pressure to hold it in close contact with the film. This caused physical damage to the film and pressure marks in the developed image. In an effort to avoid these difficulties, a knife edge was placed in a light beam and its image was projected on film by an optical system of high resolving power. In another attempt, a fine metal knife edge was placed in contact with a glass microscope slide and an extremely thin layer of aluminum was evaporated onto the unshielded portion of the microscope slide. This aluminum layer was probably only a few atoms thick. Films were exposed in contact with thin aluminum knife edge and others were exposed with the same edge blackened. All these methods were unsatisfactory as they suffered in varying degrees from problems of contact, straightness, diffraction, reflection at the front surface of the edge, and reflection of the light back-scattered from the emulsion.

It was known that if a fine "knife-edge" exposure could be made on a piece of film by any method whatsoever, this film could be used as the "knife edge" for exposures of test materials. Such an edge should be easy to use, should produce a less troublesome diffraction pattern, and should absorb the back-scattered light. It was necessary to make a metal edge to expose the film which would then be used as the working edge.

Edges of this kind were produced by clamping a piece of thin metal foil between two pieces of tool steel, then grinding and lapping this "sandwich". The foil was then placed in tight contact with a very fine-grain film and exposed to soft x-rays. X-rays were used because they produce a smaller diffraction pattern than light. Microdensitometer traces of this photographic edge showed it to be superior to edges obtained by all other methods tried.

Tantalum foil .01" thick was used to make the metal edge. Tantalum was chosen for the metal edge because of its density, mass absorption coefficient for soft x-rays, and machining characteristics.

Since tantalum foil is too flexible to be machined without support, two pieces of tool steel 105 cm x 2 cm x 1.5 cm were used to hold it and one 2 cm x 10.5 cm edge on each of these was ground and lapped optically flat. The tantalum was tightly clamped between these two flat surfaces. This steel-tantalum-steel "sandwich" was ground flat using a very fine grinding surface and lapped optically flat to .00001 inch. The flatness was checked by placing an optical glass flat over the sandwich and observing the straightness of the fringes produced by monochromatic illumination.

After machining, the foil was removed from the sandwich and placed in contact with a Kodak High Resolution Plate (649GH). The film and edge were placed in a 6 mil thick black polyethylene bag with an outlet to a vacuum pump. Black polyethylene was used since it is opaque to light and the x-ray machine is not in a dark room. It also has the advantage that it produces very little scattering of the x-ray beam. When the vacuum pump was turned on the bag was evacuated, and the foil edge was forced into very close contact with the photographic plate by atmospheric pressure.

The edge in contact with the film had to be very carefully alined in the x-ray beam so that the edge was in the center of the beam and normal to it.

A number of different developers were evaluated for their ability to produce sharp edges. It was found that the developer which produced the sharpest edge contained paraphenylene diamine as the developing agent and large amounts of sodium sulfite.

Photomicrographs and microdensitometer traces of the various edges exhibit the superior quality of the edges made by this procedure.

Reprinted from **APPLIED OPTICS**, Vol. 4, page 405, April 1965

On the Information in a Microphotograph

C. S. McCamy

The information concentration of a legible microphotograph made at a reduction ratio of 1200 is computed to be over 2×10^6 bits/mm^2. Reasons are adduced for using a mosaic of squares having sides equal to the reciprocal of the resolving power, each being capable of assuming one of two density ranges, as a model of a high contrast material for the computation of information capacity. It follows that the information capacity is equal to the square of the resolving power. This relation gives values of information capacity in good accord with the values computed by the Jones theory. The information capacity of a Kodak High Resolution Plate is computed to be 3.25×10^6 bits/mm^2. That the product of reduction ratio times quality index equals the product of resolving power times the height of the lower case **e** in type to be copied is verified at the high reduction ratio, and this relationship is considered useful for engineering computations involving the copying of commonly used styles of type. The use of literal charts designed on the basis of information theory is suggested for the evaluation of microcopying systems.

Introduction

Although graphic representation undoubtedly predated written language and written language predated electronic communication by many millenia, the formal theory of information and communication did not arise and become well developed until it was required in the field of electrical communication. The fact that graphic systems may be treated by the general theory was fully recognized by Shannon in his epoch-making treatment of the mathematical theory of communication.[1] Since the theory is formulated in terms of communication media employing time series encoding, the application of the theory to graphic media requires a scanning operation to convert the spatial distribution of optical densities to a function of time.

The frequency of a signal component in the time domain corresponds to a *spatial frequency* in the space domain. Spatial frequency is the number of cycles per unit of distance of a periodic variation in optical density (or opacity or transmittance) with respect to distance. Just as electrical communication channels are limited in the frequency of the signals which they can accommodate, photographic systems are limited in their ability to communicate spatial frequencies.

The simplest method of measuring this limitation is to photograph a series of patterns, each having a series of light and dark bands of equal width, and the series having a variation in spatial frequency from one pattern to the next. The photograph is then examined visually with the aid of a microscope of appropriate magnification, to find the highest spatial frequency resolved, known as the *resolving power* of the system. A pattern is considered resolved if the alternate bands can be counted with certainty. It is desirable to count the bands because lack of focus can produce *spurious resolution*, one result of which may be an apparent increase in the number of bands per pattern.[2] If one is attempting to evaluate the ability of the film to present detail so that it can be discriminated by eye, this method has the very important advantage of bringing into play, in the evaluation, the pertinent characteristics of the human visual system. This procedure measures the limiting spatial frequency in a practical sense, but gives no quantitative information about reproduction at lower frequencies. Nonetheless, the method is so quick and simple, requiring no more extra equipment than a microscope, that it is widely used, particularly in the microfilm industry. The National Bureau of Standards produces and supplies a Microcopy Resolution Test Chart for making such measurements.[3]

In 1937 Raymond Davis reasoned that the legibility of a microcopy of type should be related to the resolving power of the microcopying system and in 1938 he and Milo A. Durand experimented with various styles of type. They found, using common styles of type, that if the height of the lower case **e** on the film subtended three cycles of the highest spatial frequency resolved, the copy would be decipherable (that is, legible with difficulty), such letters as **e**, **c**, and **o** being partly closed. If the **e** subtended five cycles, the copy was legible without difficulty although serifs and fine details of type were not clearly defined. If the **e** subtended

The author is with the National Bureau of Standards, Washington, D. C.

Received 12 January 1965.

eight or more cycles, the details of the type were clearly defined. Davis included this correlation in the instructions for the use of the National Bureau of Standards Microcopy Resolution Test Chart. If one defines a *quality index*, his rules may be restated in the form of an equation,

$$m_r e = qR,$$

where

m_r is the resolving power in cycles/mm, on the film
e is the height of the lower case e, on the original, in mm,
q is a quality index, measured in cycles (assuming values 3, 5, and ≥ 8 in the above context), and
R is the reduction.

Taubes later confirmed the hyperbolic relationship between resolving power and legible letter height.*

There is a cooperative effort through the American Standards Association to establish a standard method of measuring the resolving powers of photographic films, plates, and papers. A high resolution camera was designed and built at the National Bureau of Standards as a part of that program.

When the resolution camera was completed, in the summer of 1960, some English text was copied to demonstrate the significance of the high resolving powers being measured and to test the legibility equation at extremely high reduction.[4] The prediction appeared to be well within the experimental limitations inherent in any judgment of legibility.

In 1961, R. C. Jones applied information theory to the computation of the information capacity of photographic films.[5] He was the first to deal with the general high-contrast case, using certain simplifications made necessary because the optical density on the film is not a linear function of exposure, the statistical distribution of variation of density with respect to distance is not independent of density, and the usable exposure range is peak limited.

In 1962 Fromm made a study of the problem of reproducing fine lines at extreme reduction and found that ideal exposure for one line-width is generally not ideal for another.[6] He noted that the height of the letter **e** alone is not the only characteristic of the type of interest and offered some evidence that line-width enters in. It has been our experience that the variation in line-width within a character or from one character to the next is the characteristic that is most troublesome, because the illumination and camera settings which are ideal for one line-width are not suitable for another. These facts were well known to Davis, and for these reasons he confined the application of his rules to "commonly used styles of type" and determined empirically what this author has called quality indexes.

Since then Altman and Zweig have conducted a study of the effect of the spread function on the storage of information on photographic emulsions.[7] The photographic material was considered to be made up of individual cells of a certain area, in each of which information could be stored as different density levels. The minimum size of cell was taken to be a square having sides equal to the diameter of the spread function at the 10% intensity level. Their analysis led to the conclusion that, although the use of several density levels contributes to higher information capacity on some materials, in general there is not sufficient gain to offset the many practical advantages of using two levels.

The computed information capacity was plotted against resolving power, and it was concluded that the relationship was linear, although the point for the high resolution plate certainly indicated an upward curvature. With respect to this relationship it was noted that "This relationship cannot be considered fundamental. On the other hand, neither is it entirely fortuitous because the reading of resolution samples is essentially the detecting of small elements of area against a noisy background. In any event, this curve does provide a simple method for estimating the maximum packing capacity of various emulsions, at least to a first order". It was concluded that the spread function diameter is a rough measure of minimum cell size when binary recording is used but that, with high contrast emulsions, somewhat smaller cells are possible. A very brief consideration of the information capacity for alphanumeric text was based upon the assumption of 1.5 bits/letter. They concluded that alphanumeric text can be stored at the same bit density as coded text.

Very recently (after the remainder of this paper was written and illustrations submitted to the publisher) De Belder, Jespers, and Verbrugghe published a study of image quality in photographic microreproduction.[8] They contributed a great deal to the solution of the problem of optimizing quality by the choice of the proper combination of characteristic curve, modulation transfer function, and granularity. They compared resolving power and information capacity as criteria of image quality. They state: "Obviously both resolving power and information capacity must be homogeneous functions in A and W, and from dimensional considerations one may write:

$$\text{information capacity} = C = A^{-1} f\left(\frac{A}{W}\right)$$

$$(\text{resolving power})^2 = m^2_{res} = A^{-1} g\left(\frac{A}{W}\right)",$$

where "A represents the surface of the section of the spread function at some definite level; and W represents the surface of the scanning aperture which gives a predetermined $\sigma(D)$".

The authors plot C and m^2_{res} as functions of two parameters and note that the striking resemblance of the two curves suggests the possibility of determining the information capacity of photographic materials by measuring their resolving powers. The shape of the curves for constant resolving power and constant legibility are said to be very similar, when plotted against the same two parameters, the resolution being

* Taubes, E. P., Reprod. Engr. (Sept. 1960), p. 9.

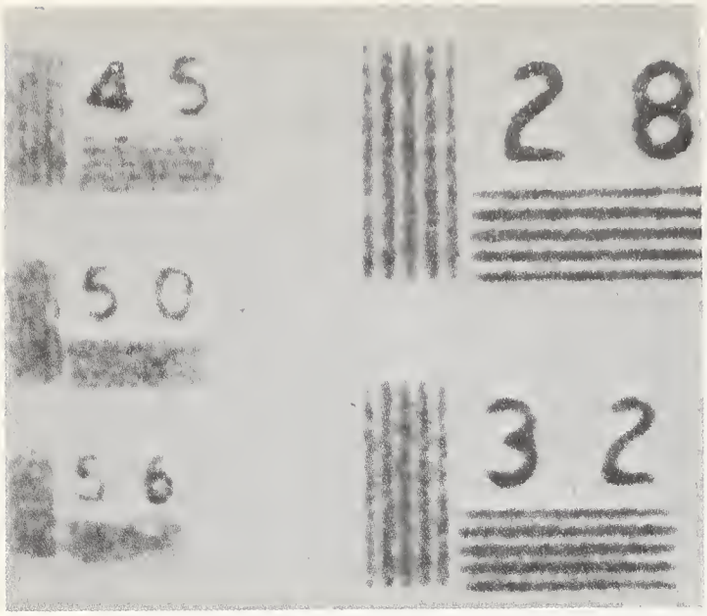


Fig. 1. Greatly enlarged copy of a resolution chart on a Kodak High Resolution Plate (modified thinner emulsion layer). The pattern marked 50 represents 2000 cycles/mm (the decimal point has been lost in the reproduction of the above figure).

based on a pattern "with a height/width ratio of about 1/4". (Presumably, "1/4" is a misprint and 4/1 was intended.)

Recent refinements in our resolution camera, the introduction of Kodak High Resolution Plates (Modified-Thinner Emulsion Layer), and our consistent attainment of 2000 cycles/mm on this emulsion prompted a new attempt to determine, for this equipment and material, the maximum reduction for a given English text and to compare the actual result with what little theory is applicable.

The "Camera"

The *camera* employed in these experiments consisted of a light source, condensing system, diffuser, transparency target, shutter, apochromatic microscope ocular and objective, (in that order), and a holder for the plate, all mounted in a massive steel cylinder to avoid relative vibration of the components. The microscope had a 10× compensating ocular and an objective with a numerical aperture of 0.65 and a focal length of 8 mm. The reduction of the optical system was 400. The emulsion surface of the plate was placed in contact with a stage, the position of which was adjustable with respect to the focal position by increments of about 0.1 μ. The source was a 100 W T–8½ incandescent lamp, used with an orange Wratten 15 (G) filter.

The Resolution Measurement

Kodak High Resolution Plates are used to establish the focus and proper condition of the camera when it is used to measure the resolving power of other films. In a proposed ASA method, a short line three-band target pattern (light and dark bands of equal width with length five times the width) is used, and the camera is considered in satisfactory condition when 1800 cycles/mm are resolved on the High Resolution Plates, using a high contrast target. With the long line five-band pattern used on the NBS Microcopy Resolution Test Chart, the system consistently resolves 2000 cycles/mm. The next pattern in the series, having a spatial frequency of 2240 cycles/mm is only rarely resolved, and then questionably so. To observe, judge, and photograph the resulting small patterns, a microscope with an oil-immersion objective having a numerical aperture of 1.30 was used at a magnification of 1125. A photomicrograph of the resolution pattern is shown in Fig. 1. Since the pattern was reduced 400 times in the camera, and the pattern marked 5.0 had a spatial frequency of 5.0 cycles/mm in the original, this pattern has a spatial frequency of 2000 cycles/mm on the microcopy. (The photomicrograph prepared for publication displayed lines which could be counted with certainty. It is not known whether this quality will withstand the processes involved in bringing it to the printed page. In reproduction, the decimal point was lost, so the pattern appears to be marked "50".)

The Microphotograph

The text of the first page of the King James version of Genesis, which is a fixed and well-known text, was first reduced in a copy camera employing a linear reduction of 3. This produced a negative which was placed in the target holder of the resolution camera and reduced 400 times, to make a positive microcopy with a total reduction of 1200. A photomicrograph of part of the microcopy is shown in Fig. 2. Since a real reduction is the matter of interest when considering the reduction of pages of English text, it is worth noting that the original page had 1,440,000 times the area of the microcopy.

A number of copies made with a reduction of 1600

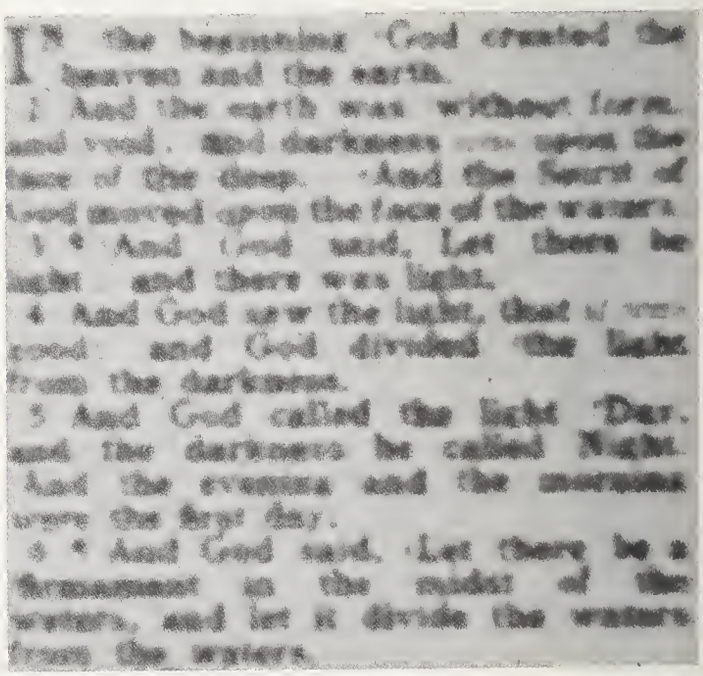


Fig. 2. Greatly enlarged copy of a microphotograph of English text. On the microphotograph, the height of the lower case e was a little less than 1.2 μ.

were barely legible to those familiar with the text. Thus, a reduction of 1200 represents the practical limit for the given text, emulsion, and equipment.

In the first experiment, the type had considerable variability in line-width so for the present experiment a style of type was chosen which was quite uniform and thus better designed for photographic reproduction.

Some of the *statistics* may be interesting in themselves, and will be used in the information analysis. Allowing small margins, each original page (other than the title pages), measured 112 × 175 mm, and there were two columns 49 mm wide separated by a center column of references. The height of the lower case **e** was about 1.4 mm in the original and a little less than 1.2 μ on the microcopy. There were 71 lines per column, 830 words per page (neglecting references), and 995 pages in the entire King James version of the Old and New Testaments. The reduced page, including margins and references, was 0.1 × 0.154 mm, so, the entire Bible reduced to this scale would occupy a square 3.93 mm on a side. Neglecting references and margins, the 830 words were contained in 0.012 mm^2. About a dozen such copies of essentially equal quality were made in succession, on Kodak High Resolution Plates developed in D-19 developer for 10 seconds at 20°C.

Information Analysis

In information theory, a source of information may be conceived as a selector which selects messages from among the various possible messages. More information is associated with a choice of one among many alternatives than one among few. Account is taken of the fact that some choices are more probable than others and more information is associated with an improbable choice than one which is likely. The amount of information is the logarithm of the number of independent choices among equally likely alternatives. The unit of information is the amount of information implied in the specification of one of two equally likely alternatives. Two alternatives can be designated by the symbols 0 and 1, and the choice indicated by specifying one or the other. The number of such choices required to select from among many equally likely alternatives can be represented conveniently in binary notation, that is, in numeration with the base 2. If there are 2^n alternatives, n binary digits are required to specify one of them. But $n = \log_2 2^n$, and therefore the number of *binary digits* is a measure of the amount of information when this logarithmic base is used. Thus, this base is used and "bit", a contraction of binary digit, is the name applied to the unit of information. The *entropy* of a source in bits is the minimum number of binary digits required per symbol or per second to transmit the messages selected by the source. The *capacity* of an electrical communication channel is the maximum number of bits which can be transmitted per unit of time. On a variable-density motion picture sound track, where density is modulated in just one direction, the information capacity is measured in bits per unit of length. On pictorial images, density is modulated in orthogonal directions, and the information capacity is measured in bits per unit of area.

The Kodak High Resolution Plates employed in the present experiment practically defy analysis with the methods used by Jones. The grain structure is so fine that the accurate measurement of the image structure characteristics required in the analysis is beyond the capability of available optical instrumentation. However, it is possible to measure the resolving power of the combined photographic system comprising the camera, film, viewing microscope, and *modus operandi*.

Given a photographic image of a high contrast resolution chart, produced by a given photographic system, if the bands can be distinguished well enough to be counted with certainty, the presence or absence of bands can be detected. If one were to use two transparency resolution charts superimposed with the bands at right angles, the transparent bands would be converted to a matrix of squares 0.5 cycle on a side separated by 0.5 cycle. In an isotropic system, at the resolution limit, one might reason that these squares could be counted with certainty, and the presence or absence of a square could be detected. There are two factors which should be considered in this line of reasoning. The first of these is that the human observer (or a microdensitometer with a long scanning slit) tends to average the density along a long band and to perceive the band even though it may be somewhat broken. Going from the one-dimensional case to the two-dimensional case, we go from a long rectangular band to a small square and the advantage of this averaging effect is lost. This would imply that the one-dimensional resolution limit would be at a frequency too high to permit the squares to be counted with certainty. A curve of resolving power as a function of the ratio of bandlength to bandwidth was given by Perrin and Altman.[9] Extrapolating the curve to lower values of this ratio, the maximum resolving power of the material appears to fall from 60 at a ratio of 5 to about 30 at a ratio of 1. This suggests that the one-dimensional frequency limit would be about twice the appropriate value. The second factor is that being able to count bands with certainty assures that the presence or absence of bands can be detected at the resolution frequency or at lower frequencies, but it does not preclude the possibility that the presence or absence of a given density level might be detected at even higher frequencies. The resolution chart is, in fact, a series of alternately present and absent bands, the frequency of the series being twice that of the alternate series ordinarily designated in photographic literature, and resolution means that the alternate presence and absence of bands is indeed detected. This would imply that the resolution limit is only half the frequency which might be employed. These two factors apparently tend to compensate one another. If so, for the purpose of information analysis, in the case of high contrast alphanumeric text copied on a high contrast material with good optics, we may consider the image to be made up of squares, with sides equal to the reciprocal of the resolving power, which may have one of two substantially different density ranges. If we assume

Table I. Comparison of Information Capacity and the Square of Resolving Power

	Resolving power,[a] m_r	m_r^2	Information capacity, C[b]
	cycles/mm	(cycles/mm)^2	bits/mm^2
Kodak Royal-X Pan	63.1	4000	4490
Kodak Tri-X Pan	100	10000	8450
Kodak Plus-X Pan	110	12100	18600
Kodak Panatomic-X	178	31700	28500

[a] The value for Royal-X was obtained at NBS, values for Tri-X and Plus-X were taken from those published by Altman,[15] and the value for Panatomic-X is the average of the values 160 reported by Altman and 196 reported by Fouquet of NBS.

[b] From ref. 5.

that the two density levels are equally probable, each square would have a capacity of one bit and the information capacity of the film in bits/mm^2 would be equal to the square of the resolving power in cycles/mm, $C = m_r^2$. In Table I, compare the square of the resolving power, based on bands having a height five times their width, with the information capacity of four films, as computed by Jones. The agreement is excellent in view of the mild claims of accuracy in Jones' paper, the approximations he had to employ, the experimental error of the measurements upon which his computations were based, and the experimental error of resolution measurement. The only discrepancy of any magnitude is with Kodak Plus-X Pan film, which Jones noted had anomalous image structure characteristics. Thus, there is a logical basis for computing the information capacity from resolving power, and the values computed are in essential agreement with the results obtained by applying the best available theory and the results of elaborate experimental procedures.

On this basis, the information capacity of the high resolution plates for the present purpose is $C = m_r^2 = (1800)^2 = 3.25 \times 10^6$ bits/mm^2. This is a little more than twice the value given by Altman and Zweig.

Now let us consider the information concentration of the recorded English text. Shannon has shown by an ingenious series of experiments, involving attempts to guess the succeeding letters in messages, that English text has an entropy between 0.6 and 1.3 bits/character.[10] Without troubling with the peculiarities of King James English, let us use 1 bit/character. Since there are about 4.5 letters/word and the space between words makes it 5.5 characters, the entropy is about 5.5 bits/word. The text then has an entropy of $5.5 \times 830 = 4560$ bits/page. We have put this information on 0.012 mm^2, giving a concentration of 3.8×10^5 bits/mm^2 or 2.45×10^8 bits/in.^{2} On this basis, about 12% of the capacity has been utilized. But bear in mind that the entropy gives the number of bits per unit of area if the English text were ideally encoded. Of course, it is not.

What has been recorded is the full text, one character at a time. The set of symbols includes at least twenty-six lower case letters, twenty-six upper case letters, twenty-six lower case italic letters, ten numbers, a space, and about eight punctuation marks for a total of ninety-seven symbols. Such a set is, of course, capable of representing not only English but many other languages and a great deal of numerical data. This great versatility could, in principle, cause great uncertainty at the receiver as to what the next character will be, and each transmitted character resolves a great deal of uncertainty. Thus, this text has more information value per character than the entropy of text strictly limited to English sentences. The set of ninety-seven symbols requires 6.6 bits/symbol to encode if one assumes the symbols occur with equal probability. Since there are about ten words per punctuation symbol, there would be another 0.1 symbol/word for punctuation, or a total of 5.6 symbols/word. This raises the information concentration to $6.6 \times 5.6 \times 830 = 30{,}700$ bits/page, that is, 2.56×10^6 bits/mm^2 or 1.65×10^9 bits/in.^{2} On this basis, the information concentration is 79% of the information capacity computed from resolving power. For this analysis to be valid, it would be required that the character by character record be of such fidelity that every character could be identified out of context. Taking such minute characters out of context presents some instrumental problems which we did not go into. A letter in context can be guessed with much higher frequency by a person familiar with the language than by a machine making random guesses and, in reading unclear type, one actually believes one sees the expected character even though it could not have been identified out of context. Because of this effect, the computed information concentration is probably somewhat high.

Finally, let us apply the legibility equation of the Introduction to this photographic system. The quality index is given by the following relation, where the resolving power for long lines is used, to accord with Davis' practice:

$$q = \frac{m_r e}{R} = \frac{2000 \times 1.4}{1200}$$

$$q = 2.34 \text{ cycles.}$$

We could not ask for better agreement with the implications of this index if we *extrapolate* from the grades of reproduction quality assigned by Davis. The quality is marginal, compared to usual microcopying practice, but the text upon which the computations were based was legible when viewed with the microscope. Even some of the prefatory and reference material in small type could be made out. (The image was magnified 3680 times for the photomicrographs submitted for publication. In such magnification, degradation is inevitable—as it is to some degree in the making of plates and in printing.)

It may be of some interest to apply this equation and the simple relationship between information capacity and resolving power to a modern microfilm, Re-

cordak Micro-File AHU Film, for which we have found a high-contrast resolving power of 631 cycles/mm. For fine print, $e = 1$ mm, let $q = 3$. Then the reduction ratio is

$$R = \frac{m_r e}{q} = \frac{631 \times 1}{3} = 210.$$

The information capacity of this material is about $C = (631)^2 = 4 \times 10^4$ bits/mm^2. This suggests that with good optics and careful control, considerably higher reduction ratios than are common today could be used for some purposes.

Since the first microphotograph was made a few months after Daguerre's disclosure, at a reduction ratio of 160, the idea of using such reductions is hardly novel.[11] Nor do increased reduction ratios appear unreasonable to the practical optical designer, judging by these remarks by Walter Mandler: "The trend in microfilming seems to be towards higher reduction ratios. It has been difficult enough to cope with a 30× standard reduction. Looking just from a side line, it is my impression that ten years from now a 50× or 60× reduction will be used for a considerable part of microfilming. The step towards larger reduction seems to be well prepared from the film manufacturers' side. A lot of work remains to be done in producing suitable lenses. From a design viewpoint the task looks difficult but feasible".[12]

Discussion

E.W.H. Selwyn recently opened a lecture on resolving power with this comment,[13]

> "Resolving power, as a means of estimating the detail-rendering capacity of a photographic system, is at the moment out of fashion. Except that it may be a mere matter of fashion it is hard to see why.
>
> In respect of ideas the theory of resolving power has all the physical content of the now prevalent theories of the modulation transfer function and much of the content of the theory of information. Moreover, only comparatively simple apparatus is necessary for measuring it.".

This is not to deprecate more elaborate forms of analysis. After all, the present study depends upon Jones' thorough analysis of information in photography. Rather, it may be said that information theory has given us new insights into the significance of resolving power. With all its pitfalls and shortcomings, it is still a very useful concept and measurement. Just as it would be bad science to neglect the more sophisticated concepts of image structure analysis, it would be bad engineering practice to employ them, however intellectually appealing they may be, if simpler and quicker methods give the desired result to the desired accuracy.

If resolving power is measured for the purpose of computing information capacity or permissible reduction, one may well ask why resolution charts should be used rather than literal charts, i.e., charts made up of alphanumeric characters. The problem of taking small characters out of context can be avoided by using an artificial text having meaningless sequences of letters. The readings of one or more observers on a number of characters would determine the percent of correct identifications. The sequences of letters could satisfy some requirements suggested by information theory and thus more truly represent a language or specialized field within a language. If character frequency, word length, sentence length, punctuation, word spacing, and capitalization were retained, much of the statistical and graphical character of the language would be preserved but the influence of context would be removed. This can be done by encoding text of the desired kind if each sentence is encoded by a *transposition cipher*, forming code words of the same length as the words in the sentence but using a sequence of letters taken one at a time from each of the words in the order in which they appear. The following sentence is this sentence so encoded: *Tfs itssehoes heonelni nc ltst ooeedwnn ei ccdneeg.*

A graded series of sizes of such code text could be used to test photographic systems. Literal charts have been the mainstay of optometry for years and have been suggested for testing photographic systems, but they have been little used in recent years.[2,9,16] As early as 1853, Rosling used a newspaper for this purpose.[14] There is the question of what style of type to use, since types vary in legibility even when the characters are well formed. There is no apparent reason why Snellen characters would not be useful test objects, but a commonly used style of type would be more representative of practice.

More time and effort would be spent reading literal charts but the result would be a quantitative evaluation for each size of type rather than a single judgment at some limit. The *literal response* may be defined as the ratio of the number of correctly identified characters to the total number observed. The *literal frequency* may be defined as the reciprocal of the height of the reduced **e**. A plot of literal response against literal frequency would have the dimensionality and much of the character of a *spatial response** curve. In the present experiments, the literal response was not measured but from the results it is estimated that it remained 1.0 out to 600 mm^{-1} or more, fell to 0.9 or so at 850 mm^{-1}, and to 0.5 or so at 1100 mm^{-1}.

It may be noted that the height of the reduced **e**, which we will call **e′**, is equal to e/R and that the literal frequency m_L is equal to R/e. Thus, the legibility equation can be rewritten in the form $q = m_r/m_L$. If we define the *literal frequency limit*, m_L, as the maximum literal frequency for which the literal response is 1, we have, for commonly used styles of type, the approximate relation: $m_r = 3\ m_L$.

Various styles of type could be evaluated for photographic legibility with respect to a given system by

---

* The author prefers this term for what has been called modulation transfer function or sine wave response.

making up literal charts and measuring m_L for each type. In the present experiment, the degree of agreement between the value for the information capacity computed from the resolving power and the value computed on the basis of considerations of the set of symbols implies that the alphabet as represented by the chosen style of type utilizes space quite well.

In adducing reasons for adopting the matrix model for the information analysis, some known factors were semiquantitatively taken into account. It is recognized that visual acuity thresholds involve very complex mechanisms and that factors such as the dynamic response of the visual system enter in. There is also a complex interplay between edge spread and granularity, and the nature of both of these depends upon the photographic material. The usual assumption of randomness of the grains is not really applicable here because even though the resolution pattern at the resolution limit displayed grain in the bands and between them, the literal characters appeared to be viewed against an essentially grainless background. Since the grains appeared only in the characters, they could hardly be said to be randomly distributed. The analysis was approached with the viewpoint that the resolution measurement involves the entire photographic system including the human visual system and partakes of all the factors influencing performance and probably in much the same way as these factors affect performance in microcopying.

In view of the tacit assumption of many experimenters that high resolution plates are essentially grainless, it should be noted that under the extreme conditions imposed here the grain was apparent and appeared to play a significant role in limiting the resolving power.

Summary

Numerous studies of the image structure of photographic materials and the storage of information on such materials have shown that a number of measures are related to the information capacity of a material. It is reasonable, for the information analysis of a high-contrast material for copying high-contrast alphanumeric characters, to consider the material to be represented by a mosaic of squares having sides equal to the reciprocal of the resolving power. This leads to the conclusion that the information capacity, C, is related to the resolving power, m_r, by the simple expression

$$C = m_r^2,$$

which is believed to be sufficiently accurate for most applications. The equation yields results that are in good accord with the Jones theory of information capacity for a number of materials and permits a computation of information capacity for a high-resolution material not readily treated by that theory. The information capacity of Kodak High Resolution Plates is found to be 3.25×10^6 bits/mm^2. A legible microcopy has been made which is computed to have an information concentration of more than 2×10^6 bits/mm^2.

The reproduction quality index, q, as here defined, the resolving power, m_r, the height of the lower case **e** in the type to be copied, e, and the reduction R, are related, for commonly used styles of type, by the expression

$$q = \frac{m_r e}{R},$$

which is confirmed in this experiment and is believed to be sufficiently accurate for most purposes. The use of carefully designed literal charts to evaluate photographic systems for microcopying is suggested.

The author wishes to express his appreciation of the skill and patience of W. R. Smallwood, who critically adjusted the camera, made the microphotographs, and made the photomicrographs at extreme magnification for this publication.

References

1. C. E. Shannon, Bell System Tech. J. **27,** 379, 623 (1948).
2. R. N. Hotchkiss, F. E. Washer, and F. W. Rosberry, J. Opt. Soc. Am. **41,** 600 (1951).
3. B. H. Fouquet, Proc. Natl. Microfilm Assn. **XII,** 67 (1963).
4. C. S. McCamy, Natl. Bur. Standards Tech. News Bull. **44,** 204 (1960).
5. R. C. Jones, J. Opt. Soc. Am. **51,** 1159 (1961).
6. H. J. Fromm, J. Phot. Sci. **10,** 147 (1962).
7. J. H. Altman and H. J. Zweig. Phot. Sci. Eng. **7,** 173 (1963).
8. M. De Belder, J. Jespers, and R. Verbrugghe, J. Phot. Sci. **12,** 296 (1964).
9. F. H. Perrin and J. H. Altman, J. Opt. Soc. Am. **43,** 780 (1953).
10. J. R. Pierce, *Symbols, Signals and Noise: The Nature and Process of Communication* (Harper, New York, 1961), p. 129.
11. Frederic Luther, *Microfilm: A History 1839–1900* (National Microfilm Association, Annapolis, Md., 1959), p. 16.
12. Walter Mandler, Proc. Natl. Microfilm Assn., **XII,** 23 (1963).
13. E. W. H. Selwyn, on Resolving Power in *Photographic Theory* Summer School on the Science of Photography, University of Liege 1962 (The Focal Press, New York, 1963), p. 223.
14. Frederic Luther, *Microfilm: A History 1839–1900*(National Microfilm Association, Annapolis, Md., 1959), pp. 22, 26.
15. J. H. Altman, Phot. Sci. Eng. **5,** 17 (1961).
16. E. P. Taubes; Reprod. Engr., (Sept. 1960), p. 9.

MC16. New Principle of Absolute Photometry. CALVIN S. MCCAMY, *National Bureau of Standards, Washington 25, D. C.*—For absolute photometry, a standard source, photometric comparator, and quantitative attenuator are required. Heretofore, quantitative attenuation has been based upon the inverse-square law or Malus' law of polarization. The principle of attenuation presented here is an optical analog of Riemann integration. A rectangular aperture in a light path is fitted with a sliding straight-edged shutter. The light flux through the aperture is then approximately proportional to the open-aperture area. A thin slit, instead of the shutter, is moved across the aperture while the light passing through the slit to a detector is compared to a constant source. If the response of the detector is constant as the slit moves, the shutter is replaced, and the flux through the aperture to the detector is known to be proportional to shutter displacement, which can be measured precisely. If this condition is not found, the edges of the aperture may be shaped or an optical compensator may be placed in the aperture to produce the required condition. Two slits may be used to establish an exponential or inverse-square relationship to shutter displacement. The method requires less space than the inverse-square method and is less subject to certain errors.

JOSA, Vol. 53, No. 4, p. 511 (April 1963)

PHOTOGRAPHIC SCIENCE AND ENGINEERING
Volume 10, Number 6, November–December 1966

Concepts, Terminology, and Notation for Optical Modulation*

C. S. McCamy, *Refractometry and Photography Section, National Bureau of Standards, Washington, D. C.*

Reflectance, transmittance, and optical density are regarded as kinds of flux modulation factors. Transmittance is defined as the ratio of transmitted flux to the incident flux, while transmittance factor is here defined for an optical system as the ratio of the emergent flux with the sample in the system to the emergent flux with the sample removed. Transmission density is defined as the negative logarithm of transmittance factor. This is essentially a new definition to accord with an old practice. Reflection, transmission, and fluorescence are considered generically as propagation and generalized modulation terms are proposed. A coordinate system and functional notation are adopted to systematize the description of optical systems which use or measure modulation. Simplified notation of the form $D(g;S:g';S')$ describes most cases. Terms, symbols, and notation are proposed for standardization.

Key Words: Modulation reflectance, transmittance, reflectance factor, transmittance factor, propagance, absorptance, optical density, contrast, symbols, and notation.

The photographic effect is usually measured and expressed in terms of transmission density or reflection density. Optical density is therefore one of the fundamental quantities in photographic science.[1–3]

American standards define various types of density and provide symbols to designate the types.[4–7] The types are differentiated on the basis of the geometrical arrangement of the optical components of the measuring system, the spectral quality of the flux incident on the sample, and the spectral sensitivity of the receiver used to evaluate the emergent flux. Certain arrangements which may be regarded as geometrically extreme cases are specified in simple terms. These include diffuse density (also called totally diffuse density), specular density, and doubly diffuse density. Two spectral types of density are defined: visual density and printing density. These spectral types correspond to the two principal uses of photographic materials, direct or projection viewing and photographic printing.

In mid-1963, the American Standards Association requested the author to form a committee to revise the existing standard on transmission density, in accordance with the customary schedule of revision. The charge to the committee was later extended to include all American standards on optical density. At the first meeting on November 18, 1963, I offered to propose a system of terms and symbols, including the functional notation which I had developed for solving problems in precise density calibration at the National Bureau of Standards.

In optical literature one often finds optical density defined as the common logarithm of the ratio of the incident flux to the reflected or transmitted flux. Although this definition is usually cited in photographic literature, in actual practice and in photographic standards, the term "density" has taken on a somewhat different connotation. Density has come to be regarded as a measure of the modulation of flux by a sheet, often bearing a graphic image, when the sheet is employed in a given type of optical system. The optical system may be as simple as the eye directly viewing a print or as complex as an optical motion-picture printer. Various methods of measuring density have been developed, using devices which simulate the important features of the optical systems in which materials are used. Thus in photographic science, and indeed in most other fields involving optical modulators, the optical density of an area on a given sample may have various values depending on the intended use and the corresponding method of measurement adopted to simulate it.

When this revision was undertaken it was not generally recognized that the usually expressed definition of density was not the one needed or used in most practical cases and, for this reason, there was no systematic terminology for the quantities involved. For years there have been differences between some of the terms and symbols used in photography and those generally accepted in photometry and colorimetry. Since there was much to be gained by bringing the standard terminologies and symbols of these disciplines into harmony, I had several conferences in June of 1965 with L. E. Barbrow and Deane B. Judd of the Photometry and Colorimetry Section of NBS. Judd had just returned from a meeting of the Committee on

Received August 10, 1966.

* The proposed functional notation presented at the Annual Conference, Cleveland, Ohio, May 18, 1965, is given here with some revisions. The other concepts were presented at the Annual Conference, San Francisco, Calif., May 12, 1966.

1. W. de W. Abney, *Phil. Mag.*, **48:** 161 (1874).
2. W. de W. Abney, *Phil. Mag.*, N. S., **11:** 38 (1887).
3. F. Hurter and V. C. Driffield, *J. Soc. Chem. Ind.*, **9:** 445 (1890).
4. American Standard Spectral Diffuse Densities of Three-Component Subtractive Color Films, PH2.1-1952.
5. American Standard Diffuse Reflection Density, PH2.17-1958.
6. American Standard Diffuse Transmission Density, PH2.19-1959.
7. American Standard Photographic Printing Density (Carbon Step Tablet Method), PH2.25-1965.

Nomenclature of The International Commission on Illumination to revise the International Lighting Vocabulary, and was therefore familiar not only with the official actions but with current attitudes of the representatives of various nations toward alternate terms. The concepts, terms, symbols, and notation are proposed in the hope that they will be acceptable in photometry as well as photography. The terms "transmission density" and "reflection density" and most of the symbols in current ASA photographic standards are retained.

Optical Modulation

When radiant energy interacts with matter, some of it is absorbed, and is converted into another form of energy, while some may be propagated by reflection, refraction, transmission, diffraction, or scattering. Some of the photons may be absorbed and then reradiated almost immediately, in less than 10^{-8} sec, with a reduction in frequency. This process is called *fluorescence*. For most purposes, that part of the energy which is reradiated in this way is effectively propagated. The term "propagation" is a useful general term for several of the processes with which we must deal. It is particularly useful in discussions involving both transmission and reflection.

We often need a general term for a measure of the effect of an object on the propagation of radiant energy from a source to a receiver. The object may change the direction, amplitude, frequency, phase, or polarization of the electromagnetic wave. The technical term for the process of varying the characteristics of a carrier wave to convey information is *modulation*. The graphic images produced by photography are mainly used as devices which modulate the flow of radiant energy and thereby convey information. Modulation has been used in about this sense for many years in sound-on-film technology.[8] This usage is standard in sensitometry.[9–14] The spatial variation of modulation of the light falling on objects around us permits us to receive visual information about these objects and to photograph them.

The modulation factor of communication theory is defined with respect to the amplitude of the signal that modulates the carrier wave.[15] Optical modulation factors, on the other hand, are usually measured in terms of ratios of fluxes and the flux (power) propagated is proportional to the square of the amplitude of the electric displacement associated with the electromagnetic wave. I suggest the term "flux modulation" to distinguish the modulation measured in terms of ratios of fluxes from amplitude modulation, phase modulation, polarization modulation, and frequency modulation; all of which have more relevance to photography and related fields than is generally recognized. We may note a few examples. Pictures printed on paper containing fluorescent dyes are effectively frequency modulators as well as flux modulators. Various effects observed when objects are placed between polarizers are utilized in high-speed shutters, stress analysis, thickness measurement, and chemical polarimetry. These applications depend on polarization modulation. The phase–contrast microscope is used to observe small objects which vary only in thickness or refractive index by utilizing the phase modulation resulting from these variations. A similar optical system is used to project photographic images produced by photoplastic recording. Thus "modulation" is a fitting generic term for a number of useful optical processes including the effects of objects such as pictures on the propagation of radiant energy. Transmittance and optical density are two of many measures of optical modulation.

Although phase modulation and polarization modulation are of interest in themselves, they are often converted by optical devices to flux modulation so that the effects may be observed visually or detected by flux-sensing devices. Since optical frequency detection is the basis of color vision and instrumental receiver sensitivities depend on frequency, spectral aspects of modulation are always considered along with the flux modulation. The remainder of this paper deals with flux modulation only, with due regard for fluorescence phenomena.

Although one could regard a whole optical, photographic, or television system, including the display, as an "object" which is irradiated on one "side" and viewed on the other, the usual conceptions of optical modulators are typified by sheet materials and the surfaces of bodies.

It should be noted that *modulation* does not necessarily mean *attenuation*. Oftentimes objects cause a gain in the flow of radiant energy from a source to a receiver.

These apparently far-ranging observations are made in the hope that optical modulation measurement can be viewed in a perspective sufficiently broad that the concepts, terms, and notation adopted will have the generality to cope with current problems and many yet to come.

Terms and Symbols for Fluxes

The measures of flux modulation most often used are ratios of fluxes or logarithms of these ratios. Therefore, the fundamental quantities to be identified and be given symbols are the fluxes. We deal in *radiant flux*, using the symbol Φ. *Luminous flux*

8. J. G. Frayne and H. Wolfe, *Sound Recording*, Wiley and Sons, N. Y. (1949), Chapters 15–19.
9. American Standard Sensitometry and Grading of Photographic Papers, PH2.2-1953.
10. American Standard Method for Determining Speed of Photographic Negative Materials (Monochrome, Continuous-tone), PH2.5-1960.
11. American Standard Method for the Sensitometry of Industrial X-ray Films for Energies Up to 3 Million Electron Volts, PH2.8-1964.
12. American Standard Method for the Sensitometry of Medical X-ray Films, PH2.9-1964.
13. The Theory of the Photographic Process, C. E. K. Mees and T. H. James, eds., Macmillan Co., N. Y., 3rd ed. p. 412.
14. T. H. James and G. C. Higgins, Fundamentals of Photographic Theory, Morgan & Morgan, Inc., N. Y. (1960).
15. W. D. Hershberger, Principles of Communications Systems, Prentice-Hall, Inglewood Cliffs, N. J. (1955), p. 77 et seq.

518-743 O - 73 - 57

TABLE I. Fluxes Involved in Optical Modulation

Term	Symbol	Definition
Incident flux	Φ_0	Flux incident on scanning aperture
Absorbed flux	Φ_a	Flux absorbed by sample
Propagated flux	Φ_p	Flux propagated by sample and utilized
Extra flux	Φ_q	Flux propagated outside spatial and spectral range of receiver
Lost flux	Φ_L	Flux not utilized
Reflected flux	Φ_r	Flux reflected by sample and utilized
Transmitted flux	Φ_t	Flux transmitted by sample and utilized
Reference propagated flux	Φ_{ps}	Flux propagated by reference standard and utilized
Reference reflected flux	Φ_{rs}	Flux reflected by reference standard and utilized
Reference transmitted flux	Φ_{ts}	Flux transmitted by reference standard and utilized
Aperture flux	Φ_j	Flux utilized with no sample in scanning aperture

is regarded as a specific case not fundamentally different from any other specific spectral distribution of flux. The symbol Φ is given an appropriate lower case subscript to identify the particular flux. The terms, symbols, and definitions of fluxes involved in optical modulation are given in Table I.

In the past, confusion and misunderstanding have stemmed from the failure of authors to state clearly whether the phrase "transmitted flux" necessarily meant all of the transmitted flux or might mean some specified part of it. To economize on symbols and terminology and at the same time make the notation completely explicit in all cases, I prefer to consider each transmitted, reflected, or propagated flux to be a specified part of the whole. In each case, one of the many possible specified parts would be *all* of the flux. Notation for specifying what part of the flux is utilized for measurement or practical purposes is given in a later section of this paper. There is usually no doubt about whether all or part of the incident flux is under consideration but since the value of modulation depends on the geometrical and spectral nature of the incident flux, the part of space and the spectrum from which the incident flux comes is denoted in a manner exactly analogous to that used to denote the part of space and the spectrum in which the emergent flux is utilized. Once the utilized part of the emergent flux is specified, the terms "transmitted," "reflected," and "propagated" mean transmitted, reflected, and propagated *in the specified part of space and the spectrum.* To provide a term and symbol for all of the flux, the prefix "total" may be used and the subscript may be capitalized.

The flux Φ_0 which is incident on a specified area (defined by a scanning aperture) on a body is either absorbed by the body, propagated to a specified receiver, or propagated in directions or parts of the spectrum other than that specified. The flux which is absorbed is called *absorbed flux* Φ_a, that propagated to the specified receiver is called *propagated flux* Φ_p, and that propagated outside the specified spatial or spectral range of the receiver is called *extra flux* Φ_q. That which is not propagated in the specified direction is called *lost flux* Φ_L and is the sum of the absorbed and extra fluxes. That which is not absorbed is called *total propagated flux* Φ_P.

When the propagated flux emerges from the surface on which the incident flux falls, it is called *reflected flux* Φ_r. When the propagated flux passes through the body and emerges from another surface, it is called *transmitted flux* Φ_t. The general term "propagation" is useful when the mode of propagation (transmission, reflection, scattering, fluorescence) is not known or it is known to be some combination.

The *aperture flux* Φ_j is the flux propagated from the source to the receiver with the sample removed from the scanning aperture but the remainder of the optical system undisturbed, as when a slide is removed from a projector.

In discussions of the propagation of flux within a medium it is convenient to have a symbol for the flux transmitted by the first interface and a symbol for the flux incident on the second interface. Sometimes several media are involved, in which case there are several interfaces. The interfaces can be numbered in the order in which they would be encountered by energy passing through the system and the terms and symbols for incident flux, reflected flux, and transmitted flux can be applied to the interactions at each interface, using the number designating the interface as a subscript. For example, in the simple case of a glass window we would have:

$$\Phi_{01},\Phi_{r1},\Phi_{t1} \text{ and } \Phi_{02},\Phi_{r2},\Phi_{t2}$$

Flux may be reflected back and forth between interfaces, in which case the subscripts for the interfaces may be appended in series. The notation "Φ_{r21435}" means that the flux was reflected by the

TABLE II. Measures of Flux Modulation

Arithmetic (modulation factor, m)	Logarithmic (optical density, D)
Transmittance, $\tau = \Phi_t/\Phi_0$	Transmittance density, $D_\tau = -\log_{10}\tau$
Reflectance, $\rho = \Phi_r/\Phi_0$	Reflectance density, $D_\rho = -\log_{10}\rho$
Propagance, $\psi = \Phi_p/\Phi_0$	Propagance density, $D_\psi = -\log_{10}\psi$
Transmittance factor, $T = \Phi_t/\Phi_j$	Transmission density, $D_T = -\log_{10}T$
Reflectance factor, $R = \Phi_r/\Phi_{rs}$	Reflection density, $D_R = -\log_{10}R$
Propagance factor, $P = \Phi_p/\Phi_j$	Propagation density, $D_P = -\log_{10}P$

second, first, fourth, third and fifth interfaces, in that order. If the number of interfaces exceeds nine, the subscript numbers would be separated by commas. To avoid confusion, it should be explicitly noted that the subscript for incident flux is a zero, not the fifteenth letter of the alphabet. This is generally accepted usage.

In discussions of instruments utilizing or measuring optical modulators, it is convenient to define two kinds of stray flux: *internal stray flux* Φ_{xi}, any flux affecting the sensor, from the source intentionally used to irradiate the sample but arriving at the receiver by any path other than the one specified for operation, and *external stray flux* Φ_{xe}, any flux affecting the sensor and coming from any source other than that used to irradiate the sample.

Terms and Symbols for Measures of Modulation

The terms, symbols, and defining equations for the various measures of flux modulation are given in Table II.

As they are defined here, transmittance, propagance, and reflectance are "absolute" measures in that they are independent of arbitrary reference standards. Reflectance factor is clearly a relative measure because a reference surface is used in the measurement.

Propagance factor and transmittance factor may be considered relative if one considers the air which normally replaces the sample as a "reference standard," but they are essentially absolute for most practical purposes. In photography we sometimes measure the modulation produced by a processed film relative to the modulation produced by the same kind of processed material with no exposure. Chemists often measure the modulation produced by a cuvette containing a solution relative to the modulation of the same or an identical cuvette containing pure solvent. One sometimes hears these measurements referred to as "relative transmittance" even though transmittance is defined, as it is here, as the ratio of transmitted to *incident flux*. Clearly, the concept employed in such instances is what we would call transmittance factor *relative to processed unexposed film* or *relative to solvent*. In the cases where a reference standard is substituted for the sample, the terms "relative transmittance factor, T_r" and "relative propagance factor, P_r" may be used and the defining equations become $T_r = \Phi_t/\Phi_{ts}$ and $P_r = \Phi_p/\Phi_{ps}$. The corresponding logarithmic terms would be "relative transmission density, D_{Tr}" and "relative propagation density, D_{Pr}" and the defining equations would be

$$D_{Tr} = -\log_{10} T_r$$

and

$$D_{Pr} = -\log_{10} P_r$$

Although it would be logically consistent to call $-\log_{10} R$ "relative reflection density," such usage would probably be much more confusing than helpful. The only "less relative" type that comes to mind would be a measurement with respect to the flux reflected by the scanning aperture with no sample or reference standard in place. Such a procedure is little used in reflectometry but if the occasion arises we have the general term "propagation density" defined in exactly this way. It is much more common, because it is useful, to use the term "relative reflection density" for measurements of reflection density using unexposed but processed photographic paper or blank printing paper as a reference standard instead of magnesium oxide or a standard plaque. When such a distinction is useful, the symbol "D_{Rr}" can be used for such relative reflection density.

Hurter and Driffield defined *opacity* as the reciprocal of transmittance and the term has been used this way in photography ever since.[3] To be consistent, we may define *opacity factor* as the reciprocal of transmittance factor, however, the use of these opacity terms is not encouraged because of frequent confusion with the concept of *hiding power*, the ability of a sheet or film to obscure a black-and-white pattern on which it is placed, as when a blank sheet of paper is laid on a printed page. In general, hiding power is not determined by transmittance alone, because the reflectance of the sheet must also be taken into account.

If we adopt the symbol m as the general symbol for any of the arithmetic forms of *optical flux modulation* τ, o, ψ, T, R, or P, and the symbol D as the general symbol for any of the logarithmic forms which we call *optical density*, we may give the general definition of optical density by the equation $D = -\log_{10} m$.

Theoretical computations can often be simplified by the use of *natural density** D_e defined by the equation $D_e = -\log_e m$. The symbols for the various types would be $D_{e\tau}$, $D_{e\rho}$, $D_{e\psi}$, D_{eT}, D_{eR}, and D_{eP}. The terms would be formed as in "natural transmission density." When necessary, the prefix "common" may be used to specify densities defined by logarithms to base ten. The two types of density are related by the equation $D = ND_e$ where, $N = \log_{10} e = 0.43429448$. Hurter and Driffield defined density in terms of $\log_e$ but then said, "For general convenience we use vulgar and not hyperbolic logarithms."[3] I have heard objection to the term "common" but it appears that we have made satisfactory progress in this regard in the past 76 years.

Terms and Symbols for Measures of Optical Attenuation

Most optical modulation measurements are made to find what part of the incident flux is propagated, but at times interest centers on the flux that is

* This name derives from the name of *natural logarithms*, often called *Napierian logarithms*. Strictly speaking, Napier's log x was not log $_e x$, see, e.g., Websters New International Dictionary, second edition, under "logarithm."

absorbed. The nature of the absorption of flux within an optical medium is an important characteristic of the medium, intimately related to its other chemical and physical characteristics. In studies of photochemical reactions, the absorbed flux is of primary importance because it provides the energy to initiate reactions. The *absorbed flux* is the difference between the incident flux and the *total propagated flux;* $\Phi_a = \Phi_0 - \Phi_P$.

In the process of modulation, flux may be lost not only by absorption but by reflection, refraction, diffraction, scattering, fluorescence, or transmission, depending on what part of the propagated flux is evaluated or utilized. The *lost flux* is the difference between the incident flux and the *specified part of the total propagated flux;* $\Phi_L = \Phi_0 - \Phi_p$. If absorbed flux is to be measured indirectly, all other losses must be taken into account.

If the nature of the medium is such that in a collimated beam the differential loss of flux in each differential layer of thickness is a constant fraction of the flux incident on that layer, that is if $d\Phi/dz = -a_b\Phi$, where z is the distance in the direction of the beam and a_b is the *thickness attenuation coefficient*, then we find by integration that the internal transmittance of a layer of thickness b is $\tau_i = \Phi_{02}/\Phi_{t1} = \exp(-a_b b)$. This equation, which holds for optical media in which there is negligible scattering and fluorescence, is known as Bouguer's law.

If the thickness attenuation coefficient is proportional to the concentration c of the substance which causes the attenuation, that is, if $a_b = ac$, where a is the *attenuation coefficient*, then the internal transmittance of a layer of thickness b and concentration c is $\tau_i = \exp(-abc)$.

The chemist, interested in the concentration variation for a constant thickness, uses a *concentration attenuation coefficient* $a_c = ab$ and the expression for internal transmittance becomes $\tau_i = \exp(-a_c c)$. This equation which is a good approximation for many materials, is called Beer's law. When a_c holds for a thickness b of one unit and c is expressed in molar concentrations, a_c is the *molar attenuation coefficient.*

The conditions imposed on these relationships should not be neglected. For some substances, Beer's law fails. If a material scatters a great deal, multiple scattering of flux invalidates the assumption of constant fractional attenuation in each differential of thickness. Nonetheless, Beer's law is useful in chemistry and Bouguer's law is useful in atmospheric optics.

In studies of the visibility of signal lights, it has been found that the atmosphere scatters very little flux in the backward direction. The flux reflected in the second, third, and other successive reflections is negligible for a receiver, such as one or a few retinal receptors of the eye, which is sensitive over a small angle. Practically all scattered light is lost, as though it were absorbed. In such cases, scattering may be treated mathematically in the same manner as absorption, the conditions for the derivation of the law of exponential attenuation hold, and we may write $\tau_i = \exp -(a_a + a_s)bc$, where a_a is the *absorption coefficient* and a_s is the *scattering coefficient.* In general, the relationship between modulation and the absorption and scattering coefficients is by no means this simple.

Since all materials have molecular density variations, all exhibit some scattering. For this reason the term *absorption coefficient* should not be used in place of the more general term *attenuation coefficient,* but should be reserved for its appropriate use.

Any of the exponential expressions for internal transmittance can be written in terms of powers of ten rather than e. I suggest that such coefficients be called "coefficients to base 10" to differentiate them from "coefficients to base e" and that such coefficients to base 10 be identified by the prime symbol, for example a'. The corresponding coefficients are related as follows: $a' = Na$, where $N = \log_{10}e = 0.43429448\ldots$.

Absorption and scattering characteristics are, of course, dependent on wavelength. Thus the various attenuation coefficients are functions of wavelength.

The various measures of attenuation are defined in Table III.

Absorptivity is usually defined as the absorptance of a layer of unit thickness. However, in some studies of radiation and modulation, absorptivity is defined as the limiting absorptance as the thickness increases infinitely, in a manner analogous to that shown for reflectivity in Table III. To differentiate these kinds of absorptivity, I suggest the following terms and symbols: "limiting absorptivity" α_∞ and "unit absorptivity" α_a.

Since the reflectance of a body depends on the condition of its surface, the measurement of the characteristic reflectivity of a material requires careful attention to the condition of the surface.

TABLE III. Measures of Optical Attenuation and Related Quantities

Properties of a body	
Internal propagance	$\psi_i = \Phi_{02}/\Phi_{t1}$
Internal propagance density	$D_{\psi i} = -\log_{10}\psi_i$
Internal transmittance	$\tau_i = \Phi_{02}/\Phi_{t1}$
Internal transmittance density	$D_{\tau i} = -\log_{10}\tau_i$
Absorptance	$\alpha = \Phi_a/\Phi_0$
Internal absorptance	$\alpha_i = \Phi_a/\Phi_{t1}$
Properties of a material	
Reflectivity	$\rho_\infty = \lim_{b\to\infty} \rho$ (b = thickness)
Transmissivity	$\tau_u = \tau_i$ when thickness is 1 unit*
Absorptivity	$\alpha_u = \alpha_i$ when thickness is 1 unit*
Attenuation coefficient (base 10)	$a' = D\tau_i/bc$ (b = thickness and c = concentration)
Attenuation coefficient (base e)	$a = -\log_e\tau_i/bc$ (b = thickness and c = concentration

* The value of transmissivity or absorptivitiy depends on the unit of thickness, which must be stated.

Contrast

Optical contrast is an evaluation of the relationship between two magnitudes of some attribute of appearance, compared simultaneously or successively. Luminance contrast and the modulations that produce it are of interest here.

Several objective measures have been used, some of which were chosen to correlate with subjective visual evaluations. According to Weber's Law, as it applies to luminance, the least preceptible difference between the luminances of two adjacent fields is proportional to the luminance at which the comparison is made. Thus the ratio of the difference of two luminances to the mean value is a measure of contrast proportional to the least perceptible difference at any mean value of luminance, over a large part of the useful range. The difference is sometimes divided by the lower of the two values instead of the mean of the two. The current edition of the International Lighting Vocabulary[16] gives three expressions for objective measures of contrast, referring to them as expressions (a), (b), and (c), without assigning symbols. I propose the use of the symbol C, always with a subscript, for contrast and the use of a, b, and c, as subscripts to correspond with the types given in the Vocabulary, thus:

$$C_a = \frac{L_2 - L_1}{L_1}$$

$$C_b = \frac{L_2 - L_1}{(L_2 + L_1)/2}$$

and

$$C_c = L_2/L_1$$

In photographic tone reproduction theory the term "contrast" usually refers to a *contrast ratio* C_c, which may be the ratio of scene luminances or it may be the ratio of two transmittance factors or reflectance factors on a photograph. This property of the photograph may also be indicated by a difference of optical densities. This measure of contrast is particularly suitable for describing a scene to be photographed or a negative to be printed. The ratio of scene luminances determines the ratio of exposures for any shutter and aperture settings so that, with appropriate settings and choice of film, the exposure scale of the film can be matched, within practical limits, to the ratio of exposures. The density difference on a negative determines the log exposure difference in the printing operation, regardless of the amount of flux in the beam modulated by the negative. The print material may be selected on the basis of the log exposure difference and the flux and exposure time are adjusted to the speed of the print material. Weber's Law has no bearing, since the negative is not being viewed directly.

One should clearly distinguish between contrast, which is a characteristic of the scene or the photograph, and gamma, the slope of the straight-line portion of the characteristic curve, which is a characteristic of the photographic process. The contrast between two parts of a picture depends on the scene as well as gamma. Incidentally, the contrast between the lightest and darkest areas may not be the contrast of interest. For example, when a person's face is photographed by artificial light outdoors at night, the contrast between parts of the face may be low while the contrast of the face to background may be extremely high.

In the case of a photograph viewed directly or by projection the mean luminance depends not only on the photograph but on the illumination. (Only with constant illumination, as in a standardized motion-picture review room, is the mean luminance determined by the photograph.) Nevertheless, the subjective contrast is determined by the photograph. If we define contrast C_m as the luminance difference $(L_2 - L_1)$ divided by the least perceptible difference ΔL at luminance L and assume Weber's Law

$$\Delta L/L_{mean} = k$$

then

$$C_m = (L_2 - L_1)/\Delta L = (L_2 - L_1)/kL_{mean}$$

$$C_m = 2(L_2 - L_1)/k(L_2 + L_1)$$

Given a transparency with areas having transmittance factors T_1 and T_2 and uniform illuminance E, then

$$C_m = 2(ET_2 - ET_1)/k(ET_2 + ET_1)$$

and

$$C_m = 2(T_2 - T_1)/k(T_2 + T_1)$$

Let $C_M = (k/2)C_m$, then

$$C_M = (T_2 - T_1)/(T_2 + T_1)$$

a quantity which I propose to call *modulation contrast* because the above derivation demonstrates that it is clearly a measure of contrast, it is computed from measures of modulation, and it has often been referred to as "modulation" in a special sense.

According to Stevens[17] the value of k is about 0.016. Therefore

$$C_M = C_m/125 = (L_2 - L_1)/125\Delta L$$

that is, C_M is a measure of contrast, the "units" of which are about 125 times the least perceptible difference. Since the constant k varies with luminance on the high and low ends of the luminance

16. International Lighting Vocabulary, 2nd Ed., Vol. 1, International Commission on Illumination Bureau Central, 57 Rue Cuvier, Paris 5, France.

17. S. S. Stevens, "Sensation and Psychological Measurement," in Edwin G. Boring, Herbert S. Langfeld, and Harry P. Weld, eds., Foundations of Psychology. Wiley, 1948, p. 268.

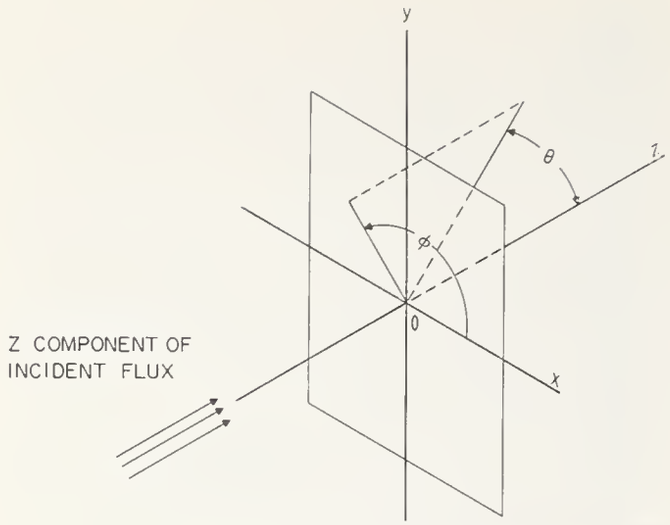


Fig. 1. Coordinate system for describing geometrical factors affecting optical modulation.

scale and with the nature of the field of view, the relationship given is approximate.

The ratio of the difference to the sum (C_M) is equal to the ratio of half the difference to the mean. This expression corresponds to one of the fundamental quantities in communication theory, the ratio of the amplitude to the mean value of a sinusoidally varying amplitude modulation of a carrier wave. It is called the "modulation factor" or simply "modulation." This specific application of the term "modulation" should be distinguished from the generic term which is used on this paper and is well established in "frequency modulation" and "phase modulation." (The "modulation index" in frequency modulation theory is quite different from the "modulation factor" in amplitude modulation.)

Coordinate System

Figure 1 provides a coordinate system for describing the geometrical factors affecting optical modulation. The coordinates relate to a reference plane, the xy plane, which is the plane in which the first surface of the sample is placed for measurement or use. The point on the sample placed at the origin 0 is the center of the area said to have its density measured. The z axis is directed with the component of incident flux normal to the reference plane.

The colatitude of a ray is the angle θ between the ray and the z axis. Unless otherwise stated, the colatitude θ of incident rays and reflected rays shall be measured from the *negative* z axis while the colatitude of transmitted and propagated rays shall be measured from the *positive* z axis. This special convention should be the most convenient for most cases but the description of some measurements and applications would be simplified by measuring all colatitudes from the positive z axis, according to the usual mathematical convention. The longitude of a ray is the angle ϕ measured in the xy plane, from the positive x axis in the direction of the positive y axis to the projection of the ray on the xy plane. The direction of a ray is given by the coordinates θ, ϕ, in that order. Angle ϕ is less than 360° and θ is 180° or less.

If the sample is thick, it may be desirable to refer the description of the efflux to a secondary coordinate system x', y', z', θ', ϕ', having its origin $0'$ displaced a specified distance b, corresponding to the thickness, along the positive z axis, so that $x'=x$, $y'=y$, $z'=z-b$ and the angles are defined in a corresponding manner.

Subscripts are used for the direction angles of rays to correspond to the subscripts for fluxes, that is, "0" for incident, "r" for reflected, "t" for transmitted, and "p" for propagated.

Functional Notation

Functional notation provides a convenient and generally understood means of denoting a dependent variable and the independent variables, known as "parameters," on which its value depends. The symbol for the dependent variable is followed by the symbols for the various parameters, or their values in a given case, in parentheses. Using this notation, we may write the fact that optical density D depends on the conditions of measurement or use, as follows:

$$D(C:C')$$

where C is the *influx function*, the complete description of the flux incident on the sample and C' is the *efflux function*, the complete description of the sensitivity of the system receiving flux from the sample and using or evaluating it. Let C and C' be given the generic name "condition functions." The influx function is, in the general case, the radiance L_e at the sample as a function of angles θ and ϕ, wavelength in vacuum λ, and the spatial coordinates x and y,

$$C = L_e(\theta, \phi, \lambda, x, y)$$

The efflux function is, in the general case, the sensitivity S_e of the receiver system as a function of angles, wavelength, and spatial coordinates,

$$C' = S_e(\theta, \phi, \lambda, x, y)$$

If the samples to be considered are uniform over the entire area of interest, the dependence of the condition functions on x and y can usually be disregarded, since the ratios of fluxes are not influenced. Then

$$C = L_e(\theta, \phi, \lambda)$$

and

$$C' = S_e(\theta, \phi, \lambda)$$

Furthermore, in most applications, the measured value of modulation does not depend on the absolute magnitudes of L_e and S_e, but only on their relative

spectral and spatial distributions. In the remainder of this paper it should be understood that relative distributions are implied, with the understanding that in all questionable cases, the constancy or variation of modulation with the absolute magnitude of L_e should be determined by measuring modulation with various values of incident radiance. The absolute sensitivity is not involved because modulations are defined in terms of ratios of fluxes.

If samples are nonuniform with respect to x or y or both, the measured modulation may depend on the product $L_e(x,y)S_e(x,y)$ known as the *aperture function*. Photographic films are usually nonuniform for any aperture function having an area less than one square millimeter. For this reason, the aperture function is of particular interest in microdensitometry, image evaluation, and applications such as sound-on-film technology involving film scanning. In such instances the aperture function should be specified. It is often regarded as constant over a rectangle or circle. In some applications, as in granularity measurement, the scanning aperture function is symmetrical with respect to the z axis and can be most easily described as a function of a polar coordinate r, the distance from the origin in the reference plane.

The condition functions C and C' describe both geometrical and spectral conditions. In very many practical cases either geometrical or spectral factors are of interest, but not both. For this reason and because it leads to a more simplified notation in most practical cases, let us separate the geometrical and spectral factors as follows:

$$C = g;S$$
$$C' = g';S'$$

where

- g = the symbol for the *influx geometry*, the spatial distribution of radiance at the irradiated surface of the sample;
- S = the symbol for the *influx spectrum*, the spectral distribution of the flux incident on the sample;
- g' = the symbol for the *efflux geometry*, the spatial distribution of the effectiveness of the optical system in utilizing radiant flux received from the sample;
- S' = the symbol for the *efflux spectral sensitivity*, the spectral sensitivity of the receiver which utilizes the radiant flux from the sample, including the effect of any filters, lenses, integrating sphere, or other optical components.

The dependence of optical density on these parameters may be denoted as follows:

$$D(g;S:g';S')$$

In the general case, g is the radiance as a function of θ, ϕ, x, and y,

$$g = L_e(\theta, \phi, x, y)$$

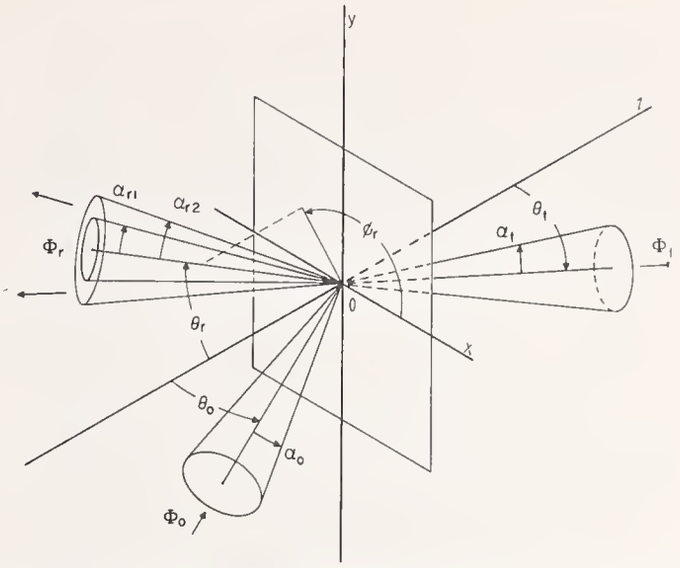


Fig. 2. Coordinate system for simplified description of geometrical factors affecting optical modulation. α_0, α_t, α_r = half cone angle of incident, transmitted and reflected pencils. θ_0, θ_t, θ_r = colatitude, also called angle of incidence, transmission, or reflection (of pencil axis) with respect to the normal (θ_0 and θ_r are measured from the negative z axis). ϕ_0, ϕ_t, ϕ_r = longitude of axis of incident, transmitted, and reflected pencils.

S is the radiance as a function of λ,

$$S = L_e(\lambda)$$

g' is the sensitivity of the receiver as a function of θ, ϕ, x, and y,

$$g' = S_e(\theta, \phi, x, y)$$

and S' is the sensitivity of the receiver as a function of λ,

$$S' = S_e(\lambda)$$

In general, the description is rather complex. The description in general form is complex even for the simple conditions most often used in photographic instruments and adopted for measurements. Fortunately, by the adoption of certain conventions the description can be very greatly simplified in most cases.

A very large percentage of practical geometrical arrangements used in photography can be described adequately in terms of uniform pencils bounded by right circular cones. In projectors and enlargers (ideally considered), the projection lens utilizes flux uniformly within the conical solid angle it subtends at the point on the film on the optical axis. In enlargers with diffuse illumination, the point is regarded as being uniformly illuminated from all angles throughout the hemisphere. When condensers are used, they are designed to image the source on the projection lens so that all rays from the condenser pass through the projection lens. It follows that the rays within the cone subtended by the projection lens on axis at the film plane would come from within an influx cone subtending the same angle. The same conical conditions are found

in many densitometers. For the description of these cases a special simplified form of notation is adopted.

The simplified notation is applicable if, for every point on the area of the sample under consideration,[1] the angular distribution of incident radiance is adequately described as uniform at all angles within a right circular cone with its apex at the point but zero at all other angles and[2] the angular distribution of the sensitivity of the receiver (including the effect of any filters, integrating sphere, or other optical components) is adequately described as uniform at all angles within a right circular cone with its apex at the point (or between two co-axial cones in the case of "annular collection") but zero at all other angles.

Then it is no longer necessary to treat the radiance L_e as a general function and the influx geometry may be described by three parameters, shown in Fig. 2; α_0, the half-angle of the incident cone, and θ_0 and ϕ_0, the direction angles of the axis of the cone:

$$g = \alpha_0,\ \theta_0,\ \phi_0$$

The efflux geometry may be described in a similar manner:

$$g' = \alpha_r,\ \theta_r,\ \phi_r \quad \text{in the reflection case}$$

$$g' = \alpha_t,\ \theta_t,\ \phi_t \quad \text{in the transmission case}$$

The geometry of the simple cases may be specified by giving these angles in degrees. The angle α is less than 180°.

As it happens, in the great majority of applications in photography, the sample, is irradiated and "viewed" normally so $\theta_0 = \theta_r = \theta_t = 0$. If $\theta = 0$, ϕ is indeterminate and, by convention, both can be omitted. Thus in most cases, the geometrical conditions may be specified adequately by one angle for the influx geometry and one for the efflux geometry. For transmission, the form would be $D_T(\alpha_0;S:\alpha_t;S')$.

Two alternatives to the specification of α in degrees are useful, particularly in applications involving projectors or microdensitometers. If the sample is irradiated by an optical system focused on the sample and the efflux is received by an optical system focused on the sample, the influx and efflux geometries may be specified by the influx numerical aperture A_n and the efflux numerical aperture A'_n, where

$$A_n = n \sin \alpha_0$$

$$A'_n = n' \sin \alpha_t$$

n is the refractive index of the medium between the irradiating optics and the sample and n' is the refractive index of the medium between the sample and the receiver optics. For systems operating in air, $n = n' = 1$. To indicate that geometries are specified by numerical apertures rather than angles, the subscript n is placed after the parenthetic statement of parameters, for example, $D_T(0.25;\ 2850°K:\ 0.40;S\text{-}4)_n$. As another alternative, the influx and efflux geometries may be specified by the influx f-number A_f and the efflux f-number A'_f where the f-numbers are those of the irradiating and receiving optics, but only when both optical systems are focused on the sample, the distance from the exit pupil of the irradiating optical system to the sample is nearly equal to the equivalent focal length of that system, and the distance from the sample to the entrance pupil of the receiver optical system is nearly equal to the equivalent focal length of that system. In the case of condenser optics, as used in projectors, where the condenser is designed to just fully irradiate the projection lens, $A_f = A'_f$. The requisite focal conditions are usually fulfilled for a system simulating a projector projecting a picture on a screen more than 25 focal lengths away from the lens. To indicate that geometries are specified by f-numbers rather than angles, the subscript f is placed after the parenthetic statement of parameters, for example $D_T(3.5;2850°K:\ 3.5,V)_f$.

Having simplified the description of geometric conditions in most practical cases, we turn to the spectral conditions and find that the same degree of simplification can be effected. The spectral distributions of many of the sources of interest in photography are satisfactorily represented by color temperature in degrees Kelvin. Others can be

TABLE IV. Notation for Parameters of Modulation

Influx function C	:	Efflux function C'	Reference/ standard s
Influx geometry g; Influx spectrum S:		Efflux geometry g'; Efflux spectral sensitivity S'/standard s	
$L_e(\theta,\phi)$; $L_e(\lambda)$	:	$S_e(\theta,\phi)$; $S_e(\lambda)$	/ s
Simplified notation when the angular distribution of irradiation and collection are uniform within right circular cones, zero elsewhere, and uniform over the sample (shown for the case of reflection):			
$\alpha_0,\ \theta_0,\ \phi_0$; S	:	$\alpha_r,\ \theta_r,\ \phi_r$; S'	/ s
Simplified notation when above conditions are met and $\theta_0 = \theta_r = 0$.			
α_0 ; S	:	α_r ; S'	/ s
For annular irradiation, when above conditions are met:			
α_{01} to α_{02} ; S	:	α_r ; S'	/ s

specified by symbols adopted by standardizing organizations for certain standardized sources, such as CIE-C, or phosphors, such as P-11. If the influx spectrum is a very narrow band of frequencies, it may be specified by the wavelength, in a vacuum, preferably in nanometers (abbreviated nm and formerly called millimicrons).

Similarly, the efflux spectral sensitivity can be denoted by the symbols for standardized spectral sensitivity functions such as V for visual, P for the standard sensitivity representing commonly used photographic printing materials, S-4 (for example) for a particular photocell response, or when appropriate, a wavelength in nanometers.

To indicate that a measurement was made relative to a reference standard, it is proposed that a symbol for the reference standard, s, be included at the end of the list of parameters and be separated from the efflux parameters by a diagonal, as follows:

$$D(C:C'/s)$$

The most common reference standards used in photography are magnesium oxide, base, and base plus fog. These may be given the symbols MgO, b, and bf, respectively. Other references may be assigned symbols or the general symbol s may be used when appropriate. The functional notation in the general and simplified forms can be used to specify the parameters of any of the arithmetic measures of modulation, as well as the various kinds of optical density, as follows:

$$\tau(C:C'),\ \rho(C:C'),\ \psi(C:C'),\ T(C:C'),\ R(C:C'),\ \text{and } P(C:C')$$

The general features of the system of notation and the various simplifications are shown in Table IV.

In some standard systems of measurements, flux is incident uniformly in the annular region between two co-axial cones or the efflux is uniformly evaluated between two such cones. In such cases the notation would be as shown in these examples for the reflection case:

$$D_R(\alpha_{01} \text{ to } \alpha_{02};S:\alpha_r;S')$$

or

$$D_R(\alpha_0;S:\ \alpha_{r1} \text{ to } \alpha_{r2};S')$$

When an angular subtense is broken into two or more discrete regions it may be described by the use of the "plus" sign, for example,

$$\alpha_{01} \text{ to } \alpha_{02} + \alpha_{03} \text{ to } \alpha_{04}$$

Similarly one may indicate a region deleted by the use of the "minus" sign, for example,

$$0 \text{ to } 360^\circ - 170^\circ \text{ to } 190^\circ$$

It should be noted that the punctuation used in the notation has been formalized in the interest of clarity. The influx function and efflux function are separated by a colon. The geometrical and spectral functions are separated by semicolons. The angular parameters of the geometrical functions are separated by commas.

If a second coordinate system for efflux description is established at a distance b along the z axis, as discussed in the section dealing with the coordinate system, the distance b, with proper unit of length, may be indicated between two colons and this group of symbols may replace the usual single colon, as follows:

$$D_T(\alpha_0;S:\ b:\ \alpha_t;S').$$

Examples of Functional Notation

The current American standard for diffuse transmission density defines two types. American standard diffuse visual density, Type Vl-b would be written in the proposed notation as follows:

$$D_T(90^\circ;3000^\circ\text{K}:\leq 10^\circ;V)$$

This notation means that the incident flux is uniformly distributed at all angles over the hemisphere and has the spectral distribution of a full radiator at 3000°K and that the flux is evaluated uniformly in a cone of half-angle less than or equal to 10° with its axis on the normal, by a receiver having a response proportional to the standard photopic spectral luminous efficiency function. The second type, American Standard Diffuse Printing Density, Type P2-b would be written in the proposed notation as follows:

$$D_T(90^\circ;3000^\circ\text{K}:\leq 10^\circ;P)$$

One can see at a glance that the only difference from the first type is that the spectral sensitivity of the receiver is that of a standard photographic print paper. The current American Standard Visual Diffuse Reflection Density, Type 3000 K, would be written:

$$D_R(40^\circ \text{ to } 50^\circ;3000^\circ\text{K}:5^\circ;V/\text{MgO})$$

which means that the incident flux is uniformly distributed at all longitudes at colatitudes from 40° to 50° and has the spectral distribution of a full radiator at 3000°K; that the flux is uniformly evaluated at all longitudes at colatitudes from 0 to 5° by a receiver having a response proportional to the standard photopic spectral luminous efficiency function; and that the measurement is made relative to a magnesium oxide reference standard.

A microdensitometer might measure $D_T(0.25;2850^\circ\text{K}:0.40;\text{S-4})_n$. The illuminating optical system has a numerical aperture of 0.25, the source is characterized by a color temperature of 2850°K, the receiver optical system has a numerical aperture of 0.40, and the receiver has the spectral sensitivity of a type S-4 photodetector.

A projector with a projection lens having an f-number of 3.5 might utilize $D_T(3.5;2850^\circ\text{K}:3.5;V)_f$.

In some studies in photography the primary interest is in the effect of geometry on modulation measurements while the spectral conditions are of

little interest. In such cases, the influx spectrum and efflux spectral sensitivity can be omitted from the notation. The general form of the notation then becomes simply $D(g:g')$.

For example, the three geometrically extreme cases in transmission densitometry mentioned in American Standards can be written as shown in Table V. It should be noted that the notation provides a means of differentiating the two types of diffuse density measurement generally regarded as equivalent.

In the proposed notation, the Callier "Q" factor is defined by the equation

$$Q = D_T(0:0)/D_T(1.00:0)_n$$

The q introduced by Altman[18] may be written as follows:

$$q = D_T(g:g')/D_T(1.00:0)_n$$

He used the symmetrical case, $g = g'$, that is, $\alpha_0 = \alpha_t$ and $\theta_0 = \theta_t = 0$. I have found this ratio useful, permitting g and g' to take on different values.

Discussion

Present day concepts in densitometry are largely based on the excellent analysis of experimental determination of photographic density by Koerner and Tuttle, thirty years ago.[19] There is, however, a basic difference between their point of view and that adopted here. They said:

> "From the definition it is apparent that, in order to measure density, it is necessary only to determine the flux incident on, and the flux transmitted by, the silver deposit. Since there are no qualifying statements attached to the above definition, it would seem quite beside the point to inquire what portion of the fluxes should be measured; one naturally supposes that all the incident and all the transmitted flux is meant. Any other interpretation would add an equivocal quality to the term "photographic density" as defining a basic physical quantity, and although it is true that Hurter and Driffield were not aware of certain properties of the photographic image, particularly the light-scattering properties, at the time the definition was formulated, it is also probably true that, had they appreciated these factors, their definition would still have remained as it now is."

No one kind of optical density is any more a "basic physical quantity" than any other. Indeed, it is the very "equivocal quality" of "photographic density" which underlies most of the problems of measurement and application. The first step in standardization was a natural one, the adoption of the proposal of Koerner and Tuttle, "that the integrating sphere used with an incident light, no ray of which departs from normal incidence by more than 10°, shall be considered as the standard optical set-up for determination of photographic density according to both letter and spirit of the Hurter and Driffield definition." There was, of course, a very great advantage to photographic science in having one standard density well defined rather than none. However, the most important justification for continuing diffuse geometry in existing standards is not that it satisfies the definition of Hurter and Driffield and is essentially unaffected by small variations in sample thickness but that it simulates the conditions of use in contact printing and in viewing transparencies against a diffuse source and that these are important applications. If the quantities to be measured are the quantities to be used, the generalized concept of density which has come into use and is here systematized, must be recognized and adopted. There is, however, some danger of "undoing standardization," in this point of view. If every kind of density were "standard" we would have no standardization other than, perhaps, a standard notation for describing the density. The reasons adduced by Koerner and Tuttle and the need for continuity are sufficient to warrant the continued use of diffuse density as the kind to be used when there is no reason to use another. Then, it seems imperative in the interest of standardization to choose and standardize representative parameters for other types of density for which there is a real need. Experience in science and industry has demonstrated a need for a projection density, a diffuse-source contact-printing density, and a reflection density which can be used to characterize photographic images on surfaces such as brushed metals, which have a considerable variation in reflection density, $D_R(5°, 45°, 0°;S:5°, 45°, 180°;S')$, as the sample is rotated about the z axis.

When the task of revising existing density standards was undertaken, it was clear that several new kinds of density would be needed but the prospect of adding to the plethora of specialized terminology, which for years has been the curse of photometry, was not inviting. The functional notation is a convenient form of expression for the values of the parameters affecting the measurement of optical modulation. It is to be hoped that the infinite flexibility of this mathematical notation will obviate the need for more words to denote the various types

TABLE V. Notation for Transmission Density with Extreme Geometrical Conditions

Name of type	Simplified notation
Regular[a]	$D_T(0:0)$
Diffuse	$D_T(0:90°)$ or $D_T(90°:0)$
Doubly diffuse	$D_T(90°:90°)$

[a] Sometimes called *specular* although many regard this term as a direct reference to a mirror and therefore inappropriate to any case other than reflection.

18. J. H. Altman, *Appl. Opt.*, **3**: 35 (1964).
19. A. M. Koerner and C. Tuttle, *J. Opt. Soc. Am.*, **27**: 241 (1937).

other than such terms as for example, "projection density" for a type of density defined in such a way that the measurement conditions simulate the conditions in a typical or "average" projector. The adoption of the present proposals should more than satisfy the requirement stated in Mees and James, "It is desirable that standard terms be adopted which would serve to differentiate the limiting true specular density from densities lying between this and totally diffuse density."[20]

New terms are proposed only because calibration and standardization activities have demonstrated a real need for some such terms. Prescriptive definitions and arbitrary systems of notation cannot be wrong, of course, but they can be good or bad. I hope these will be found useful.

20. The Theory of the Photographic Process, C. E. K. Mees and T. H. James. eds., Macmillan Co., N. Y. 3rd ed. (1966) p. 426.

The concepts of *propagance factor*, *transmittance factor*, and *reflectance factor* should appeal to the systems engineer. This "systems approach" provides solutions to a number of modulation problems which would baffle a densitometrist clinging faithfully to the definitions of transmittance and reflectance. The author did not invent these concepts but recognized them in use, though they were ill defined. Mr. L. Barbrow suggested the use of the suffix term "factor."

Acknowledgment

I am grateful to Mr. L. Barbrow and Dr. Deane B. Judd for the definitions of terms adopted or under consideration by international committees and to the members of the ASA subcommittee on densitometry, who encouraged the development of these proposals and authorized the publication of the fact that they are being considered.

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 67C, No. 3, July–September 1963

Determination of Residual Thiosulfate in Processed Film

Chester I. Pope

(May 1, 1963)

The Crabtree-Ross procedure, employing the reaction of thiosulfate with mercuric chloride to form a turbid solution is the usual test for residual thiosulfate in processed film. Qualitative tests showed that in this procedure part of the precipitate remains in the gelatin layer of the film, causing the reported concentration of thiosulfate to be too low. When residual thiosulfate was measured in the presence of image silver, the image silver contributed an additional component to the turbidity. A new test procedure was developed which also employs mercuric chloride as the precipitant but which reveals 2 to 3 times as much residual thiosulfate as does the Crabtree-Ross procedure. Residual silver causes the formation of additional turbidity which is determined as residual thiosulfate. A trace of gelatin is dissolved from the film during the thiosulfate test, the amount depending on film type. The trace of gelatin had a pronounced effect on the amount of turbidity. The addition of a known amount of gelatin to the test solution masked the effect of dissolved gelatin, making it possible to prepare one calibration curve for all types of film. A new method of measuring the residual silver in processed film was developed. A photoelectric nephelometer was used to measure the turbidity.

1. Introduction

The reaction of thiosulfate with mercuric chloride to form a precipitate is the basis for a turbidity test which has been used for the past 30 years to determine the residual thiosulfate in processed film. The test is known in the literature as the Crabtree-Ross hypo test [1].[1] In the test a sample of film is immersed in a solution of mercuric chloride and potassium bromide and the turbidity is compared visually or by a photoelectric instrument with that produced in standard solutions containing known amounts of thiosulfate.

Analytical results have been reported [3, 4] to indicate that the Crabtree-Ross test measures all of the residual thiosulfate in processed film. However, the analysis of such small quantities of thiosulfate by chemical methods is subject to considerable error. It has been suspected that some of the precipitate formed in the reaction of thiosulfate with mercuric chloride remains in the gelatin layer, giving low values for the thiosulfate content. Also the comparison or standard solutions in the Crabtree-Ross test are not always reproducible and sometimes vary as much as 50 percent for the 5 μg standard, depending on how rapidly the thiosulfate was mixed in the preparation of the solutions.

It has been shown that potassium bromide can eliminate all of the thiosulfate from film after fixation [5]. A new thiosulfate test procedure has been developed in which the residual thiosulfate is extracted from the film by treating the sample in a solution of potassium bromide. The mercuric chloride is added after removal of the film sample. In the new procedure, any residual silver in the form of a silver thiosulfate complex is removed by the potassium bromide. The silver then reacts with the halide present in the test solution to form a precipitate which adds to the turbidity formed by the thiosulfate. The effect of gelatin on the turbidity in the thiosulfate test procedure was determined.

Qualitative tests were made to show that part of the precipitate remains in the gelatin layer when the Crabtree-Ross test procedure is used. The effect of the image silver on the turbidity in the Crabtree-Ross test procedure was also investigated. A new test method was developed for determining the residual silver in processed film by measuring the transmission density of a silver image developed from the silver in the residual silver thiosulfate complex.

2. Test Methods

The solutions and experimental procedures used in this investigation for determining the residual thiosulfate in processed film are given below. The films were 35 mm. The area of the film test sample was 1 in.^{2} Round samples were punched from unperforated film with a special punch and rectangular samples were cut from perforated film using a metal stencil. Two or more specimens were tested for each determination of residual thiosulfate content. Two or more standard solutions were made in each run

[1] Figures in brackets indicate the literature references at the end of this paper.

to determine the turbidity for each thiosulfate concentration and prepare a calibration curve. The tests were made at room temperature (23 to 27 °C).

The processing solutions used in preparing the film samples were the same as those described in a previous publication [5]. Samples of 35 mm film 10 in. in length were developed 5 min, treated in a stop bath 30 sec, fixed in a fresh acid hardening bath or one containing 8 g of silver per liter, and washed in running tap water for selected times at different water temperatures to obtain a range of residual thiosulfate concentrations in the processed film. The thiosulfate-free film samples used in the tests were obtained by fixing in a fresh bath and removing the residual thiosulfate in the washing process—no thiosulfate eliminators were used. The thiosulfate concentrations are expressed in terms of anhydrous sodium thiosulfate ($Na_2S_2O_3$) in micrograms per square inch ($\mu g/in.^2$).

2.1. New Thiosulfate Test Procedure

Solution A_0

Potassium bromide	50.00 g.
Water to make	1 liter.

Solution A

Potassium bromide	50.00 g.
Water to make	980 ml.
Filter and add 0.10% gelatin solution	20 ml.

Solution B

Mercuric chloride	50.00 g.
Water to make	1 liter.

Solution A should be stored in a refrigerator and brought to room temperature before use.

The solutions were made up with distilled water and filtered through highly purified filter paper. The mercuric chloride was dissolved at room temperature. The potassium bromide, mercuric chloride, and the sodium thiosulfate ($Na_2S_2O_3 \cdot 5H_2O$) used in the thiosulfate standard solutions were of reagent quality, meeting ACS specifications. The gelatin was photographic grade.

The solutions were delivered from burets to 19 mm by 105 mm round cuvettes. The tip of the buret containing solution B was widened to permit delivery of 5 ml in about 5 or 6 sec. From the buret, 5 ml of solution A_0 or A was added to the cuvette. The test sample of film was folded twice with the emulsion side facing in and placed in solution A_0 or A for 15 min. The cuvette was agitated frequently during the immersion period. The sample was removed with a glass rod having a toe on the bottom end. The surface liquid was squeezed off as much as possible during removal. Then 5 ml of solution B was added from the buret and the cuvette was swirled immediately to insure rapid mixing of the solutions. The mixture was also stirred with the glass rod for a few seconds. The cuvette was wetted on the outside, wiped dry with soft tissue, and allowed to stand 30 min before the turbidity was measured.

Standard solutions were prepared by adding known amounts of sodium thiosulfate to 5 ml of solution A_0 or A, thoroughly mixing, adding 5 ml of solution B, and proceeding as described above in testing film. The standard solutions containing known amounts of thiosulfate were prepared as described in the American Standards Association method [2] for determining thiosulfate in processed film. No correction was made for the increase in volume of the test solution with the addition of the thiosulfate solution, since the error introduced is negligible.

2.2. Crabtree-Ross Thiosulfate Test Procedure

The following solution was made up with distilled water and filtered through highly purified filter paper:

Potassium bromide	25.00 g.
Mercuric chloride	25.00 g.
Water to make	1 liter.

The potassium bromide and mercuic chloride were of the same reagent quality as used above.

From a buret or pipet 10 ml of the above solution was added to a cuvette. The film sample was folded with the gelatin side in and immersed in the test solution for 15 min without agitation. Then the sample was removed, the solution was stirred with a glass rod, and the cuvette was allowed to stand 15 min before the turbidity was measured. The standard solutions containing known amounts of sodium thiosulfate were prepared and added as described in the American Standards Association method [2] referred to in section 2.1.

2.3. Turbidity Measurement

The turbidities of the test solutions were measured with a Coleman No. 9 nephelometer. In this instrument, a collimated beam of light from a tungsten lamp passes into the test solution, the light scattered at right angles to the exciting beam is detected by two photocells, and the output of the photocells is indicated on a meter. The nephelometer was calibrated by a Coleman turbidity standard of 74 units before each measurement. This standard did not change within a period of a year relative to other Coleman standards of 10, 18, 33, and 74 units. For readings above 130 units, the nephelometer was adjusted to make the standard of 74 units read 10 or 5 on the scale, making each unit on the scale equal to 7.4 or 14.8 units, respectively. The cuvettes were matched Coleman A grade. For the measurement, the cuvette was placed in a well containing sufficient distilled water to cover its three windows.

3. Results and Discussion

3.1. Effect of Time and Temperature on the Extraction of Thiosulfate From Processed Film

The treatment of processed microfilm and motion picture film for 15 min in solution A_0 at room temperature (23 to 27 °C) was found to be sufficient to

extract the residual thiosulfate and residual silver from the film sample. Clear (image-free) samples fixed in a fresh bath and in a bath containing 8 g of silver per liter were tested. Some of the samples were treated a second time in solution A_0 to determine if any measurable amount of thiosulfate remained. No thiosulfate was found in samples having a low residual thiosulfate concentration but 1 to 2 μg per in.^{2} was found in microfilm and 1 to 5 μg per in.^{2} was found in motion picture film having a residual thiosulfate concentration of 30 to 100 μg per in.^{2} Most of the thiosulfate was removed during a 5-min treatment in solution A_0. Extending the time to 30 min did not yield more thiosulfate than a 15-min treatment. As an example of the rapidity of the removal of the thiosulfate from the film, one sample of microfilm, containing 29 μg of thiosulfate per in.^{2} after fixation in a bath containing 8 g of silver per liter, tested 20 μg per in.^{2} after treatment for 1 min in solution A_0. Any turbidity resulting from the residual silver adds to that formed by the residual thiosulfate and is determined in the new test procedure as thiosulfate.

Tests were made to determine the amount of residual thiosulfate extracted from film by distilled water. Clear (image-free) samples were treated in 5 ml of distilled water, removed, and 0.25 g of potassium bromide added to give a 5-percent solution of potassium bromide before the addition of the 5 ml of solution B. About 30 percent of the residual thiosulfate was extracted from samples of microfilm containing 8, 14, 17, 34, 54, and 59 μg of thiosulfate per in.^{2} after fixation in a fresh bath. Only 1 to 3 μg of thiosulfate was extracted from samples of microfilm containing 14, 25, 31, 34, 44, and 46 μg of thiosulfate per in.^{2} after fixation in a bath containing 8 g of silver per liter. The bromide ion in solution A_0 decomposed the residual silver thiosulfate complex making the thiosulfate and silver soluble.

The effect of temperature on the extraction of the residual thiosulfate by solution A_0 was investigated. Tests were made at room temperature and a lower temperature on the same strip of film at the same time. After extraction, the specimens at the lower temperature were brought to room temperature before the addition of solution B. One sample of microfilm extracted at 7 to 9 °C tested 14 μg of thiosulfate per in.^{2} and at 26°, 17½ μg/in.^{2}; and a second sample at 13 °C tested 13 μg/in.^{2} of thiosulfate and at 26 °C, 15 μg/in.^{2} Both samples had been fixed in a bath containing 8 g of silver per liter. From the above data it may be inferred that a change of 3 °C in solution A_0 would cause an error of about ½ μg of thiosulfate per in.^{2}

3.2. Effect of Gelatin on the Turbidity in the Thiosulfate Test

Solution B was added to solutions of A_0 containing 5, 10, 20, and 40 μg, respectively, of sodium thiosulfate. Then the tests were repeated, except that 1-in.^{2} samples of clear (image-free) thiosulfate-free film were immersed in solution A_0 for 15 min and removed before the addition of the thiosulfate and solution B. The turbidities are shown in table 1 for different types of film and the above thiosulfate concentrations. For solutions containing more than 5 μg of thiosulfate, the turbidity of the solutions treated with film were much higher than the controls. The turbidities for microfilm specimens differed from those for motion picture film. The results suggested that the turbidity might be increased by an agent extracted from the film such as gelatin, causing a more highly dispersed precipitate. Turbid test solutions treated with film appeared bluer than those not treated. The turbidities in table 1 were reproducible to within 5 percent for each type of film. It appeared at first that a calibration curve would have to be prepared for each type of film, but this would have made the test procedure impractical.

Tests were then made to determine the effect of gelatin on the turbidity of the test solutions. Small concentrations of gelatin were added to solution A_0 containing known concentrations of sodium thiosulfate after which the required amount of solution B was added. It was found that traces of gelatin significantly increased the turbidity in the thiosulfate test solution. As little as 3 parts of gelatin in a billion caused a detectable change of the turbidity in the thiosulfate test solution!

TABLE 1. *Effect of different films on the turbidity in the thiosulfate test when solution A_0 was treated with 1 square inch of image-free thiosulfate-free film*

Thiosulfate added after removal of film	No film treatment. Data taken from calibration curve	Turbidimetric units				
		Microfilm			Motion picture film	
		A	B	C	A	B
Micrograms						
5	11	14	13	13	9	12
10	42	96	86	75	114	123
20	146	333	325	296	245	263
40	339	632	610	614	488	562

In order to evaluate the effect of gelatin concentration on turbidity, a 0.10-percent gelatin solution was made up as follows: About 10 g of photographic grade gelatin were heated in 600 ml of distilled water at 50 to 60 °C for 1 hr with frequent stirring. About 200 ml of the warm solution were filtered through 2 layers of highly purified filter paper. The solid content of the filtrate was determined by pipeting 10 ml into a tared 20 ml beaker, heating in an oven at 100±2 °C for 24 hr, and weighing the beakers and residue. From the weight of the solid residue in the 10 ml aliquot the volume of the filtrate containing 1 g of gelatin was calculated and this volume of filtrate was diluted to 1 liter with distilled water to obtain the 0.10-percent gelatin solution. A commercial food gelatin gave the same results as the above photographic gelatin, but it developed a fungus growth after storage for 8 weeks in a refriger-

ator, while the photographic grade has remained stable for 8 months in a refrigerator.

The effect of gelatin concentration on the turbidity in the thiosulfate test was determined for 10, 20, and 40 μg of sodium thiosulfate by adding different amounts of gelatin to solution A_0. Solution A_0 was made up to 1000 ml less the volume of gelatin solution to be added, filtered, and the gelatin solution added. The gelatin solution was added after filtration to avoid any effect of the filter paper on the concentration of the gelatin. In the tests, the thiosulfate was added to solution A_0 containing gelatin, followed by solution B. In figure 1 the gelatin concentrations are plotted against turbidimetric units for the 3 thiosulfate concentrations. These curves show that very low concentrations of gelatin in the test solution (1 to 10 μg per 10 ml of test solution) have a large effect on measured turbidity in the thiosulfate test. Between 50 and 200 μg per 10 ml of test solution, however, the change in measured turbidity is very slight. The data in figure 1 suggest that the effects of different amounts of gelatin dissolved from different types of film might be eliminated by increasing the gelatin concentration in the test solution.

A concentration of 100 μg of gelatin in 5 ml of the extracting solution A_0 was selected for the new thiosulfate test procedure and is hereafter designated solution A. Tests were then made to determine if treatment of different types of clear (image-free) thiosulfate-free film in solution A would give the same turbidity for a given thiosulfate concentration. Samples of 1 in.^{2} of film were immersed in solution A for 15 min and then removed before the addition of thiosulfate and solution B. Samples of three different microfilms and two motion picture films were tested at 5 thiosulfate concentrations (5, 10, 20, 40, and 80 μg per 10 ml of test solution). The measured turbidities for each concentration agreed within experimental error with the corresponding turbidities obtained for each of the above thiosulfate concentrations when solution A was not treated with a film. These results demonstrated conclusively that the effect on measured turbidity of gelatin dissolved from the film had been eliminated.

The nephelometer calibration curve shown in figure 2 was prepared by plotting measured turbidities as a function of thiosulfate concentration. The plotted values were an average of five or more determinations. In this new test procedure the rapid mixing of solutions, A and B, and the added gelatin formed a reproducible turbidity. The curve in figure 2 shows a sharp change between 6 and 8 μg of thiosulfate, but is linear from 9 to 100 μg. Figure 3 is an enlarged scale of a portion of figure 2 showing more clearly the sharp change in the calibration curve for the lower thiosulfate concentrations, those of chief concern in microfilm practice. In the Crabtree-Ross test procedure the turbidity readings for the standard solutions were not always reproducible and sometimes varied as much as 50 percent for the 5 μg standard. The rate of mixing of the thiosulfate in the test solution appears to affect the particle size of the precipitate.

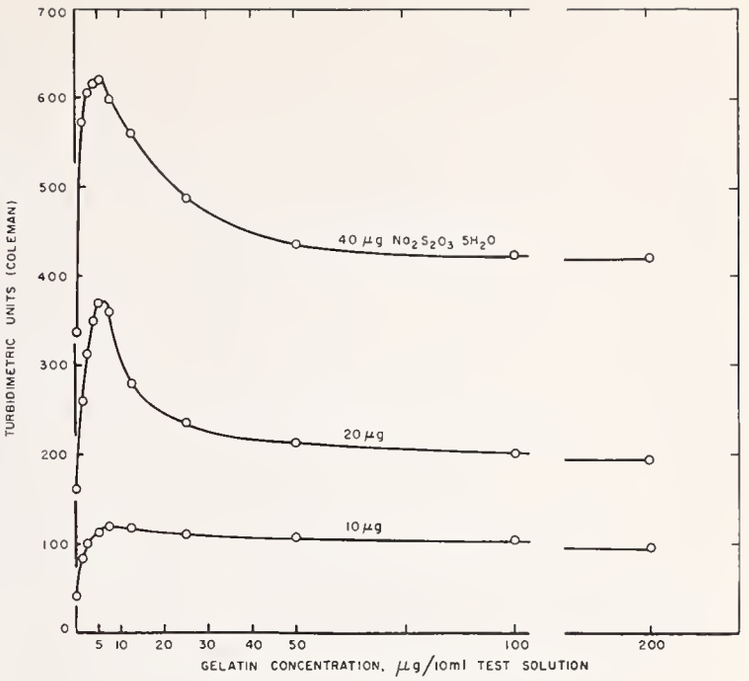


FIGURE 1. *Effect of gelatin concentration on the turbidity formed by the reaction of mercuric chloride with thiosulfate after the addition of 5 ml of solution* B *to 5 ml of solution* A_0 *for 3 thiosulfate concentrations.*

The concentration of sodium thiosulfate is shown on the curves.

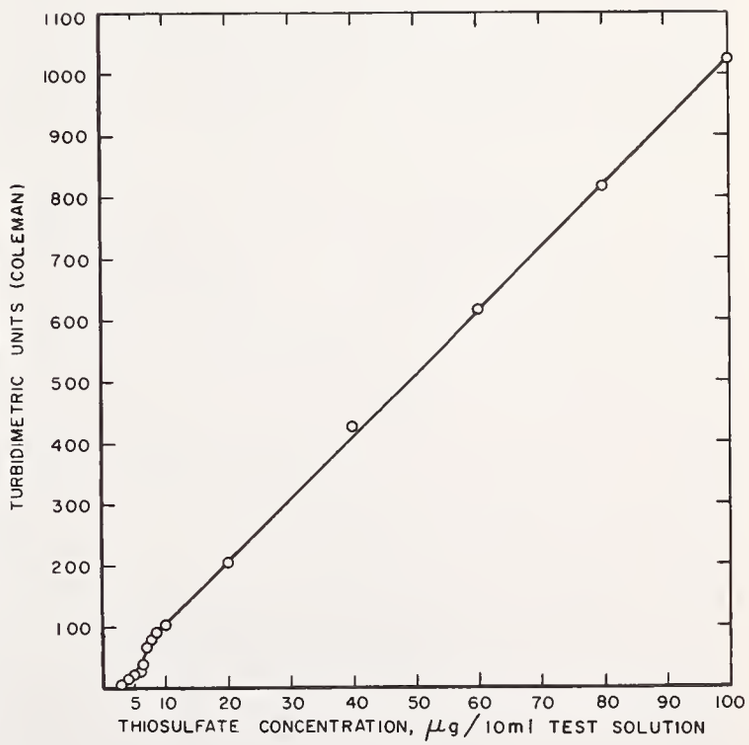


FIGURE 2. *Nephelometer calibration curve showing the turbidity formed by the reaction of mercuric chloride with thiosulfate after the addition of 5 ml of solution* B *to 5 ml of solution* A.

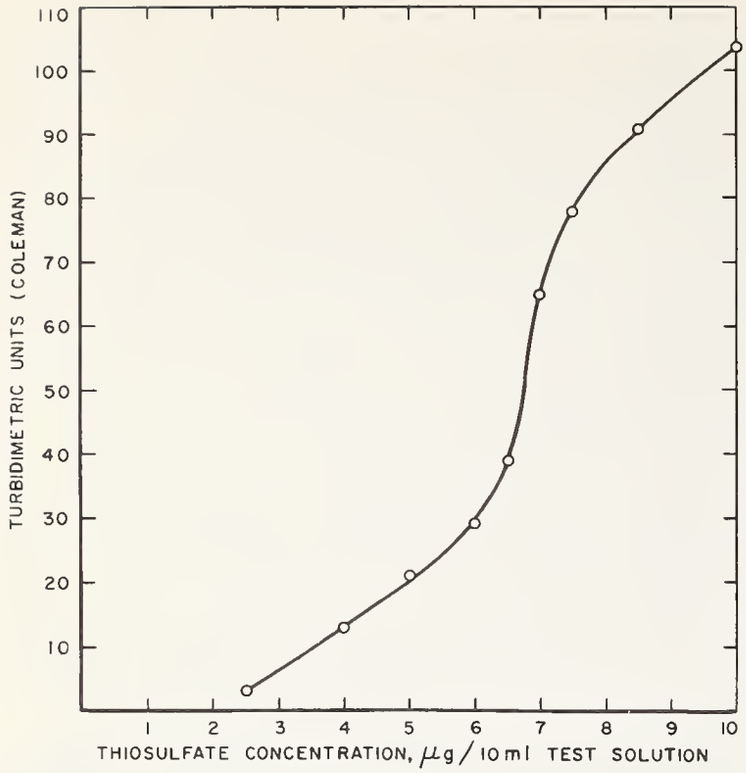


FIGURE 3. *An enlarged scale of the calibration curve in figure 2 for the low sodium thiosulfate concentrations.*

A trace of thiosulfate is held by the film in the thiosulfate test. It was found that the 1-in.^{2} microfilm samples removed 0.03 to 0.05 ml of solution A. This loss was determined by weighing, in a tared glass stoppered bottle, three 1-in.^{2} samples, before and after treatment in solution A. The loss of this amount of solution does not introduce any significant error, especially for the low residual thiosulfate concentrations. Tests were made by treating solutions of A containing 10 to 70 μg of thiosulfate with clear (image-free) thiosulfate-free microfilm to determine the error caused by removing the film sample from solution A. There was a loss of about 1μg of thiosulfate per in.^{2} regardless of the thiosulfate concentration. If the film sample was not removed before the addition of solution B, the sample sometimes interfered with the formation of the precipitate giving low turbidity values. Evidently the film absorbed the thiosulfate in the emulsion or on the film base. A correction of 1 μg might be applied. This correction is not significant for high thiosulfate concentrations, but it may be desirable to add it at concentrations below 20 μg per in.^{2} for accurate determination of residual thiosulfate.

3.3. Comparison of the New Test Procedure With the Crabtree-Ross Test Procedure

Unexposed strips of 35-mm film, cut 10 in. in length, were developed, treated in a stop bath, fixed, and washed at different water temperatures and for different times to obtain image-free samples with a range of residual thiosulfate concentrations. The film strips were stored about 2 to 3 weeks before testing. Analysis of the residual thiosulfate content was made by the two procedures at the same time on the same strip of film in each run. A calibration curve was prepared for the Crabtree-Ross test procedure and calibration curves for the new procedure are shown in figures 2 and 3. Samples of 1 in.^{2} of film were cut or punched from the strip, alternate samples 1 and 3 being used in one procedure and 2 and 4 in the other. The duplicate in both procedures generally gave the same residual thiosulfate content, occasionally differing by ½ to 1 μg in the low thiosulfate concentrations and 1 to 2 in the high. The average residual thiosulfate concentrations of the duplicate samples for the two procedures are given in table 2 for microfilm and motion picture film after fixation in a fresh bath and a bath containing 8 g of silver per liter. The thiosulfate concentrations for the new procedure were not corrected for the 1 μg of thiosulfate per in.^{2} which is normally absorbed by the film. As shown in table 2, the new test procedure measured 2 to 3 times as much residual thiosulfate in processed film as the Crabtree-Ross test procedure.

TABLE 2. *Comporison of residual thiosulfate concentrations obtained from the new thiosulfate test and the Crabtree-Ross thiosulfate test for the same sample of image-free film*

Thiosulfate, micrograms per square inch

Microfilm			Motion picture film		
Film brand	New procedure	Crabtree-Ross procedure	Film brand	New procedure	Crabtree-Ross procedure
	Fixed in a fresh bath				
A	2	0	A	28	11
A	7	4½	A	87	28
A	19	6½	A	110	32
A	22	12	B	16	7
A	55	27	B	50	18
B	5	2	B	78	28
B	7½	4			
B	19	10½			
B	43	19			
B	90	42			
	Fixed in a bath+8 g of silver per liter				
A	9½	3	A	64	18
A	25	7	A	88	23
A	38	12	B	6½	4½
A	62	18	B	52	19
B	8	2½	B	78	28
B	39	10			
B	70	18			

The lower residual thiosulfate concentrations measured by the Crabtree-Ross procedure may be attributed to the precipitation of some of the reaction product from the thiosulfate and mercuric chloride within the gelatin layer of the film. Thus, only a part of the thiosulfate is extracted to form a precipitate in the test solution. The presence of a precipitate in the gelatin layer after treatment in the Crabtree-Ross test solution was determined qualitatively by dissolving the gelatin and observing the turbidity of the solution. One-in.^{2} samples from clear (image-free) microfilm containing 0 and 50 to 70 μg of residual thiosulfate were treated in the Crabtree-Ross test solution for 15 min, washed 10

min, and treated in 10 ml of a ½-percent solution of pancreatic enzyme at about 55 °C for 30 min to dissolve most of the gelatin. On visual inspection, there was considerably more turbidity in the enzyme solution for films containing residual thiosulfate than the one containing none. The precipitate in the enzyme solutions for films containing thiosulfate appeared bluish, but the blank was brownish gray. In other tests on microfilm having a high residual thiosulfate concentration, the turbidity remaining after solution of the gelatin roughly corresponded to the difference of residual thiosulfate content measured by the new test procedure and the Crabtree-Ross test procedure. In another test, 1-in. samples of clear (image-free) microfilm containing 17, 39, and 51 μg of thiosulfate were treated in 10 ml of the Crabtree-Ross test solution for 15 min, washed 10 min, and heated in 10 ml of a 5-percent solution of citric acid at 70 to 80 °C for about 1 hr to dissolve the emulsion. There was considerably more turbidity in the solutions for films containing residual thiosulfate than for the one containing none. The degree of turbidity was in the same order as the residual thiosulfate concentrations of the films whether estimated visually or measured with the nephelometer. Again the precipitate in the solutions for the films containing residual thiosulfate was bluish, but the blank was brownish to gray. When the above solutions were made alkaline with concentrated ammonium hydroxide, most of the turbidity in the sample containing no thiosulfate was eliminated while the solutions for films containing thiosulfate became quite bluish and the turbidities were in the same order as the residual thiosulfate contents when estimated visually or measured with nephelometer.

The presence of a precipitate in the gelatin layer was also shown by measuring the transmission density of the film sample after treatment in the Crabtree-Ross test solution. Clear (image-free) microfilm samples containing 0, 60, 70, 103, and 150 μg, of residual thiosulfate per in.^{2} were treated for 15 min in the Crabtree-Ross test solution, washed 15 min, dried, and the transmission density of the samples measured. The untreated microfilm had a density of 0.07. The control sample having 0 μg of thiosulfate per in.^{2} had a density of 0.08; 60 μg, 0.09; 70 μg, 0.10; 103 μg, 0.11; and 150 μg 0.13. The above data showed that a precipitate was present in the gelatin layer and its density increased from 0.01 to 0.05 above that of the control.

The above results explain why the Crabtree-Ross test procedure gave low values for the residual thiosulfate content in processed film. The gelatin effect on the turbidity was not investigated for the Crabtree-Ross test procedure, but it is probably small since the test solution is acid.

3.4. Effect of the Image Silver on the Crabtree-Ross Test Procedure

Residual thiosulfate contents determined by the new test procedure and the Crabtree-Ross test procedure in the silver image of processed film did not show the same difference as those determined for clear (image-free) film. Flash-exposed strips of 35-mm film, cut 10 in. in length, were developed, treated in a stop bath, fixed, and washed for different times to obtain samples with a range of residual thiosulfate concentrations. The thiosulfate tests were made within 4 hr after processing. The samples were tested by the two test procedures as described in section 3.3. The comparative results for a microfilm fixed in a fresh bath and a bath containing 8 g of silver per liter are given in table 3 for an image density of 5. The residual thiosulfate contents from the two test procedures were about the same for thiosulfate concentrations of 20 μg per in.^{2} and higher, but the Crabtree-Ross test procedure gave higher thiosulfate contents in the range of 10 μg of thiosulfate per in.^{2} Other tests on microfilm for image densities of 0.02, 0.72, 1.62, and 2.48 showed that the thiosulfate contents from the Crabtree-Ross procedure were about one-half the value obtained from the new procedure at a density of 0.02, but approached those from the new procedure as the image density increased. Thiosulfate contents indicated by both procedures became about equal at image densities of 4 to 5. The above results show that the Crabtree-Ross test procedure gives higher residual thiosulfate contents in the presence of image silver than in clear (image-free) film and this difference may result from the addition of silver to the test solution forming a turbidity of silver halide. In the Crabtree-Ross test procedure the silver in the image is bleached by the mercuric chloride, forming silver halide in the film. Any residual thiosulfate present could form a soluble silver thiosulfate complex which migrates into the test solution where it decomposes to form an additional turbidity of silver halide and is measured along with that formed by the reaction of thiosulfate with mercuric chloride.

TABLE 3. *Comparison of residual thiosulfate concentrations obtained from the new thiosulfate test and the Crabtree-Ross thiosulfate test for microfim having a density of 5.*

Fixed in a fresh bath		Fixed in a bath +8 g of silver per liter	
New procedure	Crabtree-Ross procedure	New procedure	Crabtree-Ross procedure
Thiosulfate, μg/in.^{2}			
9	14	11	16
12	16	20	21
18	21	31	29
24	23	57	43

Different amounts of silver as silver nitrate were added to the Crabtree-Ross test solution and the turbidities measured. The standard silver solutions were made up and added by the same method used for the standard thiosulfate solutions. Figure 4 shows the calibration curve obtained by plotting the turbidimetric units against the silver concentration in the Crabtree-Ross test solution. The calibration curve showed that the turbidity of the silver halide varied linearly with the silver concentration and also

518-743 O - 73 - 58

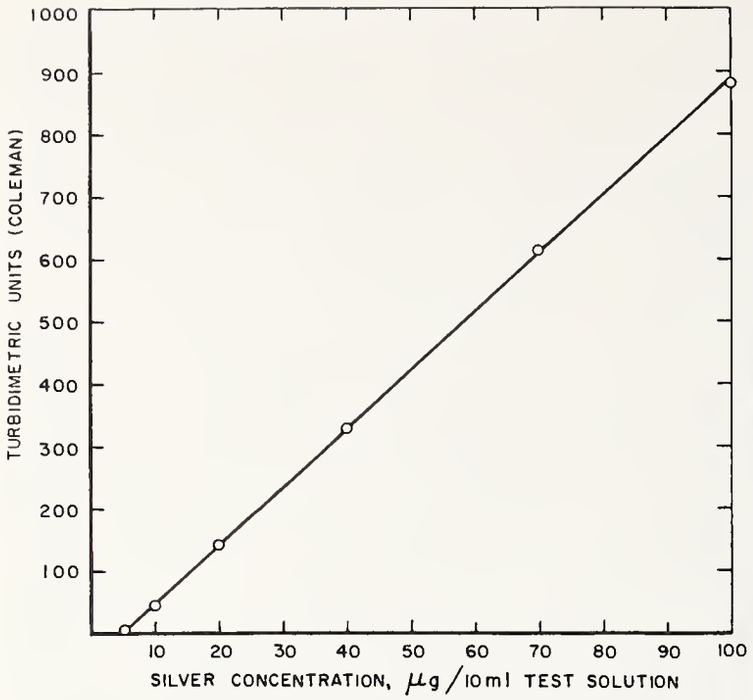


FIGURE 4. *Calibration curve showing the turbidity formed by silver added as silver nitrate to 10 ml of the Crabtree-Ross thiosulfate test solution.*

that a small amount of silver in the test solution would add appreciably to the turbidity in the thiosulfate test. For example, 20 μg of silver gave a turbidity reading of 142 as compared to 204 for 20 μg of thiosulfate shown in figure 2.

The effect of image silver on the turbidity in the Crabtree-Ross test procedure was further investigated by adding a known amount of thiosulfate to processed microfilm. A 0.05 ml solution containing 20 μg of sodium thiosulfate was pipeted onto 1-in.^{2} samples of thiosulfate-free film punched from clear (image-free) strips and strips having an image density of 5, evaporated to dryness in an oven at 50 °C, and the residual thiosulfate determined at the same time by the two test procedures. Duplicate samples were used for each test in the clear and image areas. The amount of residual thiosulfate found by the new test procedure in exposed film (image density 5) was about the same as that found in clear film. In one run 19 μg of sodium thiosulfate were measured in the clear film and 17 μg in the exposed film; in a second run, there was 17 μg in the clear film and 16 μg in the exposed film. The above results showed that the residual thiosulfate was accurately determined in the presence of image silver if allowance is made for some loss of thiosulfate due to its tendency to react with silver to form silver sulfide. A trace of thiosulfate may also be oxidized by exposure to air during oven drying.

The amount of residual thiosulfate found by the Crabtree-Ross procedure in exposed film (image density 5) was more than that found in an unexposed film. In one run, 11 μg of sodium thiosulfate were measured in the clear film and 17 μg in the exposed film. In a second run, there were 12 μg in the clear film and 19 μg in the exposed film. The results show that the exposed film tested higher in residual thiosulfate than the clear film, indicating that some of the silver in the bleached image had migrated into the test solution. The amount of thiosulfate found by the Crabtree-Ross test procedure in the clear film was less than was found by the new test procedure, due to the retention of some of the precipitate in the gelatin layer, as shown in section 3.3.

3.5. Test for Residual Silver in Processed Film

Residual silver in processed film is present as a silver thiosulfate complex. Appreciable amounts of residual silver may cause fading or yellowing of the image due to the formation of silver sulfide. A large concentration of residual silver in the film would indicate that the fixing bath contained too much silver for adequate removal of the silver thiosulfate complex during a normal washing procedure. A fresh fixing bath leaves no measurable residual silver in the film, but as silver accumulates in the fixing bath the tendency for the silver thiosulfate complex to remain in the film increases. The silver thiosulfate complex is more difficult to remove in the washing process than the thiosulfate. As the fixing bath approaches exhaustion the residual silver in the film becomes high and some of it remains after prolonged washing.

The test for residual silver in processed film [6] now in use is only qualitative. In this test the film is treated with a 0.2-percent solution of fused sodium sulfide and a yellow stain is produced if sufficient residual silver is present. Tests showed that at high residual silver concentrations a stain density of only 0.02 was produced in microfilm and 0.04 in motion picture film. A more sensitive and quantitative test was needed to indicate the amount of residual silver present in processed film. It has been shown [5] that a 0.1-percent solution of potassium iodide will react with the residual silver thiosulfate complex in processed film to precipitate silver iodide in the film. An investigation was made to determine if the silver in the silver thiosulfate complex could be converted to silver in the film and the amount of residual silver indicated by measuring the transmission density of the silver.

Samples of unexposed microfilm and motion picture film were fixed in a bath containing 8 g of silver per liter and washed for different times to obtain a range of residual thiosulfate concentrations accompanied by high residual silver concentration. It was found that the residual silver in the silver thiosulfate complex could be converted to a silver image by treating it in a developer containing 2 g of potassium iodide per liter, while exposing it to light. The optimum concentration of potassium iodide was 1½ to 3 g per liter of developer solution. If there was less than 1½ g of potassium iodide per liter, not all of the residual silver was converted to image silver and if above 3 g, some of the silver iodide passed into solution before development. For the test procedure, 2 g of potassium iodide per liter of developer solution was used. The following developers were used in the experimental work and gave the same results:

Developer A

Amidol, 2,4-diaminophenol dihydrochloride	5.0 g.
Sodium sulfite, anhydrous	30.0 g.
Potassium iodide	2.0 g.
Water to make	1 liter.

Developer B

Monomethyl *p*-aminophenol sulfate	3.1 g.
Sodium sulfite, anhydrous	45.0 g.
Hydroquinone	12.0 g.
Sodium carbonate, anhydrous	67.5 g.
Potassium iodide	2.0 g.
Water to make	1 liter.

The developer should be fresh and developer A should be made up just before use. The film sample was treated in the developer solution for 20 min with occasional stirring while exposed to a 100-w tungsten lamp located 6 in. from the surface of the solution, treated 1 min in a stop bath containing 12 ml of glacial acetic acid per liter, and washed 15 min. The densities of the silver images obtained for the different thiosulfate concentrations were very reproducible and correlated well with the thiosulfate concentrations. Silver in the image was identified qualitatively by dissolving the image in nitric acid and obtaining a precipitate of silver chloride after the addition of sodium chloride. Film samples of the silver image were immersed in water for 5 min and, after removal of the surface water, they were bleached over concentrated hydrochloric acid, dried, and redeveloped with light in developer A. The same density could be redeveloped, indicating that the image was silver.

One set each of unexposed 35 mm samples, cut 10 in. in length, from a microfilm and a motion picture film, respectively, were fixed in a bath containing 8 g of silver per liter and washed at 25° C for different times to obtain a range of residual thiosulfate concentrations. Then samples from strips having different residual thiosulfate concentrations were developed together in Developer A, one set being run for microfilm and one for motion picture film. The transmission densities of the developed samples were measured and the density of the film base subtracted from the density readings. The density of the silver image for each sample was plotted against its respective residual thiosulfate concentration and curves for the microfilm and the motion picture film are shown in figure 5. The data in the curves show that the silver density developed from the residual silver is proportional to the residual thiosulfate concentration for the lower thiosulfate concentrations and becomes nearly constant at the higher concentrations. The test method may be used to study the exhaustion of a fixing bath under controlled conditions, since the appearance of appreciable concentrations of residual thiosulfate will signal difficulty in removing the silver thiosulfate complex in the washing process.

In a microfilm with a residual thiosulfate concentration of 130 μg per in.^{2}, a density of 0.16 was obtained, while in the sodium sulfide test [6] a stain density of only 0.02 was found. In a motion picture film for a residual thiosulfate concentration of 154 μg per in.^{2}, a density of 0.27 was obtained while in the sodium sulfide test a stain density of only 0.04 was found. Clearly, the new test method is much more sensitive than the sodium sulfide test.

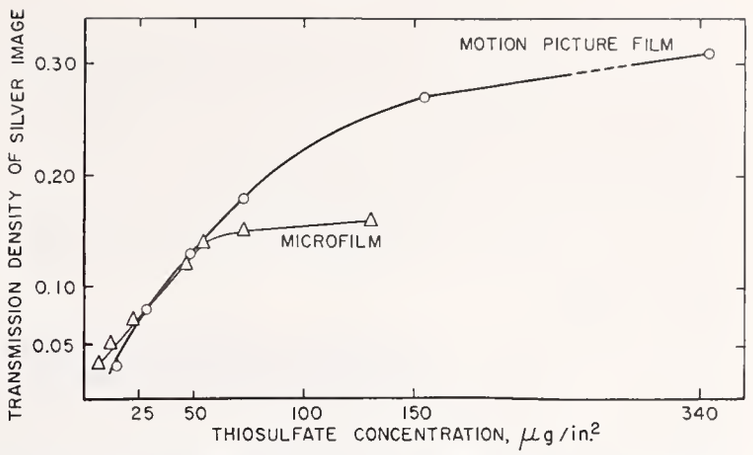


FIGURE 5. *Density of silver image formed from the silver in the silver thiosulfate complex versus the thiosulfate concentration.*

Tests were made for residual silver on microfilm and motion picture film fixed in a fresh bath by the new test method and no residual silver could be detected in microfilm samples containing up to 240 μg of residual thiosulfate per square inch and motion picture samples containing up to 500 μg per in.^{2}

4. Summary

A new test procedure was developed for the determination of residual thiosulfate in processed film. This procedure indicates that the residual thiosulfate present in clear (image-free) processed film is 2 to 3 times that indicated by the Crabtree-Ross test procedure. In the course of development of the new test procedure, it was found that a trace of gelatin dissolved from the film sample produced a considerable change in the measured turbidity. The amount of gelatin dissolved depended on the type of film being tested. A separate nephelometer calibration curve might have been prepared for each type of film, but such a procedure would be impractical. Instead the effect of gelatin concentration in the test solution was investigated. The data showed that the addition of 100 μg of gelatin to the test solution would eliminate the effect of gelatin dissolved from the film, thus making it possible to prepare a single calibration curve for all types of film. For accurate results at low thiosulfate concentrations a correction of 1 μg per in.^{2} should be added to the value obtained turbidimetrically.

In the new procedure the thiosulfate was extracted from 1 in.^{2} of film by treatment in 5 ml of a 5-percent solution of potassium bromide solution containing 100 μg of gelatin, and 5 ml of a 5-percent solution of mercuric chloride was added after removal of the film sample. A nephelometer was used to measure the turbidities. A calibration curve was prepared at 25 °C$\pm$2 °C. The temperature of the solutions

518-743 O - 73 - 59

during the thiosulfate test should be the same as that at which the calibration curve was prepared.

Qualitative tests showed that some of the precipitate in the Crabtree-Ross test procedure remained in the film during the test, causing low values of residual thiosulfate concentration to be reported.

When a silver image containing residual thiosulfate was tested by the Crabtree-Ross test procedure, there was evidence that some of the image silver migrated into the test solution, adding a turbidity of silver halide.

The new test procedure may be used for determining the residual thiosulfate in film containing image silver. However, it has been shown [5] that the test for residual thiosulfate in film should be made with an image-free area of film.

A new test method was developed for determining the residual silver in processed film. The silver in the silver thiosulfate complex was developed to a silver image and its transmission density measured.

(Paper 67C3–134)

5. References

[1] J. I. Crabtree and J. F. Ross, A method of testing for the presence of sodium thiosulfate in motion picture films, J. Soc. Motion Picture Engrs. **14,** 419–426 (1930).

[2] American Standard Method for Determining the Thiosulfate Content of Processed Black-and-White Photographic Films and Plates, PH4.8–1958 (American Standards Association, Inc., 10 East 40th St., New York 16, N.Y.).

[3] J. I. Crabtree, G. T. Eaton, and L. E. Muehler, The removal of hypo and silver salts from photographic materials as affected by the composition of the processing solutions, J. Soc. Motion Picture and Television Engrs. **41,** 14 (1943).

[4] E. P. Przybylowicz, C. W. Zuehlke, and A. E. Ballard, An evaluation of the Crabtree-Ross procedure for residual hypo in processed film, Photo. Science and Engr. **2,** No. 3, 148–153 (1958).

[5] C. I. Pope, Stability of residual thiosulfate in processed microfilm, Journal of Research National Bureau of Standards, Vol. **67C** (Eng. and Instr.), No. 1, 15–24 (1963).

[6] American Standard Method for Indicating the Stability of the Images of Processed Black-and-White Films, Plates, and Papers, PH4.12–1954 (American Standards Association, Inc., 10 East 40th St., New York 16, N.Y.).

UNITED STATES DEPARTMENT OF COMMERCE • Luther H. Hodges, *Secretary*
NATIONAL BUREAU OF STANDARDS • A. V. Astin, *Director*

Inspection of Processed Photographic Record Films for Aging Blemishes

C. S. McCamy

National Bureau of Standards Handbook 96

Issued January 24, 1964

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington D.C. 20402 - Price 25 cents

899

Foreword

Over the years the National Bureau of Standards has conducted considerable research on the permanence of record materials, including both paper and photographic film. Recently microscopic blemishes have been found on microfilms; apparently they had developed after the film was put in storage. Practically no information loss has been observed, but any potential threat to the permanence of Government records is a matter of concern not only to the National Bureau of Standards but to records officers throughout the Government. There is much to be learned from these blemishes and they are being studied in many laboratories.

This publication contains a description of the blemishes, illustrated with photomicrographs in color. It also describes methods used to inspect, sample, and report on the blemishes. It is felt that early publication of this information will serve to unify methods and terminology and thus aid in coordinating the various independent studies of the problem. This handbook should also provide the guidance needed for large-scale inspections of Government records.

A. V. Astin, *Director.*

Contents

Inspection of Processed Photographic Record Films for Aging Blemishes

C. S. McCamy

Inspections of microfilms have recently revealed blemishes which apparently developed 2 to 20 years after the films were processed. Most of the blemishes are small spots, usually reddish or yellowish in color, ranging from about 15 to 150 microns across. These blemishes have been classified on the basis of size, shape, color, and character. This publication describes and gives colored illustrations of the various types, describes the method of observing them, and recommends sampling procedures for microfilm inspectors. The cause, the exact mechanism of formation of the various types, and generally accepted preventive measures are not as yet known. This publication is intended to promote uniform terminology, inspection, and reporting.

1. Introduction

Because of its ability to record a large amount of information in a small space and its stability relative to most record materials, photographic film is one of the most common media for the storage of records. A considerable percentage of United States Government records are stored on microfilm. Even though such films have been prepared by the best known techniques and stored in the best known manner, the only assurance that such records are being well maintained is a systematic program of careful inspection. Improper processing or storage conditions have long been known to promote general fading and various kinds of biological attack. Inspections of large numbers of rolls of microfilm recently have revealed several kinds of blemishes not heretofore noted. It appears that they have developed from 2 to 20 years after the film was processed and placed in storage. The blemishes have been classified on the basis of size, shape, color, and character. The purpose of this publication is to name and describe the various blemishes, to provide illustrations of them, to describe the method of observing them, and to recommend sampling procedures for microfilm inspectors. Although the chemistry and physics of these blemishes are under investigation at the National Bureau of Standards and in research laboratories in the photographic industry,[1] the causes of the blemishes, the exact mechanisms of their formation, and generally accepted preventive measures have not yet been established. A vital part of the research program is accurate reporting of blemishes, so that the incidence of the various types can be correlated with processing and storage variables. In this stage of the investigation, it cannot be expected that the types of blemishes described necessarily exhaust the possibilities, nor can the types be regarded as necessarily mutually exclusive.

[1] Henn, R. W. and Wiest, D. G. "Microscopic Spots in Processed Microfilm: Their Nature and Prevention." Phot. Sci. Eng. 7, 253, (1963).

2. Classification of Aging Blemishes

Type 1

Type 1 aging blemishes are circular spots, usually 50 to 150 microns in diameter, with sharp boundaries. Concentric light and dark rings are common. Spots normally occur as reduced density in high-density "background" areas, but may make incursions into low-density lines or characters. They are usually brown, orange, reddish, or yellow in color. It is common to find many spots about the same size on a sample. The circular boundaries of two or more spots may intersect. The spots are often seen centered on scratches in the emulsion, sometimes closely packed like beads on a string. They sometimes occur in higher concentration at steep density gradients between high- and low-density regions, sometimes being so closely packed as to give the appearance of a continuous band. By reflected light, the spots may display a silvery sheen.

Type 2

Type 2 are defects in the light lines forming the characters themselves, rather than in the high-density "background." The lines making up the characters become lighter, yellowish, and broader. The boundaries of the defect are sharp.

Type 3

Type 3 are very small spots, about 10 to 15 microns across. When they occur, there is usually a large number per unit of area. They usually range from pale yellow to orange in color. Their boundaries are sharp. By reflected light, the spots may display a silvery sheen.

Type 4

Type 4 blemishes are spots of less regular shape than type 1 but about the same size or a little larger, usually lighter in color, and less sharply bounded. A circular central "nucleus" is common.

These spots occur in high-density "background" areas. They do *not* make incursions on low-density characters; on the contrary, their shapes may be very distorted to accommodate to the spaces between characters or parts of characters.

Type 5

Type 5 is a reddening of the dark background in the immediate neighborhood of lighter characters. The boundaries of the discoloration are very diffuse. The shape is not regular, being determined by the shape of the characters or line of characters with which the discoloration is associated.

Type 6

Type 6 is a reddish, orange, or yellow spot of reduced density, lightest at the center and gradually blending into the surrounding background. An irregular opaque or crystalline particle is commonly observed on the surface of the film at the center of the spot. The sizes of the spots may vary considerably, even within a small region.

3. Sampling Method

The first task is to divide the films into separate homogeneous groups, called "strata" in statistical sampling theory. These groups must be distinguished by differences which may be chemically or physically significant. Examples of such differences are differences in processing, differences in storage conditions, differences in film size, and differences in film type. Differences in textual content, i.e., differences in the kinds of documents photographed, may coincide with or assist in determining significant differences but are not, in themselves, of primary value in establishing separate strata. Careful attention to defining the strata to be sampled may eventually lead to the discovery of the factors upon which the formation of these blemishes depends. Once the strata have been well defined, the number of rolls in each should be ascertained.

In some cases, it may be desirable to inspect every roll of film in a stratum. This procedure may be feasible if the group is not too large, but in many cases the number of rolls makes total inspection impractical. In such cases a sample is taken from the stratum.

To be representative of the group, the sample must contain a sufficient number of rolls. The sample should be 1/1000 of the stratum but not less than 100 rolls. The whole group should be inspected if there are less than 100 rolls. This minimum sample size is based upon the desired accuracy of the estimate of total incidence as derived from the sample. The accuracy of the determination does not increase in proportion to the increase in the size of sample. Thus, additional expenditure of time and effort beyond the minimum is hardly justified.

It is necessary that sampling procedures be established which will assure that all parts of the stratum are represented. For example, if 100 rolls are to be selected from 300 drawers, 1 roll might be taken from every third drawer. Each roll should be taken from the center or some other easily identifiable position in the drawer. The establishment of the sampling pattern in advance prevents an inspector from consciously or unconsciously introducing a bias in the sampling by selecting rolls which have some distinguishing characteristic. On each roll of the sample, the whole leader, about 6 ft of the information frames adjoining the leader, and about 6 ft near the center of the roll should be inspected.

4. Inspection Methods and Equipment

Although some aging blemishes are visible to the unaided eye, it has been found necessary to use a microscope to inspect films for their presence. The very versatile instruments recently introduced by the optical industry have greatly facilitated microfilm inspections. An instrument which has been very useful is a low-power stereo microscope with continuous magnification adjustment. The magnification is continuously variable from 7× to 30×. The components of the microscope can be arranged so that the column supporting the optics is away from the inspector. This permits the inspector to place the center of a long roll of film on the stage from the side of the instrument facing him.

The film on the microscope stage is illuminated by either of two light sources which can be turned on and off independently. A small fluorescent light source with a flat diffusing surface is used to illuminate the stage from below when the film is to be viewed by transmitted light. The clarity of the image may be improved by placing a circular paper aperture directly under the stage glass, to prevent light from getting into the microscope from angles outside the field of view. The aperture should be just slightly larger than the field of view at the lowest magnification. The second illuminator is a small spotlight which is used to illuminate the stage from above, usually at a nearly grazing angle. This source is used to view the surface of the film by reflected light and is particularly useful for detecting objects or materials on the film surface or defects in the surface.

For convenience, the stage of the microscope may be equipped with guides to hold the film in viewing position but permit the film to be advanced and rewound easily. Figures 9 and 13 illustrate how

such guides can be made. The guide has two plastic rails attached by screws to a metal base. The plastic rails have grooves in which the edges of the film slide. If the screw holes through one of the rails are elongated, the spacing between the rails can be adjusted. Then the dimensions are not critical and the machining does not have to be precise. The dimensions for 16 and 35 mm film guides are given in table 1. Dimensions *b* and *c* are based upon the assumption that the grooves in the two rails are of equal depth.

Rewinds should be placed at either side of the microscope so that a roll of film can be placed on one and be taken up on the other. To avoid damaging the film, the inspector should wear cotton gloves (available from suppliers of photographic materials), the film should be handled by its edges only, inspecting equipment should be designed to avoid scratching the film, and the equipment must be kept clean. To avoid contaminating the film with chemical dust, film should be inspected in an area well removed from any area where film is processed or chemicals are handled.

TABLE 1. *Dimensions for film guide*

Dimensions (refer to fig. 13)	Nominal film width	
	16 mm	35 mm
a	65 ± 10 mm	65 ± 10 mm
b	16.5 ± 0.3 mm	35.5 ± 0.3 mm
c	14.0 ± 0.3 mm	31.0 ± 0.3 mm
d	0.7 ± 0.3 mm	0.7 ± 0.3 mm
e	2.5 ± 1 mm	2.5 ± 1 mm
f	6.5 ± 2 mm	6.5 ± 2 mm
g	6.5 ± 2 mm	6.5 ± 2 mm
h	45 ± 5 mm	45 ± 5 mm
i	75 ± 10 mm	75 ± 10 mm

TABLE 2. *Severity classification of spot blemishes*

Severity	Concentration in spots per cm ^{2}
1	less than 1
2	1 to 8
3	8 to 63
4	63 to 500
5	over 500

5. Reporting

The incidence of a type of blemish can be indicated by the percent of rolls affected, but this gives no indication of severity. In recording severity, it has been found necessary to consider the blemishes of two major kinds: *spot blemishes* (types 1, 3, 4, and 6) and *character-associated blemishes* (types 2 and 5).

The concentration of *spot blemishes* may be indicated by the number of spots per square centimeter. In practice this might be the number of spots in the field of view with a particular instrument at a given magnification. The diameter of the field of view can be measured easily by placing a millimeter scale on the stage. The area can then be computed. The instrument most often used in the preliminary survey had a field 20 mm in diameter when the magnification was 10. The inspector soon learned to recognize on sight any one of five degrees of severity defined in table 2.

The concentration, in spots per square centimeter, is not a true measure of the severity of character-associated blemishes because the concentration is dependent upon the number, size, and shape of the characters. A severity classification based upon the effect of the blemishes on the information content of the document has been devised. This classification should have some relevance for the technical study of the phenomenon and is based upon criteria of utmost concern to those involved with the practical matter of maintaining records. For convenience in tabulating the data, a scale of five degrees of severity has been defined to correspond to the five degrees of severity of spot-type blemishes. The definitions of the five degrees of severity are given in table 3.

Because of the spectral nature of the blemishes and the spectral sensitivity of photographic materials, the legibility of characters affected by type 2 blemishes is often very much better on a photographic copy than it is on the original negative film. The inspector can nearly simulate this effect

TABLE 3. *Severity classification of character-associated blemishes*

1	Blemishes barely detectable, have no effect on the original shape and size of the characters.
2	Blemishes clearly visible, coloration change clearly visible, but shape and size of characters unchanged.
3	Blemishes well developed. Lines or parts of lines making up characters changed, but general shape of characters is unchanged. (See, for example, figs. 3 and 5.)
4	Blemishes have altered the shape and size of characters to the extent that individual characters could not be identified with certainty out of context. Characters can be identified in context.
5	Blemishes have so altered the size and shape of characters that they cannot be identified with certainty even in context. This constitutes information loss, on a given roll, in the practical sense. Information from other rolls of microfilm should not be considered part of the context.

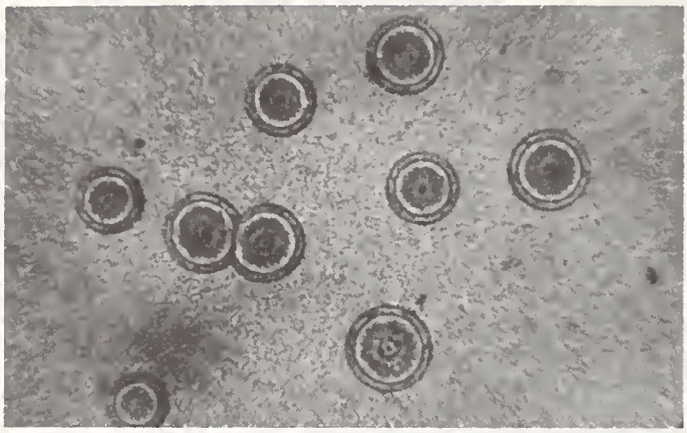
by viewing the original negative microfilm through a Wratten No. 47 blue filter. A similar increase in legibility is generally obtained with type 5 blemishes by viewing the original microfilm through a Wratten No. 29 red filter. Since these filters transmit only a small percent of the light, it may be necessary to increase the illumination when the filters are used. If characters can be identified with certainty by the use of filters, they should be considered identified with certainty for purposes of ascertaining the severity of the blemishes.

The most desirable kind of report may differ from agency to agency, and the factors of interest may change as the investigation of aging blemishes progresses. Some of the factors which have been recorded in the early stages of the investigation and which might continue to be of interest are:

roll identification (enough to locate roll later),
date of inspection,
film width in millimeters,
name of film manufacturer,
brand name of film,
type of base material,
date of manufacture of film,
date of film processing,
name of firm or agency which processed film,
type of storage can or carton,
type of reel,
range of storage temperature,
range of relative humidity in storage area,
kind of air conditioning (filtering, washing, etc.),
location of storage area (basement, for example),
use of humidification trays in storage cabinets,
whether leader was fogged or clear,
presence of splices,
frequency and type of splices,
some indication of activity of collection (such as dead storage, infrequent use, daily use),
presence of bands or strings around roll,
visual evidence of chemical residues,
presence of silvery sheen on fogged leader and other dense areas (on large areas of surface rather than just on spots),
evidence of general discoloration or fading,
presence of water spots,
incidence of various types of aging blemishes,
severity of blemishes,
blemishes of kinds not listed in this publication,
indication of correlation between incidence of blemishes and some other characteristic.

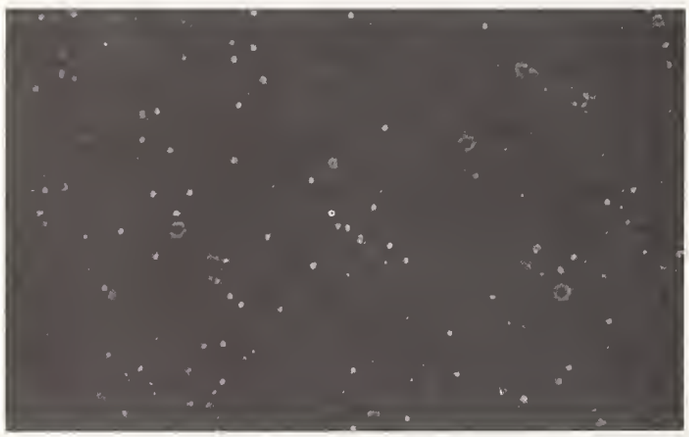
The summary of data on each stratum should include the following information of technical interest: name of stratum, location, date of inspection, name of inspector, number of rolls in stratum, number of rolls in sample, percent of rolls in sample having each kind of blemish on leader, percent of rolls in sample having each kind of blemish on information frames, and other data which appear to be indicative of strong correlations. Agencies may require other information for accounting and administrative purposes.

For counsel on the statistical sampling procedure, the author is grateful to Dr. W. J. Youden. The photomicrographs, so essential to a publication of this kind, were made by Mr. Myron Berkovitz. Miss Anna-Mary Bush did the bulk of the inspection in the preliminary program and devised the working definitions of the degrees of severity of spot-type blemishes on the basis of field of view. The author acknowledges with gratitude many exchanges of information with scientists at the Eastman Kodak Company who first identified and classified types 1, 2, and 3. The author appreciates the help of records officers in many Government agencies who have cooperated with the Bureau on this project by permitting access to records and providing research materials for the laboratory.

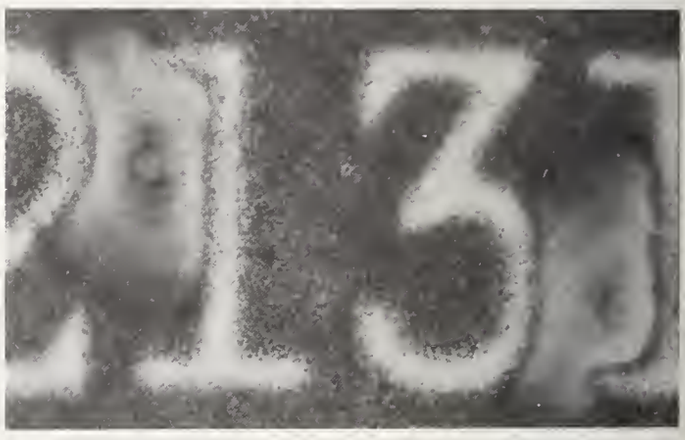


1

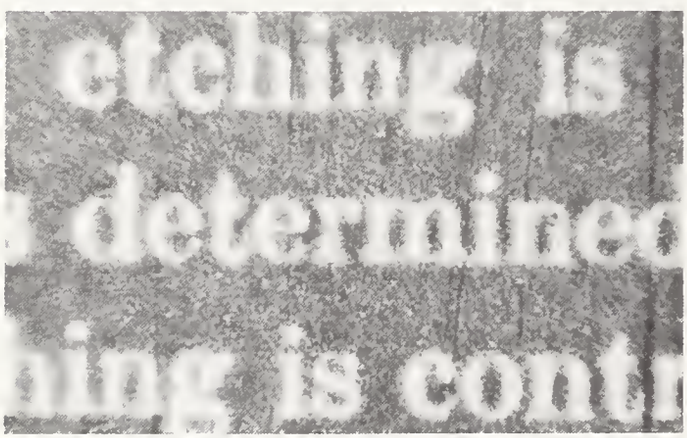
2



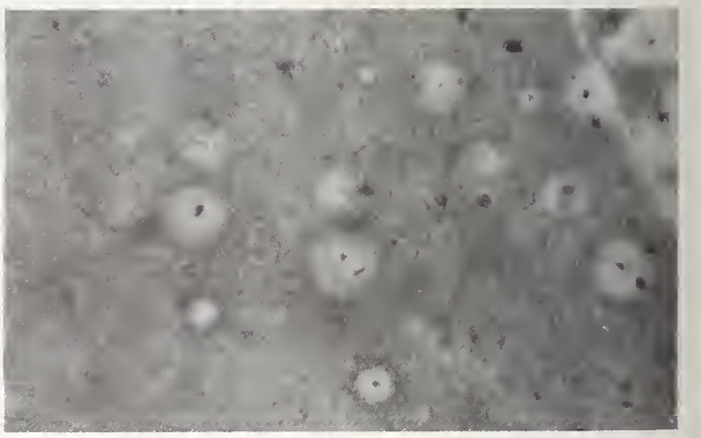
3



4



5



6

FIGURE 1. *Type 1.* On the record film, the spots shown range from 120 to 170 microns in diameter.
FIGURE 2. *Type 2.* On the record film, the height of the numbers is 180 microns.
FIGURE 3. *Type 3.* On the record film, the spots are about 15 microns across.
FIGURE 4. *Type 4.* On the record film, the height of the numbers is 240 microns.
FIGURE 5. *Type 5.* On the record film, the height of the lower case letters is 120 microns.
FIGURE 6. *Type 6.* On the record film, these spots range from 30 to 110 microns in diameter.

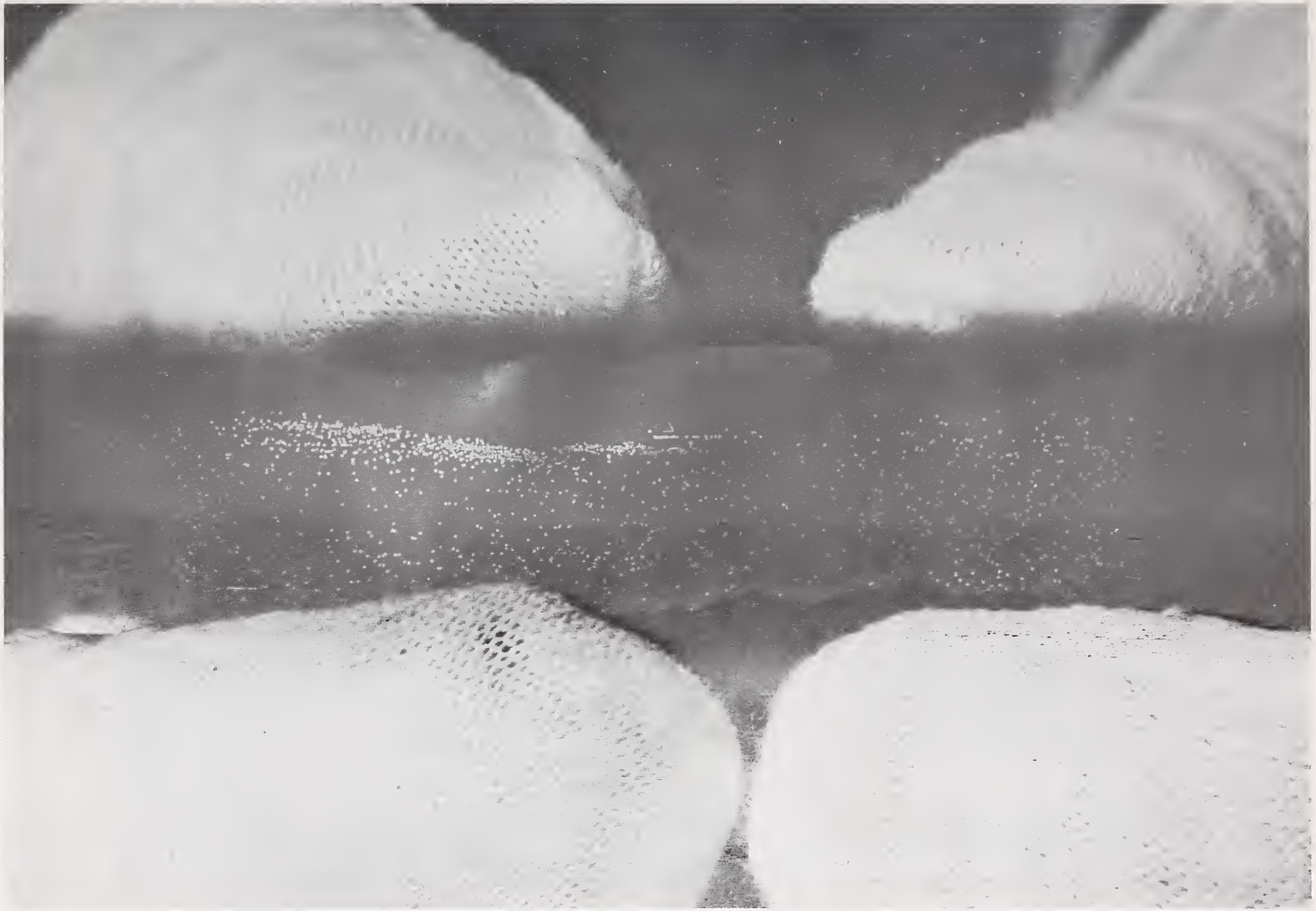


FIGURE 7. *Some spots can be seen with the unaided eye.* Here the silvery sheen of type 1 is seen on a 16 mm fogged leader by reflected light.

FIGURE 8. *The inspection of film for aging blemishes requires a binocular microscope with a magnification ranging from about 7 to about 30 times, substage illuminator, "spotlight" illuminator, rewinds, cotton gloves, and a clean working space.*

FIGURE 9. *Microscope with illumination for viewing by transmitted light.*

FIGURE 10. *Microscope with illumination for viewing by reflected light.*

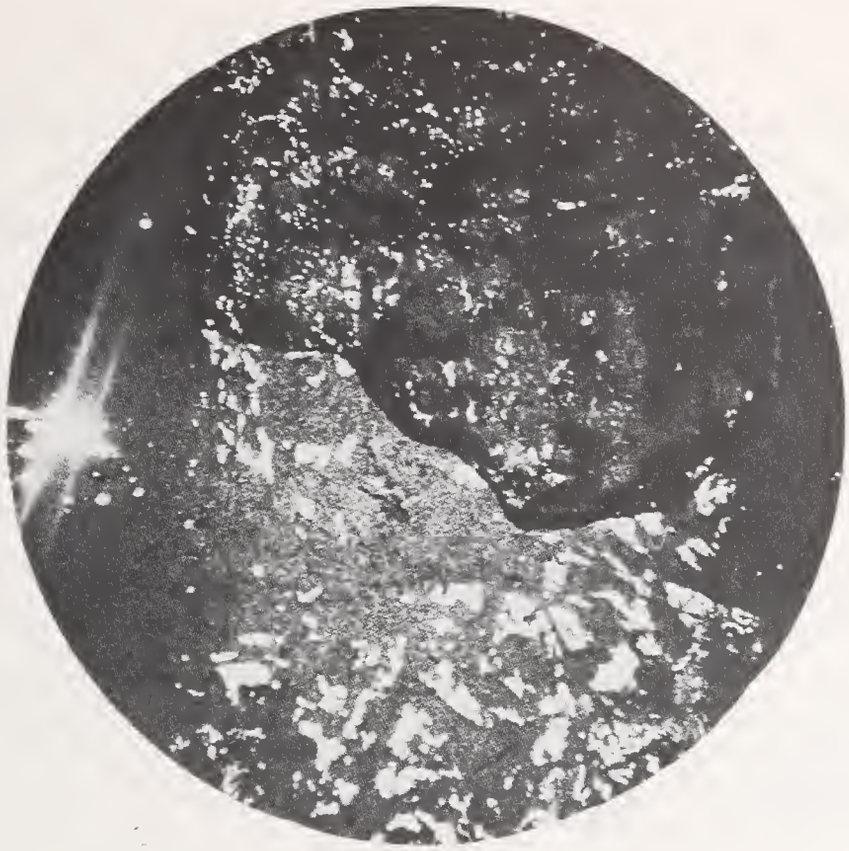


FIGURE 11. *View by transmitted light.* Differences in transmittance are made visible.

FIGURE 12. *View by reflected light.* Differences in reflectance, irregularities of the surface, and foreign matter on the surface are made visible.

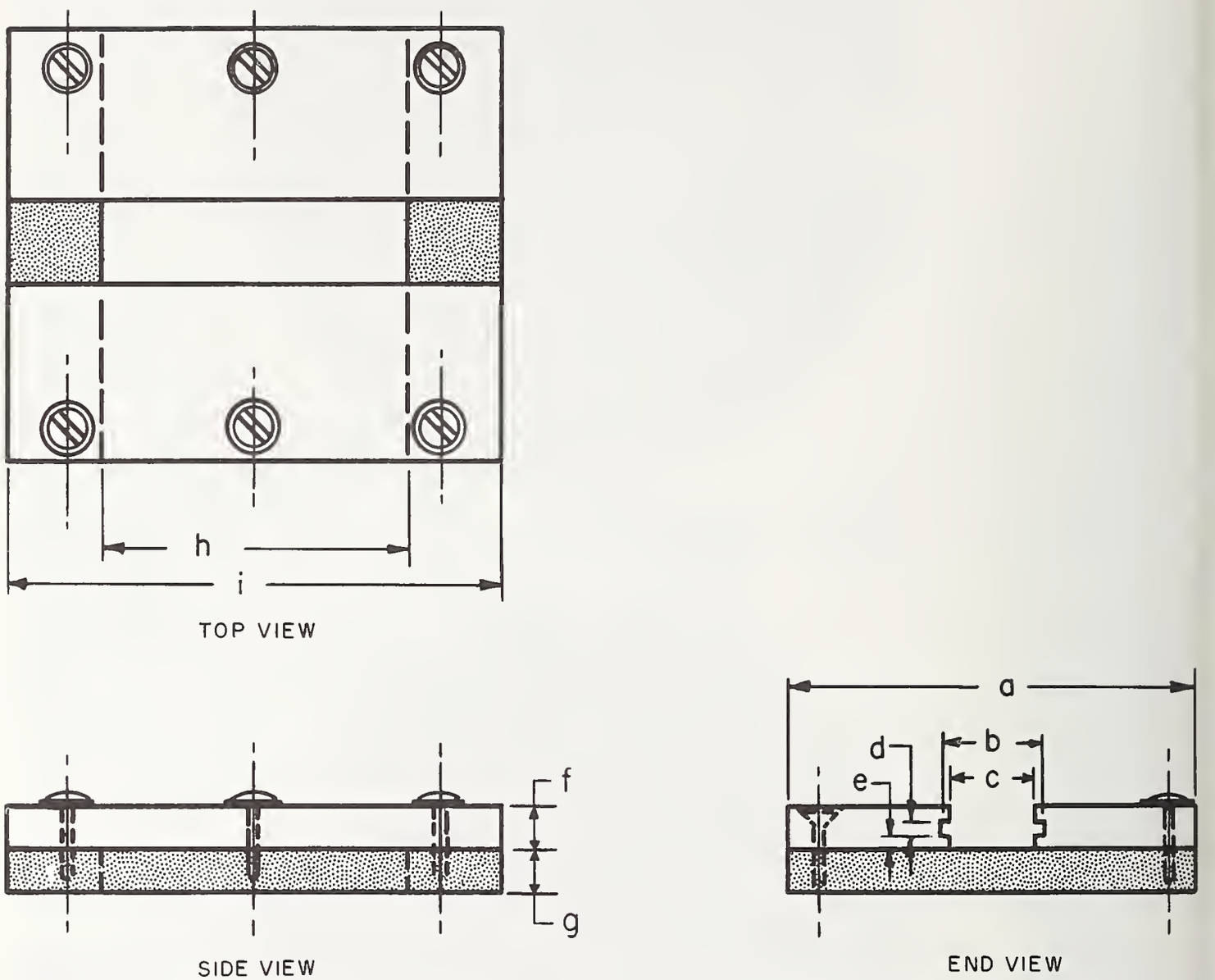


FIGURE 13. *Film guide for use on the microscope stage.* Dimensions are given in the text.

JOURNAL OF RESEARCH of the National Bureau of Standards—A. Physics and Chemistry
Vol. 72A, No. 3, May–June 1968

Blemish Formation in Processed Microfilm

C. I. Pope

Institute for Basic Standards, National Bureau of Standards, Washington, D.C. 20234

(January 9, 1968)

Oxidizing and reducing agents generated by paper cartons during storage may react with the image silver to form blemishes. The type of fixing bath, concentration of chlorine in the wash water, washing time and composition of the storage cartons can be factors that accelerate or retard the attack of the peroxide on the image silver. Microfilm washed after fixation in distilled water or chlorine-free tap water formed blemishes when exposed to peroxide paper. A trace of silver chloride in the image silver of processed microfilm augmented the formation of the natural type blemishes when exposed to peroxide paper, but high concentrations of residual silver chloride inhibited blemish formation. During storage, some storage cartons evolved formaldehyde, formic acid, and ammonia which created a chemical environment favorable for blemish formation in microfilm in the presence of peroxide. A correlation was found between the incidence of blemishes and the brand of paper storage cartons. A procedure was developed for testing processed microfilm to determine its susceptibility to blemish formation. According to the present theory, peroxide reacts with the silver in the image, forming colloidal silver which imparts a yellowish or reddish color to the blemishes.

Key Words: Archival record film, blemishes due to aging, colloidal silver, microfilm, microfilm storage cartons, permanent record film, peroxides.

1. Introduction

Blemish formation in stored microfilm has been under investigation for some time [1, 2, 3].[1] The nature of the problem and the types of blemish formation have been described [1, 2]. It has been shown that oxidizing and reducing agents, probably peroxides, emitted by paper cartons in which the microfilm is stored attack the silver in the image, causing yellowish or reddish blemishes [4]. In the present study, the processing procedure was investigated to determine its effect on blemish formation. In this paper the term "blemish" includes various types of spots and edge attack on characters and other image boundaries.

2. The Role of the Processing Procedure in Blemish Formation in Microfilm

In the investigation of processing variables, four fine grain 16 mm microfilms, made by three manufacturers, were used. The processing solutions and times of developing, stopping, and fixing are given in the appendix. The microfilm samples were processed in beakers and agitated with a glass rod. The chlorine concentration of the wash water was determined by a Taylor chlorine analyzer [7] using the *o*-tolidene dihydrochloride reagent.

[1] Figures in brackets indicate the literature references at the end of this paper.

The peroxide source for testing the microfilm was a chromatographic paper, 0.03 cm thick, which had been treated in aqueous hydrogen peroxide. The test procedure is described in section 5. From 2 to 4 cm^2 of the peroxide paper formed spots in areas of microfilm containing natural spots. All microfilms were dried at least 1 day after processing or other treatment before exposure to peroxide paper.

2.1. Effect of the Fixing Bath and Wash Water

A study was made to determine if processed microfilm formed blemishes when the thiosulfate was removed by washing in chlorine-free water. The fixing baths tested were a neutral bath, a nonhardening acid bath, and an acid hardening bath. The wash waters were distilled water and chlorine-free tap water which had a *p*H of about 8. Chlorine-free tap water was obtained by storing cold tap water until the chlorine disappeared or by adding sodium sulfite to the tap water to neutralize the chlorine. For a chlorine concentration of 1.2 ppm, 50 ml of a sodium sulfite solution containing 0.500 g per liter was sufficient for 4 liters of tap water. About 50-cm lengths of the microfilms were exposed to a test pattern having various kinds of information, developed, fixed and washed in 1-liter baths of the wash water in 1.5-liter beakers. Three baths of the wash water were used for a 3-min wash and for longer washes a bath was used for each additional 5 min of washing. The sample was trans-

ferred from beaker to beaker and agitated with a glass rod. The temperature of the processing solutions and wash water was 26 °C.

The processed microfilms developed spots and, in some cases, edge attack when exposed to peroxide paper. The microfilm fixed in the neutral and the nonhardening baths and washed in distilled water developed much larger spots than microfilm fixed in the nonhardening and the acid hardening baths and washed in chlorine-free tap water. The spots in the latter two fixing procedures were more numerous and had sharper edges than those for the first two fixing procedures with distilled water washing. The spots were almost contiguous to each other for microfilm fixed in the acid hardener bath. The spots from all fixing baths appeared white when they first developed but turned yellow or yellowish as the peroxide attack continued. The edges of the spots were slightly fuzzy. No significant differences were observed between blemishes formed after a 3-min wash and a 15-min wash.

The results of the tests are shown in table 1, where the observed effect for each processing procedure and the amount of peroxide paper necessary to induce blemish formation are indicated. The residual thiosulfate in the microfilm samples that formed blemishes ranged between 0 and 3 μg per 6 cm^2 of the image area. Microfilm D retained more residual thiosulfate than the other microfilms and exhibited only slight blemish formation for the neutral and the nonhardening baths and none for the acid hardening bath when washed in chlorine-free water. Microfilm A, fixed in the acid hardening bath and washed in distilled water, had a residual thiosulfate concentration of 10 μg per 6 cm^2, making it resistant to peroxide attack. The results show that microfilm developed blemishes when exposed to peroxide paper if the processing procedure was such that the residual thiosulfate was reduced to a very low level. A higher concentration of residual sodium thiosulfate would form additional silver sulfide in the image silver, protecting it from oxidation. The thiosulfate was more difficult to wash from the microfilm after fixation in the acid hardener bath than in the neutral or nonhardening acid bath, but chlorine-free tap water still removed the thiosulfate sufficiently to permit blemish formation. The tap water, being slightly alkaline, acted as a thiosulfate eliminator.

2.2. Effect on Blemishes of Silver Chloride Formed in Microfilm by Chlorine in Tap Wash Water

It was found that silver chloride formed in processed microfilm during processing is a factor in blemish formation. Trace amounts of silver chloride enhance blemish formation but larger concentrations inhibit it.

Water from city taps generally contains 0.1 to 1 ppm of chlorine [8]. The chlorine in tap water forms hypochlorous acid (HOCl) which is a powerful oxidizing agent because it liberates oxygen ($HOCl \rightarrow HCl + O$) [9]. The evolved oxygen oxidizes the silver in the image to silver ions which in turn combine with the ever-present chloride ions in tap water to form silver chloride.

Tests were made of the effect of chlorine in the wash water on blemish formation in microfilm A. The microfilm was fixed in an acid hardening bath. Samples were washed for 30 min in running tap water at 27 °C, containing 0.8 ppm of chlorine in one run and 1.2 ppm in a second run, and exposed, respectively, to 6 cm^2 and 8 cm^2 of the peroxide paper. No blemishes developed on the microfilm samples. Other samples were washed at 26 °C for a total of 30 min by placing them in a succession of tap water baths containing 0.3 ppm of chlorine for 2, 3, 5, 10, and 10 min, respectively. Exposure to 4 cm^2 of peroxide paper produced some

TABLE 1. *Blemish formation on exposure to peroxide paper for four microfilms and various processing procedures*

	Processing procedures			
Microfilm	Neutral fix and washed in distilled water for 3 min	Nonhardening acid fix and washed in distilled water or chlorine-free tap water for 3 min	Acid hardening fix and washed in chlorine-free tap water for 3 min	Acid hardening fix and washed in distilled water for 40 min
A	Many spots and edge attack.	Many spots and edge attack.	Many spots but no edge attack.	No spots or edge attack.
B	Many spots and edge attack.	Many spots and edge attack.	Many spots but no edge attack.	
C	Many spots and edge attack.	Many spots and edge attack.	Many spots but no edge attack.	
D	Few spots but edge attack.	Few spots but edge attack.	No spots or edge attack.	
Amount of peroxide paper used.	1 to 2 cm^2	2 cm^2	4 to 8 cm^2	8 cm^2.

spots but a considerable resistance to blemish formation was noted when compared to samples washed in chlorine-free tap water.

After fixation, microfilm A was washed in running tap water at 27 °C, containing 1.2 ppm chlorine for 3, 6, 9, 15 and 30 min. A sample was removed at the end of each washing time. The processed microfilm was exposed to 4 cm^2 of the peroxide paper. The microfilm samples fixed in the acid hardening bath developed a high incidence of yellow spots for the 3-min wash, a few in the 6-min wash, and none in the 9, 15, and 30-min washes. The microfilm samples fixed in the nonhardening bath developed many light yellow spots for the 3-min wash, with slight yellowing of the image silver, a decrease in number of spots and no yellowing for the 6-min wash, a few spots for the 9-min wash, and none for the 15 and 30-min washes.

Thus a high concentration of chlorine in the wash water formed silver chloride in the image silver making it resistant to yellowing and spot formation when exposed to peroxide. Furthermore, resistance to blemish formation increased with the washing time for 1.2 ppm of chlorine in the tap wash water and 10 to 15 min was required to impart complete resistance. At the end of the 3-min wash, the thiosulfate had reached a low concentration and the chlorine began to form silver chloride. As the washing progressed, more silver chloride was formed and it finally reached a concentration which inhibited blemish formation. The spots which formed after fixation in the acid hardening bath and a 3-min wash were more yellow than those obtained from the nonhardening bath and a 3-min wash. This would be the expected effect of acid carried over into the wash water from the acid bath, since chlorine reacts more rapidly in the presence of acid to form silver chloride. Since the tap water had a *p*H of about 8, a high concentration of chlorine and a considerable washing time was necessary to form sufficient residual silver chloride to inhibit blemish formation. Tap water containing low concentrations of chlorine, such as 0.3 ppm, may not form a sufficient amount of silver chloride to inhibit blemish formation even after a 30-min wash but the tests described above for 0.3 ppm of chlorine in wash water showed that some resistance to blemish formation was obtained.

Microfilm containing sufficient silver chloride to make it resistant to blemish formation loses this resistance if treated for 10 min in a 0.2 percent solution of ammonium hydroxide to remove the silver chloride. Microfilm so treated showed only an even yellowing of the image silver when exposed to 2 to 4 cm^2 of the peroxide paper.

It is not known how long silver chloride formed from chlorine in the tap water will impart resistance to blemish formation from peroxide attack. But in section 3 it is shown that chemicals given off by the storage cartons do react with the image silver, causing blemish formation in the presence of peroxide. Evidently the silver halide, if present, is partially reduced by the peroxide.

2.3. Effect of Iodide in the Fixer

It has been shown [12] that the silver in a photographic image is partially sulfided during fixation. If iodide is added to the fixing bath most of the sulfiding is prevented, leaving a layer of silver iodide on the silver grains.

Henn, Wiest, and Mack [14] have reported that collections of microfilm processed in iodide-containing developers or fixing baths to be relatively free from blemishes. They recommend the use of 0.2 g of potassium iodide per liter of fixing solution to give microfilm greater resistance to blemish formation.

Tests were made on four microfilms by adding 0.2 g of potassium iodide to each liter of the acid hardening fixing solution. After development, the samples were fixed and washed in chlorine-free tap water. The chlorine-free tap water was used to avoid the blemish inhibiting effect of the chlorine. The samples were exposed to 2 to 8 cm^2 of the peroxide paper. The microfilms showed more even yellowing of the image silver than the controls without iodide, indicating that iodide had made the image silver more susceptible to attack by peroxide. (If natural blemishes were caused by peroxide from the boxes, this finding seemed inconsistent with the field observations.) The resistance to spot formation was about the same as the controls. However, if the microfilms fixed with iodide in the fixer were washed in the chlorine-free tap water to which 1 ppm of chlorine was added, they, like the controls, resisted blemish formation and did not show an even yellowing. Old microfilms which had been processed with iodide in the fixing bath formed blemishes when exposed to peroxide at low levels of 2 to 4 cm^2 of the peroxide paper.

To simulate actual practice, a study was then made of the effect of storing processed microfilm in fresh storage cartons. The films in cartons were placed in a desiccator jar for 10 days at 86 percent relative humidity and 26 °C and then the samples were exposed to 2 to 4 cm^2 of the peroxide paper. The samples processed with potassium iodide in the fixer developed a few blemishes but samples processed without potassium iodide in the fixer developed many more. These tests confirmed the field observations and suggested that chemicals in addition to peroxide were generated in the storage carton and played a part in the chemistry of blemish formation.

2.4. Natural-Type Blemishes Obtained With Acid and Silver Chloride in the Image

A small amount of silver chloride was formed in the image silver of old films and freshly processed films. When these films were exposed to peroxide, sharp-edged blemishes formed which resembled very closely those found in nature, even duplicating the ring formation and edge attack. The films were exposed to about 25 cm^2 of the peroxide paper at 76 percent relative humidity and 26 °C. A higher relative

humidity did not produce the natural-type blemishes. Although these tests may not be applicable to microfilm as normally processed and stored they do suggest that a trace of silver chloride contributes to blemish formation under natural storage conditions.

For these experiments, processed films were treated 10 min in a solution containing 0.025 g of sodium chloride and 3 ml of glacial acetic acid per liter, dried and exposed to peroxide paper. (Formic acid may be substituted for the acetic acid.) The 0.025 g of sodium chloride per liter gave a solution containing 15 ppm of the chloride ion which is close to the median value of 13 ppm in the water supplies of large cities [6]. The oxygen dissolved in the sodium chloride-acetic acid solution oxidized the image silver to silver ion which combined with the chloride ion to precipitate silver chloride in the image silver. The silver chloride formed in microfilm by the sodium chloride-acetic acid solution may be removed by treating the film for 10 min in a 0.2 percent solution of ammonium hydroxide. After such treatment, the image silver showed only an even yellowing after exposure to peroxide paper. When microfilm was treated in a solution of 3 ml of acetic acid per liter without the sodium chloride and exposed to peroxide paper, no spot or edge blemishes formed in the image silver, only an even yellowing.

The microfilms A, B, C and D, reported in section 2.1, which readily formed blemishes when exposed to peroxide paper were treated in the sodium chloride-acetic acid solution and dried. Peroxide exposure produced natural type spots, some with ring structure, and edge attack. There was no even yellowing, indicating that silver chloride was formed throughout the image. The edges of the spots were sharp and the spots were deep yellow, showing the effect of silver chloride. The number of spots was much less than before the sodium chloride-acetic acid treatment since too much silver chloride had been formed at most points in the image for spot induction.

Blemish formation was induced in old processed microfilms by treatment in chlorine gas or an acid solution of chlorine and subsequent exposure to peroxide. The films were exposed to chlorine gas by suspending the samples over a chlorine solution in a 1-liter closed glass bottle. The chlorine solution was prepared by dissolving 0.250 g of technical grade calcium hypochlorite containing 70 percent calcium hypochlorite in 1 liter of water and adding 5 ml of glacial acetic acid. Approximately 400 ml of this solution was added to the bottle. The film samples were soaked in distilled water for 10 s, squeegeed with a dry glass rod and suspended immediately in the chlorine gas above the solution for 20 to 60 s, dried overnight and exposed to peroxide paper. This method produced spot and edge attack but each brand of microfilm required a different time of exposure to the chlorine gas.

To determine whether silver chloride was formed in the image silver by the chlorine gas, microfilm samples were treated in a 0.2 percent ammonium hydroxide solution for 5 min to remove any existing silver chloride, dried overnight, treated 10 s in distilled water, squeegeed, treated to the chlorine gas for 30 s and exposed to peroxide paper. These films had spot and edge attack.

Microfilm samples were immersed for 5 s in the chlorine solution described above, dried and exposed to peroxide paper. These microfilms showed spot and edge attack but were more resistant to peroxide because more silver chloride had been formed in the image silver.

The findings with respect to chlorine and chloride ion raised a question about the effects of the other halides, silver bromide and silver iodide. Old processed microfilm samples, blemish free, were treated 10 min in solutions of sodium bromide containing 0.045 g/l and 3 ml of glacial acetic acid, and potassium iodide containing 0.060 g/l and 3 ml glacial acetic acid. Strong spot and edge attack were obtained in most samples when exposed to peroxide paper. The amount of blemish formation decreased as the solubility of the corresponding silver halides decreased.

A method was developed for estimating the free silver in processed microfilm. "Free silver" as here defined includes silver ion, silver oxide and silver halides but the method does not show how much of the silver is in the form of halide salt. This method indicated as much free silver present in old processed microfilm which had been treated in a solution containing 3 ml of acetic acid per liter as in microfilm treated in a solution containing 3 ml of acetic acid and 0.025 g of sodium chloride per liter. However, when they were exposed to peroxide paper, the samples given the acetic acid and sodium chloride treatment resisted background yellowing though they had natural-like spots and edge attack while those given the acetic acid treatment showed only an even yellowing of the image silver. Apparently traces of silver in the image are oxidized by oxygen dissolved in the water, in the presence of acid, and if the chloride ion is present, silver chloride is formed.

The procedure for determining the free silver is outlined as follows. The free silver was extracted from 12 cm^2 of processed microfilm with 10 ml of a 0.2 percent solution of ammonium hydroxide containing 1 g of sodium chloride per liter. Two 6 cm^2 pieces of microfilm were cut, folded emulsion in and placed back to back in a glass vial containing the test solution and allowed to stand 15 min. The samples were removed by a glass rod having a hook at one end and 0.10 ml of a potassium bromide solution containing 100 g/l was added and mixed. Then the solution was acidified by adding 1 ml of concentrated nitric acid diluted 1 to 5 and mixed, allowed to stand 1 hr and the turbidity measured with a nephelometer. A calibration curve was prepared from known amounts of silver added to the standard solution as silver nitrate. One-fourth microgram of silver as silver nitrate gave a measurable turbidity. The values obtained are relative but were helpful in showing changes in the free silver content of microfilm after various treatments. The free silver content of old processed microfilms containing blemishes ranged from 0.2 to 1 μg for 6 cm^2 of film.

3. Formaldehyde, Formic Acid, and Ammonia From Paper Cartons

Microfilms in which blemishes had formed during natural storage were studied to determine if more blemishes would develop in the laboratory peroxide test. Several rolls of microfilm having blemishes on the outer convolutions but none on the remainder of the roll were tested. Samples taken from blemish-free areas were tested and no blemishes developed, but samples from areas where blemishes had formed during storage did develop blemishes like those already present. The blemishes present before the test were ringed with a stylus to differentiate them from blemishes produced by the peroxide. Evidently chemicals in addition to peroxide had evolved from the paper carton during storage and created a chemical environment conducive to blemish formation in the presence of peroxide.

Formaldehyde, formic acid, and ammonia have been detected in the atmosphere within many of the paper storage cartons when stored at 100 percent relative humidity at 26 °C. A glass petri dish, 60 mm in diameter and 13 mm high, containing 10 ml of distilled water was placed in the paper carton to absorb evolved chemicals. The carton was stored in a desiccator jar containing 25 ml of distilled water for 1 to 3 days, one day being generally sufficient. This extraction procedure was used in testing for the presence of formaldehyde, formic acid, and ammonia. Formaldehyde-urea and formaldyhyde-melamine resins have been widely used as wet-strengthening agents in paper manufacture for over 20 years. The chemicals detected were probably formed by hydrolysis of these resins at high relative humidities.

3.1. Detection of Formaldehyde, Formic Acid, and Ammonia

a. Formaldehyde

Two ml of the water solution which had been stored in the paper carton were added to 5 ml of a 1 percent solution of chromotropic acid (4,5-dihydroxy, 2,7-naphthalene disulfonic acid[2]) in concentrated sulfuric acid. Chromotropic acid readily detects formaldehyde in concentrations of 1 ppm (15). Chromotropic acid gave a pink to dark purple color, a positive test for formaldehyde. The less sensitive resorcinol test gave a positive indication of formaldehyde from some cartons when the same water solution was stored successively in 3 cartons for a period of 3 days each. This test was made by adding 0.1 ml of 1 percent solution of resorcinol to 5 ml of the water solution and pouring the mixture carefully down the side of a test tube containing 5 ml of concentrated sulfuric acid to form two layers. A pink color formed at the junction of the two layers. The formaldehyde was also identified by microwave absorption analysis.

[2] Obtained from Matheson Coleman and Bell, Cat. No. P5425, East Rutherford, New Jersey.

b. Formic Acid

The water stored in the paper cartons had a *p*H range of 3.5 to 4.5 after storage for 3 days at 100 percent relative humidity at 26 °C. Titration with 0.001 *N* sodium hydroxide solution indicated that the water stored in the cartons was about 0.001 *N* with respect to formic acid for some storage cartons. The water stored in the cartons decolorized a light pink solution of potassium permanganate, indicating that formic acid was present. When 5 ml of the water solution was added to 5 ml of saturated mercuric chloride and warmed in hot tap water for 10 min, a precipitate formed, indicating reduction of the mercuric chloride. This test indicated the presence of formic acid. A precipitate of reduced silver formed when .2 ml of silver nitrate solution containing 10 g per liter were added to 10 ml of the water solution from the paper cartons and warmed 15 min in hot tap water. The reduced silver indicated the presence of formic acid which is capable of reducing the silver ion. This procedure detected 5 ppm of formic acid in a known solution. The formic acid is evidently formed by oxidation of the formaldehyde by the peroxide generated in the storage cartons and possibly by aerial oxidation of formaldehyde, catalyzed by the silver in the image. A water solution of formaldehyde was found to increase in acidity when stored over silver dust for 4 hr showing that formic acid was formed by the oxidation of formaldehyde.

c. Ammonia

About 10 ml of the water stored in the paper cartons were placed in a 50-ml beaker and made alkaline with sodium hydroxide. The beaker was covered with a watch glass having a strip of moistened red litmus paper attached to the underside and the beaker was warmed in hot tap water. After 30 to 60 min the litmus paper turned blue suggesting the presence of ammonia. A control test run on the distilled water was negative.

3.2. Tests of Old and New Storage Cartons

Some new storage cartons readily evolved formaldehyde when stored for 3 hr in the drawer of a metal laboratory desk. A strong indication of formaldehyde was obtained after a new carton was stored at 100 percent relative humidity at 3 °C for 1 day, indicating that a low temperature does not stop the evolution of formaldehyde if the relative humidity is high. Most of the old storage cartons tested evidenced formic acid evolution.

Storage cartons from one collection of blemish-free microfilms evolved no formaldehyde or only a very small amount. Cartons from a second collection where blemishes have not been found evolved an appreciable amount of formaldehyde, but this collection is located in an arid climate which would minimize the evolution of formaldehyde and the formation of formic acid. All the tests of paper storage cartons from collections where blemishes have been found, even collections that were ten years old, gave a strong indication of formaldehyde.

3.3. Reactions of Formaldehyde, Formic Acid, and Ammonia on Silver and Gelatin

During natural storage, blemishes generally form on the outside leader and edges of the microfilm roll, which are the areas most accessible to gaseous chemicals evolved by the storage carton. Formaldehyde will harden gelatin by reacting with the amino groups attached to the amino acids in the gelatin. Blemish formation was inhibited in a blemish forming microfilm after an exposure of 24 hr to the gaseous phase of a 2 percent solution of formaldehyde. This severe test may not be relevant to natural storage conditions but suggests that formaldehyde, alone, may inhibit blemish formation. Formaldehyde and formic acid neutralize the alkalinity of the amino groups making the gelatin more acidic.

Formic acid and chloride ion reacted with the image silver at high relative humidities to form silver chloride in the same manner as described in section 2.4 for the sodium chloride-acetic acid solution, making the high-density uniform image more resistant and the edges of the blemishes sharp, when the film was exposed to peroxide evolved by the storage cartons.

In section 2.2 it was noted that tap wash water having as much as 1 ppm of chlorine formed a sufficient amount of silver chloride in the microfilm to make it resistant to blemish formation. The formic acid from cartons stored at high relative humidities appears capable of reducing the silver chloride concentration to a level favorable for blemish formation. Formic acid readily reduced the silver ion in an aqueous solution of silver nitrate, forming a gray precipitate of silver. Solid silver chloride in a solution of formic acid turned slightly gray after standing in the dark for 7 days. A control without formic acid did not exhibit this effect. The gray color was attributed to reduced silver.

Experiments were performed to see whether or not residual silver chloride in the image silver of processed microfilm is susceptible to reduction by formic acid and peroxide. To incorporate silver chloride into clear (image-free) processed microfilm, the film was washed 20 min in distilled water, treated 10 min in a solution containing 5 g of silver nitrate per liter, treated 10 min in a second solution containing 5 g sodium chloride per liter, washed 20 min in distilled water and dried. When exposed to 24 cm^2 of paper treated with formic acid, samples of this microfilm developed a yellow color in the gelatin layer, indicating reduction of the silver chloride. The test was made at 86 percent relative humidity and 26 °C and in the same type of desiccator jar used for peroxide tests. The paper was prepared by treating Whatman No. 3MM filter paper in a 10 percent formic acid solution for 30 min drying it 3 hr at 26 °C and storing it at approximately −10 °C. Microfilm placed in a new storage carton for 12 days at 86 percent relative humidity and 26 °C showed evidence of reduction of silver chloride. Samples exposed to 6 cm^2 of the peroxide paper also evidenced such reduction. Thus residual silver chloride reduction in stored microfilm should decrease the blemish resistance.

It was found that the silver-chloride-ammonia complex was more easily reduced by formic acid and peroxide than silver chloride precipitated in the gelatin layer as described above. Tests were made on clear (image-free) processed microfilm treated 5 min in a solution containing 0.5 g of silver chloride dissolved in one liter of water containing 8 ml of concentrated ammonium hydroxide. The results indicated that ammonia evolved by storage cartons may make the silver chloride more sensitive to reduction by formic acid and peroxide.

The above data shows how microfilm resistant to blemish formation when stored may lose its resistance to peroxide in the presence of other chemicals evolved by the paper storage carton. Although both peroxide and formic acid reduce silver chloride, formic acid may be the principal reducer since blemishes have generally been found in microfilm stored in paper cartons which evolve relatively large amounts of formaldehyde and formic acid.

4. The Chemistry of the Formation of Blemishes in Processed Microfilm

4.1. Theory

Paper cartons in which microfilm is stored undergo degradation at high relative humidity and high temperature to form peroxide. The silver in the image serves as a catalyst to decompose the peroxide, forming oxygen which oxidizes the image silver to the silver ion. The chloride ion catalyzes the oxidation of silver by peroxide. The peroxide [11] and formic acid reduce the silver ion to yellow colloidal silver.

Processed microfilms exposed for 2 min to the gas phase of 20-percent hydrochloric acid formed spot blemishes when exposed to peroxide paper. The hydrochloric acid reacted with the image silver to form silver chloride [10]. Many white spots formed, a few with a sharp ring structure, but they were colorless because the gelatin was too acid for the formation of the yellow to red colloidal silver found in naturally occurring spots. Under these conditions only oxidation of the silver occurred but concomitant oxidation and reduction must occur to form yellow to red colloidal silver. When the samples with the white spots were given a printout exposure and treated in a developer solution the spots were still colorless. This indicated that the silver ion had not remained in the colorless area but had migrated to the edge of the spots.

The silver ion may be reduced in an aqueous solution to form colloidal silver [9]. The reduction of silver ion in the gelatin layer of processed film by peroxide forms yellow colloidal silver. A proprietary product containing 19 to 23 percent silver in the colloidal state in gelatin is known in the trade as "Solargentum" [13]. Processed microfilm having no silver image was washed in distilled water, treated in a dilute solution of silver nitrate or silver acetate and dried. When samples were exposed to peroxide at 86 percent relative humidity and 26 °C, the gelatin layer turned yellow. Reducing agents evolved by paper storage

cartons formed the same yellow color in the gelatin of other samples stored in the cartons for 10 days at 86 percent relative humidity and 26 °C.

There is another type of colloidal silver formed when ionic silver is reduced in the presence of an insoluble silver salt such as silver chloride. This type of colloidal silver appears to be silver dispersed in a silver halide and gives colors ranging from pink, red, yellow, brown to gray when formed in an aqueous solution but yellow or reddish when formed in the gelatin layer of microfilm. Carey Lea claimed that this colloid was a mixture of silver chloride and a subhalide of silver. Mees [5] has pointed out that it may be regarded as a mixture of silver in silver halide.

Tests showed that the blemishes contained ionic silver in addition to colloidal silver. The black specks in the centers of many spots appear to be reduced silver since they are readily dissolved by strong nitric acid. In section 2 it was noted that large spots, with light yellowing and even attack on the image silver, were obtained when the microfilm was fixed in a neutral fixer or a nonhardening fixer and washed in distilled water. In these processing procedures no chloride or bromide ions should be present to form silver halide and the yellow color would be attributed to the silver-gelatinate type of colloidal silver. However, when the microfilm was fixed in the nonhardening fix or the acid-hardening fix and washed in chlorine-free tap water the spots were sharper, more yellow, attended by no edge attack or even attack on the image silver, and required more peroxide for spot formation. The chloride ion introduced by the tap water formed silver chloride which led to the formation of the silver-silver chloride type of colloidal silver in spots and prevention of even attack on the image silver. The silver-silver chloride colloidal silver probably predominates in blemish formation.

4.2. Formation of Silver-Silver Halide Colloidal Silver in Solution and on Silver Strips

It was found possible to form the silver-silver chloride type of colloidal silver in water solution. Thirty percent hydrogen peroxide was added to a solution containing varying concentrations of sodium chloride in the presence of solid silver oxide. A small amount of silver oxide dissolved making the solution alkaline. Reduced gray silver accumulated in the bottom of the beaker, leaving the solution clear when no chloride ion was present. When chloride ion was added, a colloidal suspension formed, ranging in color from pink to yellow and sometimes a reddish precipitate formed along with the gray reduced silver. Treatment of the reaction mixture with nitric acid dissolved the reduced silver and left a yellow or reddish colloidal suspension and a reddish precipitate. The pink, yellow and reddish color in the colloidal suspension resembled the color of natural spots. The optimum condition for producing the colloidal silver was a 60-ml solution containing 0.05 g of sodium chloride and about 0.5 to 1 g of solid silver oxide. The hydrogen peroxide was added dropwise with stirring.

Equal mol concentrations of potassium bromide (0.102 g) and potassium iodide (0.142 g) equivalent to 0.05 g of sodium chloride were used in the above procedure to test for the formation of colloidal silver. The potassium bromide solution produced a light yellowish colloidal suspension with a trace of reddish color and some reddish precipitate. The potassium iodide solution gave no visual evidence of the formation of colloidal silver in the solution although a faint trace may have been present. A small amount of sodium chloride in the potassium iodide solution did produce a slightly yellowish colloidal solution indicating that colloidal silver will form in the presence of potassium iodide and the chloride ion.

The silver-silver chloride type of colloidal silver was also formed on silver strips. The silver strips were cleaned with nitric acid, washed in distilled water and immediately immersed in a 50-ml solution containing 0.05 g of sodium chloride, to which 30 percent hydrogen peroxide was added dropwise. A heavy reddish coating of colloidal silver formed on the silver strip showing that the silver on the strip was both oxidized and reduced by the peroxide. When no sodium chloride was added, the silver strip remained bright and served only as a catalyst to decompose the peroxide into water and oxygen. There was no precipitate of reduced silver. These results showed that the chloride ion played an important part in the oxidation of the silver by the peroxide.

5. Test for the Formation of Blemishes in Processed Microfilm With Peroxide Paper

5.1. Preparation of the Peroxide Paper

It was found that paper treated in a solution of hydrogen peroxide was a good source of gaseous peroxide for the formation of blemishes in processed microfilm. All of the types of blemishes described in NBS Handbook 96 [2] have been produced by exposing films to such paper. The paper used was Whatman No. 3MM filter paper which is used in chromatography. The paper was treated 30 min in a 5-percent solution of hydrogen peroxide at room temperature with occasional stirring, hung in a ventilating hood for 2 hr, dried for 24 hr in a desiccator over anhydrous calcium sulfate at 26 °C and stored in a sealed glass bottle in a refrigerator at −10 °C.

5.2. Apparatus

The tests were made in a desiccator jar containing a saturated salt solution to control the relative humidity. The jar had an inside diameter of 160 mm at the ground flange and a volume of approximately 2300 ml. The air and salt solution were not stirred but the relative humidity value reported is that recorded in the literature for equilibrium conditions. The jar was covered with a flat glass plate about 0.6 cm thick and sealed with petroleum jelly. A porcelain plate perforated with small holes was located just above the saturated salt

solution. Saturated solutions of sodium chloride and potassium chloride were used to obtain relative humidities of about 76 and 86 percent, respectively, at 26 °C.

The test samples, about 6 cm in length, were fastened at each end with black photographic masking tape to a thin glass plate about 8 × 10 cm in size. Four or five samples could be mounted on a plate. The glass plate was then fastened at two opposite edges to the underside of the cover with a strip of the tape, so that the emulsion faced downward. A glass dish, 60 mm in diameter and 13 mm high, was centered in the middle of the porcelain plate to hold the peroxide paper. The mounted microfilm samples were about 8.5 cm from the peroxide paper. In testing microfilms for blemish formation the peroxide paper was added to the glass dish and allowed to react 24 hr at 26° ± 1 °C. Two cm^2 of peroxide paper was sufficient for some microfilms. If no blemishes formed, 2 cm^2 more of peroxide paper was added and the test was continued. Most of the blemish forming microfilms required 2 to 4 cm^2 of the peroxide paper.

6. Summary

Microfilms fixed in a neutral or nonhardening acid bath and washed in distilled water or chlorine-free tap water, formed blemishes when exposed to peroxide paper. Microfilms fixed in an acid hardening bath and washed in chlorine-free tap water, also formed blemishes when exposed to peroxide paper but no blemishes formed after washing in distilled water. The distilled water wash did not reduce the concentration of the residual thiosulfate in the microfilm sufficiently to permit blemish formation. The silver sulfide formed in microfilm during fixation affords some protection against oxidation of the image silver by peroxide but its concentration is not large enough to prevent oxidation of the silver after washing in chlorine-free tap water. Residual thiosulfate reacts with the image silver in processed microfilm to form additional silver sulfide, giving a greater resistance to peroxide attack. Tap wash water containing a chlorine concentration of 1 ppm formed a sufficient amount of silver chloride in the image silver to make microfilm resistant to blemish formation when exposed to peroxide paper. One microfilm required a 10-min wash after fixing in a nonhardening acid bath to inhibit blemish formation. The resistance to blemish formation by the silver chloride in the image silver can be removed by treating the processed microfilm in a 0.2 percent solution of ammonium hydroxide. The image silver showed an even yellowing after this treatment and exposure to peroxide paper. Some resistance to blemish formation was found when microfilms were washed 30 min in tap water containing a chlorine concentration of 0.3 ppm. A trace of silver chloride in the image silver promoted blemish formation. Thus the type of fixing bath, concentration of the chlorine in the wash water, and the washing time after fixation can be factors in the formation of blemishes in processed microfilm stored in paper cartons.

Formaldehyde-urea and formaldehyde-melamine type resins are incorporated in microfilm paper storage cartons as wet-strengthening agents. These resins hydrolyze at high relative humidity to give formaldehyde and ammonia. Formic acid may form by the oxidation of the formaldehyde by the peroxide evolved by the storage carton and aerial oxygen in the presence of silver in the image. Microfilm containing sufficient residual silver chloride to prevent blemish formation loses this resistance when exposed to formic acid evolved in storage cartons. It was found that formic acid was capable of reducing silver chloride and the silver chloride ammonia complex in the gelatin layer of microfilm. The silver chloride ammonia complex was more easily reduced than silver chloride, suggesting that ammonia evolved by storage cartons may sensitize the reduction of the silver chloride in the image silver. A correlation has been found between the incidence of blemishes and the brand of storage cartons. Heavy blemish formation was found on a collection of 35 mm step tablets stored with paper sheets for 4 years in a desk drawer in this laboratory. These films had been washed 30 to 45 min in running tap water. The paper sheets were found to evolve formaldehyde and formic acid after 4 years in storage. These results show that chemicals evolved by paper cartons during storage may react with the image silver to form blemishes.

Microfilms fixed in a bath containing iodide ion formed blemishes when exposed to peroxide paper. However, the microfilms showed good resistance to blemish formation after storage in some paper cartons at 86 percent relative humidity. The same microfilms fixed in a bath without iodide showed appreciable blemish formation. The residual silver iodide in the image made the film resistant to attack by chemicals evolved by the storage cartons. These results are consistent with the finding of a lower incidence of blemish formation in stored microfilm containing iodide.

The present theory is that peroxide oxidizes the silver in the image and also reduces the silver ion. The reduced silver may form colloidal silver of the silver-gelatinate type or combine with a silver halide, forming the silver-silver halide type of coloidal silver.

7. Appendix. Formulas for Developing and Fixing the Silver Image in Microfilm

7.1. Developer

Monomethyl *p*-aminophenol sulfate	2.0 g.
Sodium sulfite, anhydrous	90.0 g.
Hydroquinone	8.0 g.
Sodium carbonate, monohydrate	52.5 g.
Potassium bromide	5.0 g.
Water to make	1 liter.

The development time was 3 min.

7.2. Stop Baths

No. 1 for Neutral Fixing Bath:

Treated sample 1 min in each of two 2-liter lots of distilled water.

No. 2 for Nonhardening Acid Fixing Bath:

Sodium bisulfite	60.0 g.
Water to make	2 liters.

Treated sample 2 min.

No. 3 for Acid Hardening Fixing Bath:

Glacial acetic acid	30 ml.
Water to make	2 liters.

Treated sample 30 secs.

7.3. Fixing Baths

Neutral Fixing Bath:

Sodium thiosulfate ($Na_2S_2O_3 \cdot 5H_2O$)	245.0 g.
Water to make	1 liter.

Fixed sample 4 min.

Nonhardening Acid Fixing Bath:

Sodium thiosulfate ($Na_2S_2O_3 \cdot 5H_2O$)	240.0 g.
Sodium bisulfite	25.0 g.
Water to make	1 liter.

Fixed sample 5 min.

Acid Hardening Fixing Bath:

Sodium thiosulfate ($Na_2S_2O_3 \cdot 5H_2O$)	300.0 g.
Sodium sulfite	10.0 g.
Glacial acetic acid	10 ml.
Boric acid	5.0 g.
Potassium aluminum sulfate	10.0 g.
Water to make	1 liter.

Fixed sample 4 min.

7.4. Chemical Specifications

All of the solutions were made up with distilled water. The following chemicals were of reagent quality meeting ACS specifications: sodium sulfite, acetic acid, boric acid, potassium bromide, and potassium iodide. The monomethyl *p*-aminophenol sulfate, hydroquinone, sodium carbonate, sodium thiosulfate, and potassium aluminum sulfate conformed to the USA Standards Institute Specifications for Photographic Grade Chemicals. The sodium bisulfite was analytical reagent grade.

8. References

[1] R. W. Henn and D. G. Wiest, Microscopic spots in processed microfilm: their nature and prevention, Phot. Sci. Eng. **7**, 253 (1963).

[2] C. S. McCamy, Inspection of processed photographic record films for aging blemishes, National Bureau of Standards Handbook 96 (U.S. Government Printing Office, Washington, D.C., 1964).

[3] C. S. McCamy and C. I. Pope, Summary of current research on archival microfilm, National Bureau of Standards Technical Note 261 (U.S. Government Printing Office, Washington, D.C., 1965).

[4] C. S. McCamy and C. I. Pope, Current research on preservation of archival records on silver-gelatin type microfilm in roll form, J. Res. NBS **69A** (Phys. and Chem.) No. 5, 385 (1965).

[5] C. E. K. Mees, The Theory of the Photographic Process, p. 103, (The Macmillan Co., New York, N.Y., 1954).

[6] Charles N. Durfor and Edith Becker, Public Water Supplies of the 100 Largest Cities in the United States, 1962, U.S. Geological Survey Water-Supply Paper 1812 (U.S. Government Printing Office, Washington, D.C.).

[7] W. A. Taylor and F. R. McCrumb, Modern *p*H and Chlorine Control, p. 42 (W. A. Taylor & Co., Baltimore, Md. 1931).

[8] William J. Ryan, Water Treatment and Purification, p. 197 (McGraw-Hill Book Co., New York, N.Y., 1946).

[9] J. R. Partington, A Text-Book of Inorganic Chemistry, pp. 334 and 804 (The Macmillan and Co., Limited, London, 1933).

[10] J. W. Laist, Copper, Silver and Gold, Vol. **II**, Comprehensive Inorganic Chemistry, p. 155 (D. Van Nostrad Company, Inc., New York, N.Y. 1954).

[11] W. M. Latimer and J. H. Hildebrand, Reference Book of Inorganic Chemistry, p. 24 (The Macmillan Co., New York, N.Y. 1935).

[12] C. I. Pope, Formation of silver sulfide in the photographic image during fixation, J. Res. NBS **64C** (Engr. and Instr.) No. 1, 65 (1960).

[13] The Condensed Chemical Dictionary, 3d edition, p. 592 (Reinhold Publishing Corp., N.Y., 1942).

[14] R. W. Henn, D. G. Wiest, and B. D. Mack, Microscopic spots in processed microfilm: The effect of iodide, Photo. Sci. Eng. **9**, 167 (1965).

[15] Eugene Sawicki, T. R. Hauser, and Sylvester McPherson, Spectrophotometric determination of formaldehyde and formaldehyde-releasing compounds with chromotropic acid, 6-amino-1-naphthol-3-sulfonic acid (J acid), and 6-anilino-1-naphthol-3-sulphonic acid (phenyl J acid), Anal. Chem. **34**, 1460 (1962).

(Paper 72A3–499)

PHOTOGRAPHIC SCIENCE AND ENGINEERING
Volume 13, Number 5, September–October 1969

A Simplified Method for Determining Residual Thiosulfate in Processed Microfilm*

C. I. Pope, *Image Optics and Photography Section, National Bureau of Standards, Washington, D.C. 20234*

In the silver nitrate test for residual thiosulfate in processed film the silver ion reacts quantitatively with thiosulfate to form silver sulfide *in situ*. The excess silver ion is removed by treating the sample in solutions of sodium chloride and sodium thiosulfate followed by washing. The method is simplified by using a solution of ammonium hydroxide and sodium chloride to remove the excess silver ion; no washing is necessary. The new method takes half the time previously required.

Residual thiosulfate in processed photographic papers is usually determined by treating the paper with silver nitrate, which reacts with the reactive sulfur in thiosulfate to form silver sulfide in situ.[1,2] The silver nitrate also reacts with the reactive sulfur in any trithionate or tetrathionate present to form silver sulfide. Mattey and Henn[3] adapted the silver nitrate test to determine the residual thiosulfate in processed film. They removed the excess silver ion by treating the sample in a sodium chloride solution, then a sodium thiosulfate solution, and then washing. Any excess silver ion left in the film would darken and increase the density of the film. The purpose of this study was to simplify the processing of the sample after treatment in the silver nitrate solution. It was found that a 1% ammonium hydroxide solution containing 5 g of NaCl per liter adequately removes the silver ion, and no washing is required. This treatment should be effective on all types of silver films, but its effectiveness on papers is unknown.

Experimental

Processed microfilm test samples were prepared by treating 20-cm strips of unexposed film 1 min in a developer solution, fixing them for 4 min in a thiosulfate fixing solution, washing ½ to 1½ min in chlorine-free tap water, removing the surface water immediately with wiping tissue, and drying. The film strips were treated 4 min in a solution containing 10 g of silver nitrate and 30 ml of glacial acetic acid per liter. In some test strips the excess silver ion was removed by treating them for 4 min in a solution containing 45 g of sodium chloride per liter, then for 4 min in a solution containing 15 g of sodium sulfite and 45 g of sodium thiosulfate per liter, and washing according to the recommendations of Mattey and Henn.[3] Control samples not treated in the silver nitrate solution were placed in the sodium chloride solution. In other test strips the excess silver ion was removed by treating them for 2 min in a 1% solution of ammonium hydroxide containing 5 g of sodium chloride per liter, removing the surface solution with wiping tissue and drying. Control samples not treated in the silver nitrate solution were placed in the ammonium hydroxide-sodium chloride solution. The transmission densities of the precipitated silver sulfide in the gelatin layer was read on a commercial densitometer equipped with a Kodak Wratten No. 18A filter.

Several tests were made by the two procedures for concentrations of residual sodium thiosulfate ranging from 0.1 to 15 $\mu g/cm^2$ in processed microfilm. The transmission densities were generally slightly higher for the ammonium hydroxide-sodium chloride fixation. For example, the transmission densities of the silver sulfide in three test samples of single layer film were 0.05, 0.13, and 0.21 after treatment in the ammonium hydroxide-sodium chloride solution but were, respectively, 0.04, 0.12, and 0.20 after treatment in the sodium chloride solution and the sodium sulfite-sodium thiosulfate solution. The cause of this difference is not known. Prolonged treatment of the samples in the sodium chloride solution or in the sodium sulfite-sodium thiosulfate did not decrease the transmission density of the silver sulfide, indicating no removal of the silver sulfide by these solutions. However, calibration curves may be prepared for either procedure by plotting the residual sodium

Received March 24, 1969; revised May 20, 1969.

1. J. I. Crabtree, G. T. Eaton, and L. E. Muehler, A review of hypo testing methods, *J. Soc. Motion Picture Engrs.* **42**: 34–57(1944).
2. American Standard Method for Determining Residual Thiosulfate and Thionates in Processed Photographic Papers, PH4.30-1962 (United States of America Standards Institute, 10 East 40th Street, New York, N. Y. 10016).
3. D. A. Mattey and R. W. Henn, Determination of thiosulfate and thionates in film with silver nitrate, *Photogr. Sci. Eng.* **10**, No. 4, 202–208 (1966).

thiosulfate content against the density of silver sulfide in processed microfilm of known sodium thiosulfate content.

Test samples treated in the ammonium hydroxide-sodium chloride solution did not show a measurable increase in transmission density when aged 16 hr at 50°C or treated 4 min in a developer solution, washed and dried. These tests indicated that the residual silver ion had been adequately removed.

Discussion

The simplified procedure for the residual thiosulfate test in processed microfilm uses two solutions instead of three and washing of the test sample is eliminated. The procedure should be applicable to other types of film. Thicker films may necessitate increasing the ammonium hydroxide concentration or the time of fixation in the ammonium hydroxide-sodium chloride solution. A calibration curve may be made by the procedure given by Mattey and Henn[3] by either the methylene blue test[4] or the modified Crabtree-Ross test.[5]

4. C. D. Warburton and E. P. Przybylowicz, A new test method for the measurement of residual thiosulfate in processed film based on borohydride reduction to sulfide and methylene blue formation, *Photogr. Sci. Eng.* **10**: 86–92(1966).
5. C. I. Pope, Determination of residual thiosulfate in processed film, *J. Res. NBS 67C* (Engr. and Instr.), **3**: 237 (1963).

Redox Blemishes—Their Cause and Prevention*

C. S. MCCAMY AND C. I. POPE

C. S. McCamy

Mr. McCamy received the Bachelor of Chemical Engineering degree and Master of Science degree in physics from the University of Minnesota. He taught mathematics at the University of Minnesota (1947-1950) and physics at Clemson University (1950-1952). He joined NBS in 1952, was Chief of the Photographic Research Section (1958-1966), and the Image Optics and Photography Section since that time. His publications include contributions in flame physics, radiometry, theory of color filters, color vision, photometry, densitometry, photographic image evaluation, and preservation of photographic records. His laboratory provides reference standards and calibrations in photography, calibrations of aerial mapping cameras and other optical instruments, and precise measurements of refractive index, and conducts associated research. He is a consultant to all agencies of Government and is a member of the National Research Council. He has participated in developing numerous national and international standards for photography, in the American National Standards Institute, the American Society for Testing and Materials, and the International Organization for Standardization. He is a member of the Royal Photographic Society and the American Society of Photogrammetry, Fellow of the Society of Photographic Scientists and Engineers, in which he has held chapter and national offices, Fellow of the Optical Society of America, and Fellow of the Washington Academy of Sciences.

Chester I. Pope

Mr. Chester I. Pope received the A.B. degree in chemistry from the University of Kansas in 1935 and the A.M. degree in chemistry from George Washington University in 1939. Since 1936 he has conducted research on the chemistry of paper and the photographic process at the National Bureau of Standards. His publications include a new light-sensitive resist for etching fine lines on glass, preservation of archival records, the chemistry of photographic fixing, washing, and fading, and blemish formation on film. He has participated in developing standards for photographic films and chemicals. He is a member of the American Chemical Society, Fellow of the American Institute of Chemists, and member of the National Association for the Advancement of Science. In 1969 he was awarded the Silver Medal of the U. S. Department of Commerce for "outstanding contributions to the chemistry of preservation of archival records on photographic film".

*Presented at the National Microfilm Association in Boston, Massachusetts, May 7, 1969.

REDOX BLEMISHES, THEIR CAUSE AND PREVENTION

ABSTRACT

A type of small spots and character-associated defects on some microfilm were investigated by a variety of techniques. The blemishes result from the displacement of image silver by an oxidation-reduction reaction caused by peroxides and other gaseous products of degradation of the paper cartons in which the films are stored. A method of producing blemishes for test purposes was developed. Recommendations cover the materials, processes, and storage conditions for microfilm preservation of records of permanent value. KEY WORDS: archival records, microfilm, redox blemishes.

In 1963, NBS received, from Kodak Research Laboratories, reports of spots and other blemishes of unknown cause on microfilms in private and public collections. In the next few months, several thousand rolls of microfilmed Government records were inspected by NBS personnel. Blemishes of six general types were observed. Colored photomicrographs of these blemishes were published in NBS Handbook 96, which described procedures for statistical sampling of collections, microscopic examination of films, and reporting of observations.[1]

The possibility of biological cause of blemishes was thoroughly investigated, not only at NBS but at the Armed Forces Institute of Pathology and Kodak Research Laboratories. No evidence of biological activity was found and all attempts to induce blemish formation by biological incubation failed.

The National Microfilm Association shared our interest in this problem and solicited funds to support research. The NMA hired Mr. Donald Lehmbeck to work for one year at NBS as a research associate.

About 100 inspectors were trained to inspect the microfilm collections in the various government agencies. They examined 7,411 rolls of film and answered about 50 questions on each. The Bureau of the Census encoded the 370,000 observations on punched cards. The correlations were analyzed and the statistics were tabulated in the NBS computation laboratory, with the assistance of the Statistical Engineering Section. Mr. R. E. Wiley and Miss J. A. Speckman conducted the analysis.[2]

Mechanism

Microscopy, electron microscopy, electron probe analysis, and several chemical tests all indicated that the normal silver grains were dissipated and replaced by general deposits of colloidal silver, silver-gelatinate, and metallic silver. The silver was apparently oxidized to an ionic form, which was free to migrate, and reduced again to silver. In 1963, Henn and Wiest proposed this oxidation-reduction mechanism and suggested peroxide reactions, among others, but had not identified the oxidizing and reducing agent or its source.[3] As a preventive measure, they proposed treating films in a gold chloride solution to deposit a protective gold layer on the silver grains.[3,4]

We subjected microfilms to a variety of likely dry chemical contaminants and to various sources of peroxides. Consideration of many possible sources of peroxides led to the hypothesis that the principal source was the cardboard boxes in which the films were stored. Marraccini and Kleinert had found that measureable amounts of peroxides form in the natural aging of commercial pulp papers.[5,6,7] They found peroxide formation to increase with humidity and the rate of generation of peroxides to increase linearly with aging. The field survey revealed that the incidence of blemishes likewise increased with humidity and with the time of storage. Significantly greater numbers of blemishes were associated with the use of cardboard or fiber reels, rather than metal or plastic ones. Except for films known to have been poorly processed, those stored in metal cans were almost entirely free of blemishes.

To test the hypothesis that peroxides from the paper attacked film, films were stored at 50°C and 81% relative humidity with the emulsion in contact with the cardboard carton. These films developed blemishes in 14 to 21 days.

Although much research remained to be done, at the request of the NMA, an interim report, NBS Technical Note 261, was issued in April 1965 to make these results immediately available to other government agencies and the microfilm industry.[8] This publication was soon followed by a more detailed paper in the NBS Journal of

Research.[9] The recommendations given in those two publications were conservative, but are generally valid in view of the findings since that time.

Test Method

One of the objectives of our work was the development of a procedure for measuring the tendency of films to form blemishes. Such a procedure would have direct practical applications and would be of great value in research on blemish formation. Pope found a convenient source of peroxide for such test purposes. Small pieces of paper, such as is used for chromatography, were immersed a half hour in 5% hydrogen peroxide, and dried. When films were placed in a jar maintained at a relative humidity of 86% at room temperature, at a distance of about 8 cm from a 6 cm^2 piece of such paper, blemishes formed. All of the naturally occurring types of blemishes were produced by this method, the type depending on the film processing conditions and the storage conditions prior to testing. At first there was no doubt that this procedure would be a good practical test for the tendency of a film to form blemishes, but soon a question arose.

Henn and Wiest found that films fixed in solutions that happened to contain small concentrations of iodide ion resisted blemish formation in actual storage.[10] Such films, however, did form blemishes in the laboratory exposure to peroxide-treated paper. This suggested that there were other factors involved and led to the investigation of other gases evolved by real paper cartons. It was found that most commercially available cartons released measurable quantities of formaldehyde and formic acid. These chemicals were detected in water stored for one day in an open petri dish inside a storage box. Further laboratory work confirmed the finding that films fixed in a sodium thiosulfate solution containing a small amount of iodide ion do resist attack when stored at high humidity in most, but not all, commercially available paper containers. This work emphasized the fact that the susceptibility of films to blemish formation in the presence of peroxides is conditioned by many factors, including the presence of other gases evolved by carton stock, the concentration of chlorine in the wash water, and the storage conditions before testing. The availability of the peroxide-paper technique has greatly accelerated our research, by providing results quickly under controlled conditions.[11]

Stability of the Silver Image

The silver grain, as formed in the gelatin matrix during development, is inherently unstable.[12] The silver develops as very fine filaments only a few atoms in diameter. In the normal photographic process, these filaments become coated with sulfur atoms during fixation. It is this coating which stabilizes the grain structure. Films fixed a relatively long time and films retaining a small amount of hypo after washing acquire more sulfur and are quite resistant to blemish formation when they are exposed to peroxides at high humidity.

Pope had shown, some years ago, that a small amount of iodide in the fixer prevents the sulfiding of the image.[13] Apparently an iodide layer is formed instead of a sulfide layer. According to James, the iodide layer stabilized silver grains.[12] This corresponded well with the field experience with films processed in a fixer containing iodide ions. We have developed the technique of cutting a thin transverse section of film having a thickness less than a wavelength of light. Electron micrographs of such thin sections of blemishes, at a magnification of 750,000 diameters, reveal the division of the silver filaments into numerous smaller particles of the form observed in experimental emulsions by James.

When silver is oxidized and reduced in the presence of chloride ions, the reduced silver becomes incorporated with silver chloride in a colloidal form. Pope demonstrated the formation of reddish colored colloidal silver by the reaction of hydrogen peroxide with pure metallic silver in the presence of a small concentration of chloride ion. This colloid is yellowish or reddish in color. We attribute the characteristic color of the blemishes to the presence of this colloidal material. Spectrophotometry and microspectrophotometry over the spectral range from 350 to 750 nanometers confirmed the close similarity of the spectral absorption of blemished areas and experimental coatings bearing colloidal silver. The fact that the formation of this form of silver is associated with chlorine suggested an investigation of what effect the concentration of chlorine in the wash water had on the stability of the film. Films washed in distilled water were faded by peroxide attack but typical blemishes did not

form. A small concentration of chlorine appeared to be essential to the formation of typical blemishes, but a concentration considered high for tap water, over one part per million, had an inhibiting effect. The mechanism of inhibition was probably the same as that involving iodide.

The distributions of silver, sulfur, and chlorine in typical blemishes were measured by the technique of electron probe analysis. The emulsion was mounted on a titanium disk, the blemish was scanned with the electron beam, and a characteristic x-ray emission line was monitored. We used the $L\alpha_1$ line of silver, the $K\alpha_1$ line of sulfur, and the $K\alpha$ line of chlorine. Silver and chlorine concentrations tended to correlate with the morphology of the blemishes. Silver was usually somewhat depleted from the central area and concentrated at the edges but often a silver speck appeared at the center. Sulfur distributions did not follow a fixed pattern, but sometimes increased in the spots. White and Weston of E. I. du Pont de Nemours did such analyses for silver and sulfur on about 50 of our samples and corroborated our findings.[14] All of the probe results are consistent with our theory of blemish formation.

Morphology

The circular ring patterns in blemishes are typical of the patterns produced in chemical reactions in a gelatin medium, patterns known as Liesegang rings.[15,16] We repeated the classical Liesegang experiment, placing a drop of silver nitrate solution on a gel layer containing potassium dichromate. Unlike Liesegang, we used a microscopic droplet and, as a consequence, the ring pattern was not of the usual large size but the size of a Type 1 blemish, about 0.1 mm in diameter. In another experiment, 8 concentric rings formed in 30 days under the action of a single minute droplet of a metol-hydroquinone developer on a gel containing silver sulfide. Thus the Liesegang phenomenon is demonstrable on the small scale encountered in blemishes and is adequate to explain the ring structure. On the other hand, this explanation may not be necessary, because in some experiments, Pope has controlled the number of rings by exposing films to a series of mild attacks of peroxide.

Why some blemishes should conform to characters and other imagery as they do has not been investigated in detail, although experiments with hardening and non-hardening fixers demonstrated a connection between edge effects and gelatin hardness. It is well known that the hardness of the gelatin layer is greatly affected by the density of the silver image, as are the concentrations of residual chemicals. The gas permeability of the gelatin may also be modified by the image. These factors are adequate to explain the image-dependence of the blemishes.

The very diffuse Type 6 blemishes usually had foreign matter at the center. Some of these central specks were silver. Others appeared to be flecks of black film backing which had been removed in processing and had become embedded in or adhered to the gelatin. Even if small particles do not penetrate the surface of the gelatin, they may act as small condensation centers where the effective humidity exceeds that in the ambient air and gaseous absorption is accelerated. Minute pores in the gelatin may act in the same way.

Density

Generally speaking, in the field and in the laboratory, the incidence of spots tended to increase with the optical density of the image. In one typical experiment a step tablet exposed to peroxide-treated paper developed blemishes on all steps having densities of 0.58 or more, while none formed at densities of 0.43 or less. The incidence increased for each step above 0.58. Blemishes do not form on microfilms having dark characters on a light background, even though the same films have blemishes in dense areas around frames. Image silver catalyzes the reactions which we have described, so the relationship of blemish formation to density should be expected.

Observations Supporting the Theory

Sometimes one observes a silver sheen on high density areas. This is a deposit of reduced silver and silver sulfide. The observation of such a sheen was statistically associated with a great increase in blemish formation. The suggested oxidation-reduction reaction accounts for both effects.

Spots were very frequently found on the outer end of the roll and at the edges. When there was no leader on the film, the incidence of blemishes on information frames was 2 to 3 times that on rolls with leaders. The incidence began to drop when the leader length exceeded 1.5 meters. These effects should be expected if peroxides

generated in the paper carton cause blemishes, because the areas nearest the paper are most affected. Peroxides are notably short-lived species so they are active only a short distance from a feeble source. Fewer blemishes were found in well-ventilated storage areas and there were fewer in active files than in dead files. All this evidence favors the theory that blemishes are caused by paper-generated peroxides that concentrate in the still air in the storage containers.

Type 1 blemishes very commonly form in a continuous line along a scratch in the film surface. The break in the gelatin exposes the silver more directly to gaseous reactants.

In all of the experiments, humidity has been an important factor. Temperature has been less important but significant. In the survey, we found a significantly lower incidence of blemishes in air-conditioned storage areas.

The blemish rate reported in the survey was directly related to reports of discoloration, fading, water spots, and other observed defects. When any chemical residue was found, the blemish rate tripled and the rate on information areas was high. The differences in blemishing on various brands of film was not significant, but such was not the case with processing equipment. Practically all film processed on one particular kind of machine eventually developed blemishes. In fact, the data from government agencies which had used that machine so out-weighed the other statistics that it became necessary to run a second analysis without these data. In one agency, this type of processor had been replaced by another type, years ago, and the film processed by the new machine has not become blemished. The troublesome machine had small tanks, inadequate squeegee action, and quick washing. It was customary to load dry chemicals into the tanks without first dissolving them. As one might expect, chemicals were left on the film. Such machines are no longer on the market.

Aperture Cards

Reports on films mounted in aperture cards indicate a rather low incidence of blemishes. However, one should bear in mind that such forms have not been in use as long as the roll form. It is reasonable to suppose that the better ventilation and the fact that the film surfaces do not usually touch one another are beneficial. Our findings would suggest caution in the choice of papers for use in aperture cards.

"Redox"

The blemishes have gone by a variety of names, "microspots", "measles", etc. One never knows what to call a thing until he knows what differentiates it from other similar things. We have called them "aging blemishes", but now that we can produce them in the laboratory in a few hours, that name no longer seems appropriate. The strength of the evidence for an oxidation-reduction mechanism is sufficient to warrant calling them "redox blemishes". "Redox" is a standard term in chemistry for reactions of this kind.

Conclusions

One of the most important results of the study is the clear demonstration of the fact that judgements of the suitability of materials for the preservation of permanent records cannot be made with respect to the materials considered independently. The archival process must be regarded and evaluated as a complete system including materials, procedures, and storage conditions.

In concluding this study, we must recommend appropriate action to archivists. All evidence adduced in past research, recent research, and in practice indicates that images on silver-gelatin materials which are properly processed and stored, will last as long as records on high-grade record paper. In addition to the obviously necessary safeguards against damage or destruction of records by flood, fire, riots, and other such incidents, we recommend that archivists take the following precautions to guard against blemish formation on silver-gelatin type films for the preservation of records of permanent value:

1. Use safety base permanent record film as specified in the United States of America Standards Institute Specifications for Photographic Films for Permanent Records.
2. Use no higher densities than are required for the intended purposes and use dark characters on a light background if this is feasible.
3. The residual thiosulfate concentration should not exceed 1 microgram per square centimeter, but one should not attempt to reduce it to zero. The optimum concentration appears to be about 0.5 microgram per square centimeter in a clear area.
4. Keep processing machinery and film clean.
5. Avoid scratching film.

6. Store films in containers made of inert materials, such as metals or plastics of proven quality. With good ventilation and clean air, the containers need not be sealed.
7. Do not permit storage temperature to exceed 70° F nor the relative humidity to exceed 40%.
8. Avoid wide-range cycling of temperature and humidity, since these accelerate the imbition of gaseous contaminants.
9. Inspect films on a regular schedule, using proper sampling and microscopic inspection techniques.
10. Maintain records of processing and storage conditions so that the next generation of archivists will be able to understand our successes or failures and act on the basis of facts.

The authors are grateful to the personnel of the many government agencies that cooperated with us and the National Microfilm Association for its support.

REFERENCES

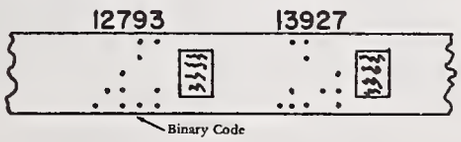
1. C. S. McCamy, *"Inspection of Processed Photographic Record Films for Aging Blemishes,"* Handbook 96, National Bureau of Standards, Washington, D. C. (1964).
2. C. S. McCamy, S. R. Wiley, and J. A. Speckman, *"A Survey of Blemishes on Processed Microfilm,"* J. of Research, NBS 73A, 79 (1969).
3. R. W. Henn and D. G. Wiest, *"Microscopic Spots in Processed Microfilm: Their Nature and Prevention,"* Phot. Sci. Eng. 7, 253 (1963).
4. R. W. Henn and Bernadette D. Mack, *"A Gold Protective Treatment for Microfilm,"* Phot. Sci. Eng. 9, 378 (1965).
5. L. M. Marraccini and T. N. Kleinert, *Spectrophotometric Estimation of Peroxide in Cellulosic Materials,* Svensk Papperstidning 65, 78 (Jan-June 1962).
6. L. M. Marraccini and T. N. Kleinert, *Aging and Colour Reversion of Bleached Pulps,* ibid, 66, 189 (Jan-June 1963).
7. T. N. Kleinert and L. M. Marraccini, *Aging and Colour Reversion of Bleached Pulps,* ibid, 66, 189 (Jan-June 1963).
8. C. S. McCamy and C. I. Pope, *"Summary of Current Research on Archival Microfilm,"* NBS Tech. Note 261 (1965).
9. C. S. McCamy and C. I. Pope, *"Current Research on Preservation of Archival Records on Silver-Gelatin Type Microfilm in Roll Form,"* J. of Research, NBS 69A, 385 (1965).
10. R. W. Henn, D. G. Wiest, and Bernadette D. Mack, *Microscopic Spots in Processed Microfilm: The Effect of Iodide,* Phot. Sci. Eng. 9, 167 (1965).
11. C. I. Pope, *"Blemish Formation in Processed Microfilm,"* J. Res. NBS 72A, 251 (1968).
12. T. H. James, *"The Stability of Silver Filaments,"* Phot. Sci. Eng. 9, 121 (1965).
13. C. I. Pope, *"Stability of Residual Thiosulfate in Processed Microfilm,"* J. Res. NBS 67C, 15 (1963).
14. D. R. White and N. E. Weston, *Electron Probe Examination of Aging Blemishes in Microfilm,* Phot. Sci. Eng. 11, 107 (1967).
15. R. E. Liesegang, Naturw. Wichschr. 11, 353 (1896).
16. R. E. Liesegang, Phot. Archiv. 21, 221 (1896).

Correction

In the *"Glossary Standard for Computer Output Microfilm"* published in the last number of the *Journal* (Vol. 3, No. 3., page 140) the illustrations to accompany definitions for Bibary Digital Code, Cine (Motion Picture) Oriented Images and Comic Strip Oriented Images were transposed. The correct definitions and illustrations have been brought together as follows:

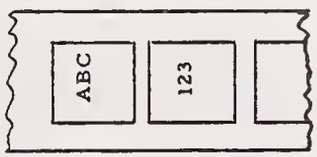
BINARY DIGITAL CODE

An optical pattern of clear and opaque rectangles machine encoded for random access retrieval used to index one or more images.



CINE (MOTION PICTURE) ORIENTED IMAGES

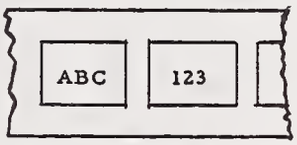
Jargon used with intent to reference images oriented on microfilm as follows:



See American National Standard PH5.3, IA oriented images in Fig. 1.

COMIC STRIP ORIENTED IMAGES

Jargon used with intent to reference images oriented on microfilm as follows:



See American National Standard PH5.3, IB oriented images in Fig. 1.

AUTHOR INDEX

Author Index—Continued

Authors Index—Continued

SUBJECT INDEX

Subject Index—Continued

Subject Index—Continued

T

U

V

W

U.S. GOVERNMENT PRINTING OFFICE : 1973 O - 518-743

FORM NBS-114A (1-71)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBS-SP-300-V. 10	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE

Image Optics

5. Publication Date: November 1973

6. Performing Organization Code

7. AUTHOR(S)
Calvin S. McCamy

8. Performing Organization

9. PERFORMING ORGANIZATION NAME AND ADDRESS

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

10. Project/Task/Work Unit No.

11. Contract/Grant No.

12. Sponsoring Organization Name and Address

Same as 9.

13. Type of Report & Period Covered
Final

14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

16. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)

This volume is one of an extended series which brings together the previously published papers, monographs, abstracts, and bibliographies by NBS authors dealing with precision measurement of specific physical quantities and the calibration of the related metrology equipment. The contents have been selected as being useful to the standards laboratories of the United States in tracing to NBS standards the accuracies of measurement needed for research work, factory production, or field evaluation.

Volume 10 deals with image optics, including photography. It contains 62 reprints assembled in four sections: 1) Refractometry and Optical Homogeniety, 2) Interferometry in Image Optics, 3) Optical Design and Image Evaluation, and 4) Photographic Science. Each section is introduced by an interpretive foreword.

17. KEY WORDS (Alphabetical order, separated by semicolons)
Camera calibration; image evaluation; image optics; image stability; interferometry; lens testing; light filters; light sources; photography

18. AVAILABILITY STATEMENT

[X] UNLIMITED.

[ ] FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NTIS.

19. SECURITY CLASS (THIS REPORT)
UNCLASSIFIED

21. NO. OF PAGES
953

20. SECURITY CLASS (THIS PAGE)
UNCLASSIFIED

22. Price
$11.10

USCOMM-DC 66244-P71

Official SI Unit Names and Symbols

[For a complete statement of NBS practice, see NBS Tech. News Bull. Vol. 55, No. 1, Jan. 1971]

Name	Symbol	Name	
meter	m	newton	N
kilogram	kg	joule	J
second	s	watt	W
ampere	A	coulomb	C
kelvin[1]	K	volt	V
candela	cd	ohm	Ω
radian	rad	farad	F
steradian	sr	weber	Wb
hertz	Hz	henry	H
lumen	lm	tesla	T
lux	lx		

Additional Names and Symbols approved for NBS use

curie[2]	Ci	mole	mol
degree Celsius[3]	°C	siemens[4]	S
gram	g		

[1] The same name and symbol are used for thermodynamic temperature and temperature interval. (Adopted by the 13th General Conference on Weights & Measures, 1967.)
[2] Accepted by the General Conference on Weights & Measures for use with the SI.
[3] For expressing "Celsius temperature"; may also be used for a temperature interval.
[4] Adopted by IEC and ISO.

Table for Converting U.S. Customary Units to Those of the International System (SI)[5]

To relate various units customarily used in the United States to those of the International System, the National Bureau of Standards uses the conversion factors listed in the "ASTM Metric Practic Guide", designation: E 380–70. These are based on international agreements effective July 1, 1959, between the national standards laboratories of Australia, Canada, New Zealand, South Africa, the United Kingdom, and the United States.

To convert from:

(1) inches to meters, multiply by 0.0254 exactly.
(2) feet to meters, multiply by 0.3048 exactly.
(3) feet (U.S. survey) to meters, multiply by 1200/3937 exactly.
(4) yards to meters, multiply by 0.9144 exactly.
(5) miles (U.S. statute) to meters, multiply by 1609.344 exactly.
(6) miles (international nautical) to meters, multiply by 1852 exactly.
(7) grains (1/7000 lbm avoirdupois) to grams, multiply by 0.064 798 91 exactly.
(8) troy or apothecary ounces mass to grams, multiply by 31.103 48 . . .
(9) pounds-force (lbf avoirdupois) to newtons, multiply by 4.448 222 . . .
(10) pounds-mass (lbm avoirdupois) to kilograms, multiply by 0.453 592 . . .
(11) fluid ounces (U.S.) to cubic centimeters, multiply by 29.57 . . .
(12) gallons (U.S. liquid) to cubic meters, multiply by 0.003 785 . . .
(13) torr (mm Hg at 0 °C) to newtons per square meter, multiply by 133.322 exactly.
(14) millibars to newtons per square meter, multiply by 100 exactly.
(15) psi to newtons per square meter, multiply by 6894.757 . . .
(16) poise to newton-seconds per square meter, multiply by 0.1 exactly.
(17) stokes to square meters per second, multiply by 0.0001 exactly.
(18) degrees Fahrenheit to kelvins, use the relation $T_{68} = (t_F + 459.67)/1.8$.
(19) degrees Fahrenheit to degrees Celsius, use the relation $t_c = (t_F - 32)/1.8$.
(20) curies to disintegrations per second, multiply by 3.7×10^{10} exactly.
(21) roentgens to coulombs per kilogram, multiply by $2.579\ 760 \times 10^{-4}$ exactly.

[5] Système International d'Unités (designated SI in all languages).