A UNITED STATES DEPARTMENT OF COMMERCE PUBLICATION

NBS SPECIAL PUBLICATION 300
VOLUME 4

Precision Measurement and Calibration

Electricity—Radio Frequency

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards

UNITED STATES DEPARTMENT OF COMMERCE • Maurice H. Stans, Secretary
U.S. NATIONAL BUREAU OF STANDARDS • Lewis M. Branscomb, Director

Precision Measurement and Calibration

Selected NBS Papers on

Electricity — Radio Frequency

Arthur J. Estin, Editor
Institute for Basic Standards
National Bureau of Standards
Boulder, Colorado 80302

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 4

Nat. Bur. Stand. (U.S.), Spec. Publ. 300, Vol. 4, 456 pages, (June 1970)
CODEN: XNBSA

Issued June 1970

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price $5.50

Abstract

This volume is one of an extended series which brings together some of the previously published papers, monographs, abstracts, and bibliographies by NBS authors dealing with the 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 4 contains reprints through June 1967 on radio-frequency electrical measurements covering the following topics: Power, Sinusoidal Voltage and Current, Electromagnetic Fields and Antennas, Radar and Baseband Pulses, Noise, Attenuation and Phase, Impedance, Radio Frequency Materials, Quasi-optics and Millimeter Waves, and Applications to Measurement Systems.

Key Words: Admittance; antenna; attenuator; bolometer; calorimetry; horn; impedance; interferometry; measuring system; phase; power; radiometry; resonant cavity; voltage; waveform; waveguide junction.

Library of Congress Catalog Card 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 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 and references 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 fill responsible positions in standards laboratories.

LEWIS M. BRANSCOMB, *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 12 volumes having the following titles and editors:

Statistical Concepts and Procedures, H. H. Ku
Frequency and Time, A. H. Morgan
Electricity—Low Frequency, F. L. Hermach and R. F. Dziuba
Electricity—Radio Frequency, A. J. Estin
Heat, D. C. Ginnings
Temperature, J. F. Swindells
Mechanics, R. L. Bloss and Mary J. Orloski
Dimensional Metrology—Length and Angle, H. K. Hammond, III
Radiometry and Photometry, H. K. Hammond, III
Colorimetry and Image Optics, I. Nimeroff and C. S. McCamy
Spectrochemical Analysis, B. F. Scribner
Ionizing Radiation, E. H. Eisenhower

This division of subject matter has been chosen to assure knowledgeable selection of content rather than to attain uniform size. It is believed, however, that the larger volumes, of approximately 600 pages, will still be small enough for convenient handling in the laboratory.

The compilation consists primarily of original papers by NBS authors which have been reprinted by photoreproduction, with occasional updating of graphs or numerical data when this has appeared desirable. In addition, some important publications by non-NBS authors, as well as publications by NBS authors that are too long to be included, are represented by abstracts or references; the abstracts are signed by the individuals who wrote them, unless written by the author.

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 the volume, the page numbers in the original publications also being retained and combined with the volume page numbers, for example 133-10. The index notation 4-133 refers to volume 4, page 133 of this volume. A convenient list of SI (Systeme 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 NBS Clearinghouse for Federal Scientific and Technical Information, Springfield, Va. 22151. Reprints from the NBS Journal of Research or from non-NBS journals may sometimes be obtained directly from an author.

Suggestions as to the selection of papers which should be included in future editions will be welcome. Current developments in measurement technology at NBS are covered in annual seminars held at either the Gaithersburg (Maryland) or the Boulder (Colorado) laboratories. These developments are summarized, along with a running list of publications by NBS authors, in the monthly NBS Technical News Bulletin.

H. L. Mason,
Office of Measurement Services,
NBS Institute for Basic Standards.

Editor's Note

This volume of NBS Special Publication 300, together with the one entitled "Electricity—Low Frequency," is an outgrowth of the first volume of Handbook 77, *Precision Measurement and Calibration—Electricity and Electronics* (1961). The present work similarly consists of reprints, abstracts, and references to work in the field of Radio Frequency Measurements. In this, we have included primarily work by the NBS staff, but the results contained herein do not exclusively represent the output of the NBS.

The criteria of selection of papers and the order in which they are presented deserve a brief explanation. In general, those which contribute more general information, such as review papers and conceptually fundamental papers, have been accorded first choice for verbatim reproduction. Other significant papers are given only by abstract, or by title alone if an abstract does not appear in the publication. Some relaxation of this criterion is made for length of the paper vis-a-vis space limitations of the volume. No papers appear if they were included in Handbook 77. Therefore, the present volume is a supplement and not a replacement for the earlier one. To be sure, the state of the art has advanced in the intervening years, and to that extent some of the earlier work may be outmoded or obsolescent. Subsequent papers and the reader's own judgment will settle this question without comment from the editor.

The ten sections in this volume represent an attempt to categorize problems in Radio Metrology. They proceed from the fundamental to the derived; from the active to the passive in a generalized Thevenin sense; from the basic to the applied; from the conceptually simple to the complicated. Thus, we have 1. *Power,* 2. *Voltage and Current,* and 3. *Electromagnetic Fields and Antennas,* all of which are fundamental, sinusoidal, and active; then 4. *Radar and Baseband Pulses,* and 5. *Noise,* these being also active, but nonsinusoidal and somewhat less fundamental. Next, 6. *Attenuation and Phase*—together comprising complex attenuation—and 7. *Impedance* are sinusoidal, but are derived quantities and are Thevenin passive. Finally, in the last three sections, we have passed from the conceptually simple idea of a "quantity" to physical and technological utilization of more than one "quantity" in an application. Hence, we have 8. *Radio Frequency Materials,* 9. *Quasi-Optics and Millimeter Waves,* which is a composite technology, and lastly 10. *Components and Subsystems.*

In no sense should an inference be drawn from this arrangement regarding relative difficulty. This is entirely a subjective matter for the reader and his problem. If he can solve it, it is easy; if not, it is difficult!

Within each section, the *full papers* are arranged in a subject-related sequence consistent with the above groupings. The *abstracts* are alphabetical with respect to first author, and chronological as a sub-category, and a letter, upper or lower case. The upper case letters designate papers reprinted here in their entirety, and the lower case letters those which appear only by abstract or title. An asterisk in either event designates a paper falling into the category of journal "correspondence." Finally, the cutoff date for all publications appearing in this volume is July 1967.

A. J. Estin, *Editor.*

Contents

2. Sinusoidal Voltage and Current

Papers

Abstract

3. Electromagnetic Fields and Antennas

Papers

Abstracts

See also 7.e.

4. Radar and Baseband Pulses

Papers

4. Radar and Baseband Pulses—Continued

5. Noise

See also 7.5.

6. Attenuation and Phase

6. Attenuation and Phase—Continued

7. Impedance

7. Impedance—Continued

* Private communication.

8. Radio Frequency Materials

8. Radio Frequency Materials—Continued

Abstracts

9. Quasi-Optics and Millimeter Waves

Papers

9. Quasi-Optics and Millimeter Waves—Continued

Abstracts

* Correspondence

10. Components and Subsystems

Papers

Abstracts

* Private communication.

1. Power

Papers

Abstracts

See also 7.e.

Radio Frequency Power Measurements

ANNE Y. RUMFELT, MEMBER, IEEE, AND LYMAN B. ELWELL, MEMBER, IEEE

Abstract—**The need for improved accuracy in and understanding of all kinds of measurements has come with the recent rapid advances in modern technology. Radio frequency power measurement is no exception to this requirement. The basic principles of bolometric, calorimetric, and certain other types of power meters are reviewed. The methods for making accurate RF power measurements are discussed in detail. Emphasis is given to the techniques for eliminating or accounting for the errors due to mismatch, dc or LF substitution, and bolometer mount efficiency.**

1. Introduction

As the frontiers of science and technology have been pushed outward, accurate measurements of the quantities used to describe the transmission of electromagnetic energy have become a necessity.

Power is one of these quantities. Power measurements may be used in the determination of current, voltage, impedance, energy, and efficiency [1].

At high frequencies (HF), voltage, current, and power measurements are common. However, power is often a more convenient and a more meaningful quantity to measure because the end use depends as much on the energy as on voltage or current levels. In the microwave-frequency region, voltage and current tend to lose their practical significance, while power is one of the important quantities that can be measured directly.

This introduction to radio frequency (RF) power-measuring techniques reviews briefly the basic methods of CW power measurement in the HF and microwave-frequency ranges. Emphasis is placed on the application of these methods, the calibration of power meters, and the sources of error in the measurement process.

Methods for measuring voltage, current, pulse power, power in millimeter-wave and laser-frequency regions, and swept-frequency techniques are discussed elsewhere in this issue of the PROCEEDINGS, and will not be repeated here.

2. Common Types of Power Meters

Power meters may be classified as either feed-through or terminating. The former type consists of a section of transmission line with provision for sampling the power flow; it absorbs only a small fraction of the transmitted power. The latter, as the name implies, terminates the transmission line and ideally absorbs all the incident power.

Most power meters used at HF and microwave frequencies are of the terminating type. The most commonly used feed-through power meter is the directional coupler with a power detector attached to the side arm. Other feed-through power meters have been designed; for example, radiation-pressure wattmeters and power meters based on the Hall effect in semiconductors. The following discussion includes power meters whose designs are based on calorimetric and bolometric principles, as well as other thermal, mechanical, and a few entirely electronic devices.

2.1 Calorimetric Power Meters [1]–[4]

The underlying mechanism of the calorimetric measurement of power is the conversion of a fraction of the electromagnetic energy into heat. Flow and static calorimeters measure power in terms of temperature change, mass, and time and are used in most frequency ranges from dc to 40 GHz and above. The flow type is ordinarily used to measure medium (10^{-2} to 1 watt) and high (1 to 10^{6} watt) power. The static type is customarily used to measure low (10^{-8} to 10^{-2} watt) and medium power. The uncertainty in the measurement varies between 0.2 and 5 percent depending on power level, frequency range, auxiliary instrumentation, and thoroughness of the analysis of the sources of error.

The advantages of using calorimetry lie in its potentially high accuracy. Its disadvantages come from the complexity of the design and construction and the long time constants (several minutes to several hours).

2.1.1 Static Calorimeters [1], [3], [4]: The basic static calorimeter consists of a thermally isolated load, which converts the absorbed electromagnetic energy into heat, and a device for measuring the load-temperature rise, as shown in Fig. 1. The rate of temperature rise, $\Delta T/\Delta t$, in a thermally isolated body of mass M and of known specific heat C is proportional to the power absorbed by that body. The average power P averaged over the time interval Δt is given by

$$P = kMC\Delta T/\Delta t, \tag{1}$$

where k is a proportionality constant.

In principle, the static calorimeter is simple. However, for accurate calorimetric measurements, the heat capacity MC of the calorimetric body must be accurately known, and the heat lost to the surroundings must be small and corrected for in the measurement.

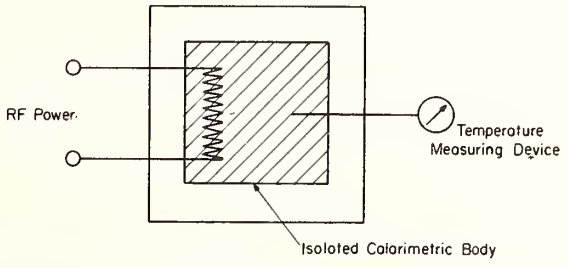


Fig. 1. Basic static calorimeter.

Manuscript received March 1, 1967.
The authors are with the National Bureau of Standards, Boulder, Colo.

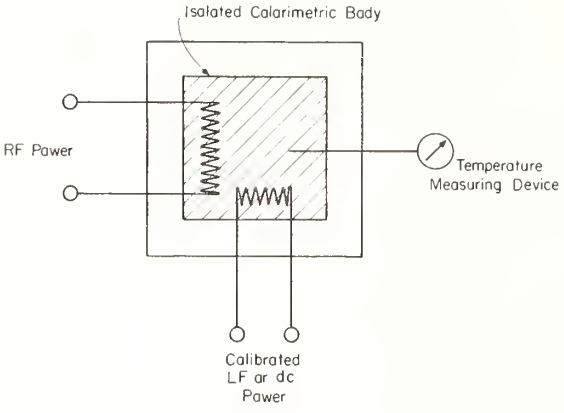


Fig. 2. Basic static substitution calorimeter.

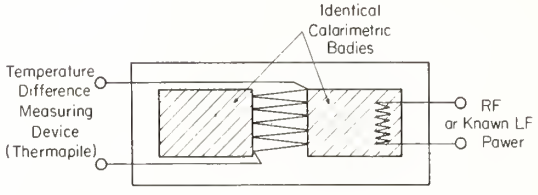


Fig. 3. Basic twin calorimeter.

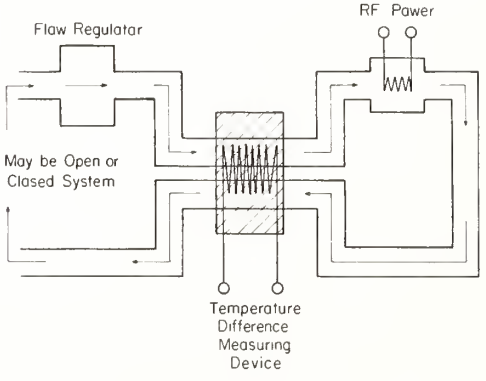


Fig. 4. Basic flow calorimeter.

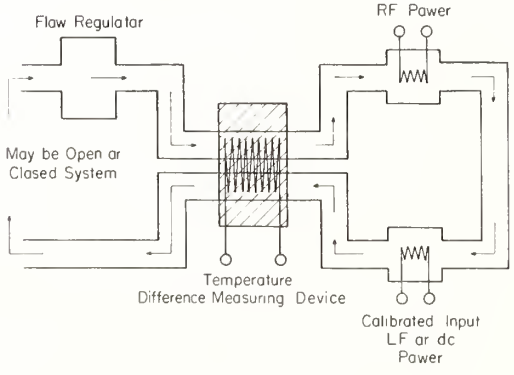


Fig. 5. Basic substitution flow calorimeter.

Some of these problems are avoided by using a known dc or low-frequency (LF) power to calibrate the temperature-measuring device. Then, the need for thermal isolation is reduced, and the requirement for an accurate value of the heat capacity is nominally eliminated. In this substitution method, the calorimeter becomes a transfer device measuring RF power in terms of the accurately measured dc or LF power. The substitution static calorimeter, as it is termed, is identical to the basic static unit except for the addition of the dc or LF heater, as shown in Fig. 2. The assumption that equal quantities of dc or LF power and RF power cause equal heating must be evaluated for each substitution calorimeter.

Another substitution calorimeter, the twin calorimeter, consists of two identical calorimetric bodies partially isolated from their surroundings, as shown in Fig. 3. One of the bodies absorbs the RF power while the other acts as a reference of temperature. The steady-state temperature difference between the two loads is a measure of the RF power. The thermal symmetry of the system reduces the effect of ambient temperature variations, and observation of smaller temperature changes is possible.

2.1.2 Flow Calorimeters [1], [3], [4]: The basic flow calorimeter consists of a load for converting the electromagnetic energy into heat in a liquid, a system for circulating the liquid, and a means for measuring temperature differences within the circulating liquid, as shown in Fig. 4. A liquid of specific heat C and known specific gravity D, flowing at a known rate F, passes through the load, where its temperature T is raised by transfer of heat from the load. The temperature-measuring device is located to measure the difference in temperature ΔT between the liquid entering and leaving the load. The RF power equals the rate of energy absorption, and is given by

$$P = kFDC\Delta T. \tag{2}$$

To make accurate measurements, several parameters must be known; for example, two of these are the flow rate and the specific heat of the circulating liquid over the operating temperature range. In addition, other parameters not accounted for in (2) must be known, such as the rate of certain heat losses from the system and the Joule heating of the flowing liquid due to friction.

As with the static calorimeter, some of these problems are eliminated by using dc or LF substitution. The substitution flow calorimeter consists of the same components as the basic calorimeter with an additional load for the introduction of the dc or LF power, as shown in Fig. 5. Ideally, the substitution method eliminates the need to know the flow rate, temperature scale, density, and specific heat of the liquid.

2.2 Bolometric Power Meters [1]–[5]

Most power meters presently in use have a bolometer as a power-sensing element. A bolometer is a temperature-sensitive resistance element. Its change in resistance, due to heating by the absorbed RF power, is measured with an external meter or bridge circuit [4]. More often, the ele-

ment's resistance is maintained constant by means of dc or LF substitution power, as discussed in Section 2.2.6. Bolometric power meters have a relatively high sensitivity and are used to measure low power levels (10^{-6} to 10^{-1} watt). The element is installed in a waveguide mount that has been designed for minimum reflections. A lead, insulated from the waveguide, is accessible for the resistance measurement; the waveguide acts as the second lead.

The uncertainty in a bolometric measurement of RF power varies from 0.5 to 3 percent depending upon the mount efficiency and impedance match, and the auxiliary equipment used to measure the resistance change of the element. The problems associated with the impedance match and efficiency are discussed in Sections 3 and 4, respectively.

The advantages of bolometer power meters in general are their small size and consequent portability, ease of application, and relatively high sensitivity. Their main disadvantage is that they must conduct the energy they absorb to their surroundings and thus are difficult to isolate from external temperature changes. Temperature compensation is provided in some power meters to counteract the effect of changes in ambient temperature. This usually involves the use of a second bolometer element and dual bridge circuits. The temperature sensitivity (i.e., change in power reading versus change in ambient temperature) of these compensated power meters is typically one or two orders of magnitude less than for uncompensated meters.

A few types of bolometers are: the barretter, the thermistor, the film bolometer, the load lamp, and the waveguide wall bolometer. Among these, the use of barretters and thermistors predominates.

2.2.1 Barretter [2]–[4]: The barretter consists of a short length of thin Wollaston resistance wire, mounted in a dielectric capsule with metal end-caps. It is characterized by a positive temperature coefficient, a short time constant (50 to 400 μs), and a linear change in resistance for low RF power levels. It is used to measure low RF power levels directly. The disadvantages of the barretter are its mechanical fragility and the ease with which it is destroyed by power overload.

2.2.2 Thermistor [2], [4]: The second commonly used bolometer element is the thermistor. It consists of a small bead of semiconducting material with thin lead wires. The thermistor is distinguished from the barretter by its negative temperature coefficient and its much longer time constant (approximately 0.1 second). It has a wide range of operating resistance. It is mechanically rugged and difficult to damage or destroy by power overloading.

2.2.3 Thin-Film Bolometer [4]: The thin-film bolometer consists of a thin metallic film vacuum deposited on a glass or mica substrate. The film can be tailored in shape and resistance to present a good impedance match to the waveguide. Since these resistive elements are extremely thin, the skin depth at the highest frequencies is greater than the film thickness, and since they are physically short compared to the wavelength, the current distribution is nearly uniform. Consequently they are assumed to have the same RF and dc resistances. The time constant of one of these units is about 5 seconds. Power levels up to about 100 mW can be measured.

2.2.4 Load Lamp [2]–[4]: The load lamp is similar to the barretter except that a much larger wire size is used for a greater power-handling capability. They may be used in bridge circuits, but are more often used with optical detection schemes which measure the brightness of the wire. The accuracy of the load lamp is quite poor; matching its impedance to the transmission line is difficult, and the error due to the skin effect can be large. The load lamp may be used to indicate roughly RF power levels in the 100-mW to 20-W range.

2.2.5 Waveguide Wall Bolometer [4], [6], [7]: The waveguide wall bolometer is also known as the enthrakometer [6]. It consists of a temperature-sensitive resistive film deposited on an insulating substrate that forms a part of the narrow wall of a rectangular waveguide. The resistive film absorbs only a small fraction of the RF power. The resulting change in film resistance is measured using bridge techniques. This is a feed-through power meter that may be used to measure RF power levels from 1 mW to 1 W over a frequency range of 1 to 40 GHz with a measurement uncertainty of 2 to 5 percent.

2.2.6 Bridges for Accurate Bolometric Power Measurement: One could measure the resistance change of a bolometer element and calculate the power using the resistance law of the bolometer element, which relates the operating resistance r to the absorbed power P. For the barretter the resistance law is [3]

$$r = r_0 + JP^n, \tag{3}$$

where r_0 is the resistance of the element without bias current, and J and n are characteristic constants of the element. However, this technique is not used to measure RF power because r_0, J, and n are not constant with time, and, more important, because the impedance of such a power meter varies with power level. These sources of measurement uncertainty can be avoided by maintaining the bolometer element at a constant resistance value.

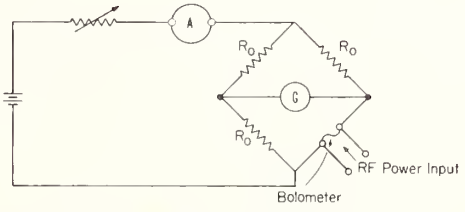


Fig. 6. Basic bolometer bridge circuit.

The bolometer resistance is kept constant by connecting it as one arm of a Wheatstone bridge, as shown in Fig. 6. The RF power (or substituted dc power) is given by

$$P_{RF} = \frac{R_0}{4}(I_1^2 - I_2^2), \tag{4}$$

where I_1 and I_2 are the dc currents required to balance the bridge without and with RF power present, respectively.

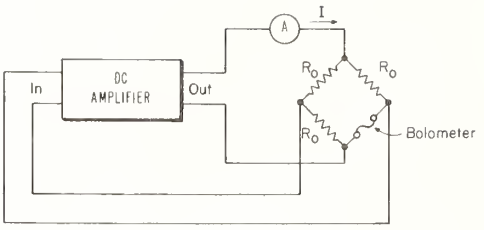


Fig. 7. Basic circuit of self-balancing dc bolometer bridge.

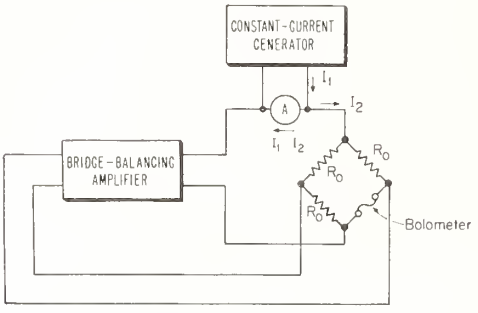


Fig. 8. Application of constant-current generator to power-measurement problem.

One of the difficulties with the manual bridge is that the operator must take the precaution to reduce the dc bias power before applying the RF power to the bolometer element, in order to avoid damaging or destroying the bolometer element with excessive power.

For convenience, protection of the element, and in some cases better accuracy, a self-balancing bridge is generally preferred. Both dc and audio-frequency (AF) self-balancing bridges are used. The self-balancing dc bridge [8] uses a stable high-gain dc amplifier in the feedback loop to balance the bridge, as shown in Fig. 7.

Power levels measured range from about 100 μW to about 10 mW. The uncertainty in a power measurement using either a manual or a self-balancing dc bridge ranges from about 0.2 to 5 percent, depending upon the power level and the dc system limitations. The accuracy deteriorates at the lower power levels, where an accurate measurement of a small change in a relatively large dc power is required.

Low power measurements (below 5 mW) can be improved by using a constant-current generator in conjunction with a bolometer bridge. Equation (4) may be rewritten in the form

$$P_{RF} = \frac{R_0}{4}(2I_1 - \Delta I)\Delta I, \tag{5}$$

where $\Delta I = I_1 - I_2$. A high-stability constant-current generator [8] connected as shown in Fig. 8 permits a direct and accurate measurement of ΔI. Its output is adjusted to equal I_1 without RF power. When the RF power is applied, the meter A will read ΔI directly. The RF power is then calculated by (5).

2.3 Other Types of Power Meters

2.3.1 Thermocouple and Thermoelement Power Meters [2], [4]: There are two basic types of thermocouple devices used to measure RF power: a directly heated unit in which the RF current passes through the thermocouple, and an indirectly heated unit, called a thermoelement, in which the RF power heats a resistive film or wire, and the thermocouple measures the temperature rise. Directly heated thermocouple detectors are available for frequencies to about 18 GHz, while thermoelements have been designed for frequencies as high as 40 GHz [9].

The thermoelement response and stability are improved by enclosing the element in an evacuated envelope [10]. Vacuum elements have been built to match 50-ohm transmission lines at fixed frequencies in the range of 10 to 1000 MHz. They are used to measure from 1-mW to 5-mW RF power directly with an uncertainty of one percent. Directional couplers are used to extend the power range to as high as 1000 watts [10].

The advantage of the two types of thermocouple detectors is the simplicity of the indicating equipment, the ease with which they can be calibrated using dc or LF substitution techniques, and their ability to take a 50-percent overload without damage. Their disadvantages are low sensitivity, poor RF impedance match, and nonlinearity of power versus voltage indication.

In general, the time constant of these detectors ranges from 0.1 to 5 seconds, and the directly measured power varies from 1 to 150 mW with an uncertainty of one to two percent.

2.3.2 Crystal-Diode Power Meter [4]: The crystal diode is one of the commonly used power detectors. The diode detects the RF voltage across the transmission line. The power level may be read directly from a meter with a nonlinear scale. Meters using crystal-diode detectors are rugged, have a wide dynamic range (microwatts to milliwatts), have a high power sensitivity (5 mV/μW into a high impedance load), and a fast response time. The uncertainty of a crystal-diode power meter may vary from 1 to 20 percent. It is not considered suitable for a laboratory standard because crystal characteristics change with time, ambient temperature, and mechanical shock. Its principal use is as a detector or indicator of the presence of RF power in a transmission line.

2.3.3 Mechanical and Hall Effect Power Meters:

a) Mechanically actuated power meters: The mechanical RF power meters include radiation pressure [11], [12], torque-vane [13], [14], and vibration [5], [15] devices.

The first two depend on the interaction of the RF magnetic [11] or electric field [13] and a conducting vane. In the torque-vane power meter the torque on the vane is proportional to the RF power and is measured in terms of a known restraining torque. Since these forces are very small, their measurement requires delicate apparatus. As a result, such instruments tend to be vibration-sensitive. A torque-vane wattmeter is commercially available. It is used to measure RF power over a range of 1 to 200 watts, with an uncertainty ranging from 2 to 5 percent.

There are two types of vibration power meters: the first is based on the oscillation of a metallic rod suspended in a resonant cavity [5], the second on the oscillation of a quartz cantilever beam suspended in a waveguide [15]. The amplitude of the oscillation is proportional to the RF power. For

greater sensitivity the RF power must be chopped at the resonant frequency of the mechanical system in the first design and at the quartz-beam resonant frequency in the second. Experimental meters that measure power over a range of 10 μW to 1 W with an estimated uncertainty of 2 to 5 percent have been built.

b) Hall effect power meter [4], [5], [16]–[18]: The Hall effect in semiconductors constitutes the basis of an experimental RF power meter designed to measure the net power in a transmission line. While the Hall effect has been used in a commercial power meter for the power frequencies (50 to 60 Hz), it requires additional experimental work and evaluation at RF.

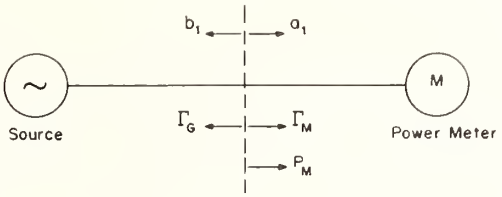
3. Applications

3.1 Introduction

Impedance considerations are a part of every power measurement. Reflections from the waveguide components can cause serious power-measurement errors, called mismatch errors. These errors can be greater than the error inherent in the power meter. Understanding of mismatch errors and of special methods for eliminating them is essential to the proper application of power meters.

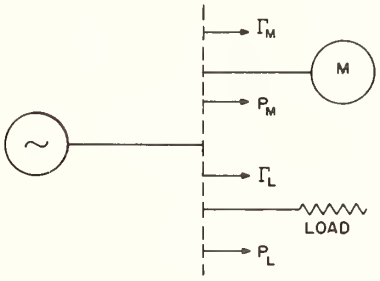
3.2 Direct Power Measurement

The simplest application of a terminating power meter is a direct measurement. The power output of a source is measured as shown in Fig. 9.



$$P_M = P_0 \frac{1-|\Gamma_M|^2}{|1-\Gamma_G \Gamma_M|^2}$$

Fig. 9. Direct power measurement.



$$\frac{P_L}{P_M} = \frac{1-|\Gamma_L|^2}{1-|\Gamma_M|^2} \cdot \frac{|1-\Gamma_G \Gamma_M|^2}{|1-\Gamma_G \Gamma_L|^2}$$

Fig. 10. Comparison power measurement.

The source can be any combination of waveguide components that includes a signal generator with some of its energy coupled to the designated waveguide port. If the signal generator is stable and isolated from load changes, the normalized amplitude a_1 of the traveling wave incident upon the load is given by [19]

$$a_1 = \Gamma_G b_1 + b_G, \tag{6}$$

where b_1 is the amplitude of the reflected wave, b_G is the amplitude of the wave the source would deliver to a nonreflecting load ($b_1 = 0$), and Γ_G is the generator reflection coefficient.

In the direct measurement, the net power P_M absorbed by the power meter, in terms of the wave amplitudes, is $P_M = (|a_1|^2 - |b_1|^2)/Z_{01}$ where Z_{01} is the characteristic impedance (real). This equation and (6) are used to obtain a more revealing equation for the ratio of P_M to the power P_0 that the source would deliver to a nonreflecting load. This ratio is

$$\frac{P_M}{P_0} = \frac{1-|\Gamma_M|^2}{|1-\Gamma_G \Gamma_M|^2}, \tag{7}$$

where Γ_M is the reflection coefficient of the power meter.

The output power P_M is sometimes referred to the available power P_A. The available power is the maximum power the generator will deliver when the load impedance is the complex conjugate of the generator impedance. The available power may be determined by using the relationship $P_0 = P_A(1-|\Gamma_G|^2)$. The primary purpose of this measurement is to determine a signal level that is a property of the source and not dependent upon the particular power meter used. For this reason, the uncertainty introduced by the term involving the reflection coefficients is called a mismatch error [20], [21]. The usefulness of this power measurement depends upon having some knowledge of the reflection coefficients Γ_G and Γ_M.

3.3 Comparison Power Measurement

A second and probably more common type of measurement is one made to determine the power absorbed by a given load. This may be done by the comparison measurement scheme shown in Fig. 10. If a load and a power meter are alternately connected to the same stable generator, the relationship between the power absorbed by the load P_L and that absorbed by the power meter P_M is given by [20]

$$\frac{P_L}{P_M} = \frac{1-|\Gamma_L|^2}{1-|\Gamma_M|^2} \cdot \frac{|1-\Gamma_G \Gamma_M|^2}{|1-\Gamma_G \Gamma_L|^2}, \tag{8}$$

where Γ_L is the load reflection coefficient. In effect, the power meter is used to measure the source power P_0, as in (7), and the load power is determined from this knowledge of the source. The load could be another power meter being calibrated. The right-hand side of (8) gives the mismatch error in the comparison measurement. It introduces a measurement uncertainty to the extent of the uncertainty in the values of the reflection coefficients. Much of the effort to improve power measurements has been directed toward eliminating the mismatch errors.

3.4 Mismatch Errors

The magnitudes of the reflection coefficients in (7) and (8) are frequently known or are readily measured. Their phase angles are usually unknown and are more difficult to measure. Because of the difficulty in accurately measuring these phase angles, they are seldom used in the evaluation of mismatch errors. Other approaches are used; the particular approach depends upon the allowable uncertainty in the application.

One method that can be simply applied is the calculation of the limit of the uncertainty, assuming that the reflection coefficient phase angles have those values that maximize the error. The maximum and minimum mismatch error can be calculated using the inequality

$$1 - |\Gamma_G\Gamma_L| \le |1 - \Gamma_G\Gamma_L| \le 1 + |\Gamma_G\Gamma_L|$$

and known values of the reflection coefficient magnitudes $|\Gamma_L|$ and $|\Gamma_G|$. For example, if $|\Gamma_G| = 0.13 (\sigma_G = 1.3)$, $|\Gamma_L| = 0.2 (\sigma_L = 1.5)$, and $|\Gamma_M| = 0.1 (\sigma_M = 1.21)$, the uncertainty in the comparison measurement is about ± 4 percent. The voltage standing-wave ratio (VSWR) is given by $\sigma = (1 + |\Gamma|)/(1 - |\Gamma|)$.

The uncertainty in evaluating the comparison measurement (8) due to the unknown arg $(\Gamma_G\Gamma_M)$ and arg $(\Gamma_G\Gamma_L)$, is eliminated, first, when $\Gamma_L = \Gamma_M$, including the case $\Gamma_L = \Gamma_M = 0$; or, second, when $\Gamma_G = 0$. The load and power-meter reflection magnitudes must be known in the second case but not in the first case. Similarly, if $\Gamma_L = 0$ or $\Gamma_G = 0$, the uncertainty in the direct measurement (7) is eliminated.

Although a tuner could be added to the load or the power meter to obtain the first case, $\Gamma_L = \Gamma_M$, the uncertainty introduced by the unknown power loss in the tuner generally prohibits its use. A power meter which has an adjustable impedance but is free of this uncertainty will be discussed later. The second case, $\Gamma_G = 0$, can also be obtained in practice.

3.4.1. The Generator Reflection Coefficient, Γ_G: Most RF oscillators require some isolation from the load to prevent changes in output frequency, signal amplitude [b_G of (6)], and source reflection coefficient Γ_G. This isolation is usually provided by an attenuator of 20 dB or more or by a ferrite isolator. The source reflection coefficient is essentially determined by the attenuator (or isolator) and other components connected between it and the output port. A simplified circuit with an attenuator is shown in Fig. 11. The generator reflection coefficient Γ_{G2} at the output of the attenuator is given in terms of the generator reflection coefficient Γ_{G1} at the input, and the scattering coefficients of the attenuator by

$$\Gamma_{G2} = S_{22} + \frac{S_{12}S_{21}\Gamma_{G1}}{1 - S_{11}\Gamma_{G1}}. \tag{9}$$

The product $|S_{12}S_{21}| = 10^{-(A_1 + A_2)/20}$ where A_1 and A_2 are the "forward" and "reverse" attenuations, in dB, of the isolating device. For a reciprocal 20-dB attenuator or a 40-dB isolator, $|S_{12}S_{21}| = 0.01$. In most cases S_{22} will be, by far, the largest term in (9). A tuner connected to the output of the attenuator can be adjusted to reduce $|\Gamma_{G2}|$ (by reducing $|S_{22}|$).

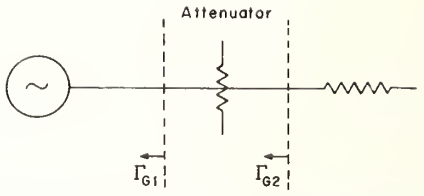


Fig. 11. Isolation of source from load using an attenuator.

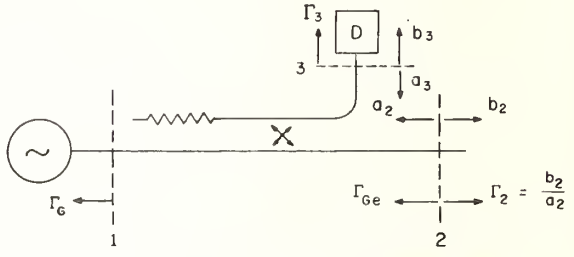


Fig. 12. Directional coupler used to simulate a low- (or zero-) reflecting source.

A directional coupler can be used to realize the equivalent of a low-reflection (or if tuned, essentially a zero-reflection) source [22]. With a coupler connected as shown in Fig. 12, the response of this system in terms of the normalized traveling-wave amplitudes, b_2, a_2, b_3, and of the scattering coefficients for the coupler is given by [22]

$$b_2 = \left(S_{22} - \frac{S_{21}S_{32}}{S_{31}}\right)a_2 + \frac{S_{21}}{S_{31}}\left[1 - \Gamma_3\left(S_{33} - \frac{S_{31}S_{23}}{S_{21}}\right)\right]b_3, \tag{10}$$

where $\Gamma_3 = a_3/b_3$ is the reflection coefficient of the detector at arm 3. If b_3 is maintained constant, then (10) has the same form as (6) and can be rewritten as

$$b_3 = \Gamma_{Ge}b_2 + b_{Ge}, \tag{11}$$

where b_{Ge} represents the term involving b_3 in (11). The equivalent generator reflection coefficient Γ_{Ge} is given by

$$\Gamma_{Ge} = S_{22} + \frac{S_{21}S_{32}}{S_{31}}. \tag{12}$$

The equations for direct and comparison power measurements made at arm 2, with b_3 constant, are the same as in (7) and (8) with Γ_{Ge} replacing Γ_G. The measurements are improved by use of the directional coupler for the following reasons: 1) b_3 is kept constant, providing a stable generator, 2) Γ_{Ge} is often smaller with this method than when an isolator or attenuator is used, and the large signal loss suffered when using attenuators is avoided, and 3) a tuning transformer added to arm 2 of the coupler can be tuned to obtain $\Gamma_{Ge} = 0$.

The maximum magnitude $|\Gamma_{Ge}|_{max}$ can be estimated in terms of the directional coupler's main-line VSWR, cou-

pling, and directivity. The maximum magnitude of (12) is

$$|\Gamma_{Ge}|_{\max} = |S_{22}| + \left|\frac{S_{21}S_{32}}{S_{31}}\right|.$$

The term $|S_{22}|$ is related to the main-line VSWR σ by the equation $S_{22}=(\sigma-1)/(\sigma+1)$. In the second term, $|S_{31}/S_{32}| = 10^{D/20}$, where D is the coupler directivity in dB. The remaining coefficient, S_{21}, which gives the coupling between arms 1 and 2, satisfies the inequality $|S_{21}|<1$. For example, broadwall multihole directional couplers with a main-line VSWR less than 1.05 ($|S_{22}|<0.025$) and a minimum directivity of 40 dB are commercially available. These result in a $|\Gamma_{Ge}|_{\max}$ of less than 0.035. A $|\Gamma_{Ge}|$ of 0.025 is typical for WR 90 (rectangular waveguide) directional couplers. However, even with $|\Gamma_{Ge}|=0.025$, the uncertainty in the direct measurement (Fig. 9) is 0.44 percent for a Γ_M equal to $0.2(\sigma_n=1.5)$.

A tuning transformer connected to arm 2 of the coupler can be tuned for $\Gamma_{Ge}=0$ as follows. With the junction excited as shown in Fig. 13, tuner T_A is adjusted to produce a null in the detector on arm 3. Under this condition, the reflection coefficient looking into arm 2 is $\Gamma_{2i}=\Gamma_{Ge}=S_{22}+(S_{21}S_{32})/S_{31}$. Tuner T_B is then adjusted for no reflection ($\Gamma_{2i}=0$), as indicated by a standing-wave machine or reflectometer connected to arm 2. With the source reconnected to arm 1, the equivalent generator ($b_3=$constant) as seen at arm 2 has the reflection coefficient $\Gamma_{Ge}=0$, as long as T_B remains undisturbed. Of course, the desired condition is only approached in practice. A method is available for measuring the Γ_{Ge} actually obtained [22].

3.5 *Directional-Coupler/Power-Meter Combinations*

The combination of a directional coupler with a power meter attached to its side arm is often used as a feed-through power meter or to extend the power range of an existing power meter. If, in Fig. 12, the detector on arm 3 is a power meter, the ratio of the power delivered to a load on arm 2 to the power reading at arm 3, P_3, is given by [20]

$$\frac{P_2}{P_3} = \left|\frac{S_{21}}{S_{31}}\right|^2 \cdot \frac{1-|\Gamma_2|^2}{1-|\Gamma_3|^2} \cdot \left|\frac{1-\left(S_{33}-\frac{S_{23}S_{31}}{S_{21}}\right)\Gamma_3}{1-\left(S_{22}-\frac{S_{21}S_{32}}{S_{31}}\right)\Gamma_2}\right|^2. \quad (13)$$

This equation may be written in the form

$$\frac{P_2}{P_3} = K\frac{1-|\Gamma_2|^2}{|1-\Gamma_{Ge}\Gamma_2|^2} \quad (14)$$

which is similar to that for the direct measurement (7). The constant factors in (13) are represented by K.

The value of K is determined primarily by the coupling ratio $|S_{21}/S_{31}|$ of the directional coupler. It is measured by a calibration performed with another power meter. If P_{2M} is the power measured at arm 2 by a power meter with a reflection coefficient Γ_M, and P_{3M} is the corresponding reading at arm 3, then K is given by the solution of (14):

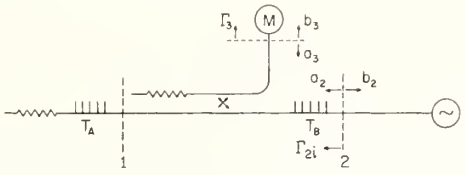


Fig. 13. Adjustments for $\Gamma_{2i}=0$.

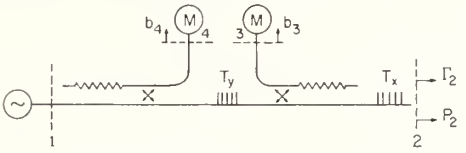


Fig. 14. Four-arm waveguide junction used for power measurements.

$$K = \frac{P_{2M}}{P_{3M}}\frac{|1-\Gamma_{Ge}\Gamma_M|^2}{1-|\Gamma_M|^2},$$

and the resultant expression for P_2/P_3 is

$$\frac{P_2}{P_3} = \frac{P_{2M}}{P_{3M}} \cdot \frac{1-|\Gamma_2|^2}{1-|\Gamma_M|^2} \cdot \left|\frac{1-\Gamma_{Ge}\Gamma_M}{1-\Gamma_{Ge}\Gamma_2}\right|^2.$$

The similarity of this expression to that for the comparison measurement (8) is apparent.

The advantages of the coupler/power-meter combination are: 1) as a feed-through power meter, it provides a continuous measurement of the power delivered from arm 2; 2) by choice of coupling ratio, the power range of a power meter can be shifted to a higher or lower range (a lower range is obtained when the roles of the coupler arms 2 and 3 are interchanged); 3) the calibration of the combination is independent of the characteristics of the source connected to arm 1; 4) a small equivalent generator reflection coefficient Γ_{Ge} is obtained; and 5) a tuner can be added to arm 2 and tuned for $\Gamma_{Ge}=0$, as described above. This tuner becomes part of the combination and is adjusted before calibration.

3.5.1 Combinations Employing a Pair of Couplers: Each of the above measurement techniques requires, at least, knowledge of the magnitudes of the load and power-meter reflection coefficients. This requirement can be eliminated by using a tuned system employing a pair of directional couplers. The net power out of arm 2, P_2, of the system shown in Fig. 14 is given by [23]

$$P_2 = \frac{|A|^2-|B|^2}{|\Delta|^2}|b_4|^2 - \frac{|D|^2-|C|^2}{|\Delta|^2}|b_3|^2 + 2\mathcal{R}e\left[\frac{1}{|\Delta|^2}(BD^*-AC^*)b_3^*b_4\right], \quad (15)$$

where A, B, C, D, and Δ contain the scattering coefficients of the four-arm junction and the power-meter reflection coefficients, Γ_3 and Γ_4; $\mathcal{R}e$ designates the real part of the quantity in brackets; the asterisk denotes the complex conjugate. The definitions of A, B, C, D, and Δ are not required to understand this discussion and are therefore not given here.

When the last term in (15) is zero, the power P_2 is obtained from the power-meter readings at arms 3 and 4 by

$$P_2 = k_1 P_4 - k_2 P_3. \tag{16}$$

The values of the proportionality constants, k_1 and k_2, are found when the system is calibrated. Although P_2, P_3, and P_4 are, in general, functions of the load reflection coefficient Γ_2, its value is not needed. This desired relation (16) results when in (15)

$$BD^* - AC^* = 0 \quad \text{or} \quad B = C = 0. \tag{17}$$

The condition $BD^* - AC^*$ is obtained [24] if tuner T_y is tuned for the result $|b_3/b_4|$ = constant, while adjusting the phase of a short circuit in arm 2 (tuner T_x is not needed here). That is, the ratio $|b_3/b_4|$ is independent of the phase ϕ_2 of $\Gamma_2 = e^{j\phi_2}$, when $BD^* - AC^* = 0$.

An added advantage accrues when the second method, $B = C = 0$, of obtaining the desired response (15) is realized. The system then becomes a tuned reflectometer [24] capable, when calibrated, of accurately measuring the reflection coefficient magnitude of the load on arm 2. For this application the second tuner T_x is tuned in conjunction with tuner T_y. The details of the tuning procedure have been reported [24]. The reflectometer is discussed in the section on microwave impedance measurements elsewhere in this issue.

3.5.2 Adjustable Impedance Power Meter [23]: An adjustable impedance power meter has many applications, especially in comparison measurements. The mismatch error in (8) is eliminated if the reflection coefficients of the power meter and the load are equal. The tuning of T_y or of T_x and T_y of Fig. 14 and the calibration to obtain values for k_1 and k_2 in (15) are independent of the characteristics of the termination at arm 1. Also, the validity of (15) is independent of the direction of the net energy flow at arm 2.

Two-coupler systems (e.g., Fig. 14) that satisfy either of the conditions of (17) can be used as adjustable impedance power meters by connecting a variable impedance (T_z and its load) to arm 1, as shown in Fig. 15. The tuner T_z is adjusted to obtain the desired value for the reflection coefficient Γ_M. If the system satisfies the second condition of (17), $B = C = 0$, it will still, with suitable calibration, provide an accurate measurement of $|\Gamma_M|$.

In the comparison measurement shown in Fig. 15, the ratio of P_L to $P_M (P_M = -P_2)$ is given by (8) and reduces to $P_M = P_L$ when $\Gamma_M = \Gamma_L$.

The adjustment of T_z for $\Gamma_M = \Gamma_L$ is done with the circuit shown in Fig. 16, where T_z and M represent the adjustable impedance power meter shown in Fig. 15. First, tuner T_A is adjusted for a null at the detector with the load Γ_L attached. The desired condition, $\Gamma_M = \Gamma_L$, is then obtained by adjusting T_z, with T_A undisturbed, to again produce a null with the adjustable impedance power meter attached.

The two-coupler systems can provide accurate power measurements. The available discussion of their applications includes: criteria for the choice of coupler ratios, alternate methods of performing the calibration, and equations for calculating the limits of the uncertainty due to imperfect tuning and other causes [23].

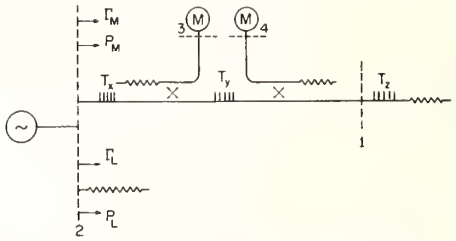


Fig. 15. Comparison measurement with adjustable impedance power meter.

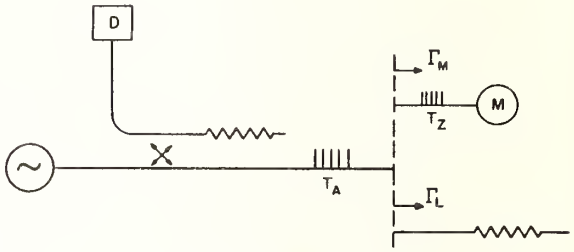


Fig. 16. Adjustment for $\Gamma_M = \Gamma_L$.

In systems where directional couplers can not be employed, (7) and (8) can be used to calculate the limits of measurement uncertainty in terms of the magnitudes of the power-meter, load, and generator reflection coefficients. The power-measurement uncertainty is reduced when directional couplers can be used to reduce the generator reflection coefficient or to eliminate the need to know the load reflection coefficient.

4. Pitfalls

4.1 Introduction

The total uncertainty in bolometric RF power measurements is the result of the uncertainties in associated instrumentation, mismatch, bolometer-mount efficiency, and substitution errors [4]. In coaxial mounts, one must also consider the dual-element bolometer error [25], [26].

The errors due to instrumentation are caused by inability to measure exactly the substituted dc or LF power. For the most widely used bolometric power meters, this instrumentation error varies from ± 0.5 to ± 3 percent, depending on whether dc or ac power is substituted. The range of error is reduced to ± 0.1 to ± 0.5 percent by using either a manual or self-balancing dc bolometer bridge with precision dc measuring instruments.

The mismatch error is caused by wave reflections from the load and generator. The problems associated with and techniques for minimizing this source of error have been discussed in Section 3.

Bolometer-mount efficiency of less than unity comes about because not all of the RF power is absorbed by the bolometer element, some being dissipated in the mount walls and supporting structures.

The substitution error arises because the dc and RF powers do not have identical distributions along or in the bolometer element. In dual-element bolometers, there is another source of substitution error, which is due to differences

in the resistances and the ohms-per-milliwatt coefficients of the two elements.

The sources of error are discussed in more detail in the following.

4.2 Efficiency of Bolometer Mounts

A bolometric power meter indicates the power absorbed by the detecting element. The efficiency figure of a bolometer mount accounts for that part of the net RF power dissipated in the unit in places other than the detecting element. The substitution error results from the difference in the effectiveness of the RF and dc in heating the element. It is difficult to separate these two sources of error in a measurement process, so they are both included in the concept of effective efficiency. The efficiency of a bolometer mount is defined as the ratio of the RF power absorbed by the bolometer element to the net power input to the bolometer mount [27]. The effective efficiency is defined as the ratio of the substituted bias power in the bolometer element to the net RF power input to the bolometer mount [28]. Note that both of these quantities are independent of the mount reflection coefficient.

The calibration factor, another useful figure of merit for a bolometer mount, includes the effective efficiency η_e and the magnitude of the mount reflection coefficients $|\Gamma_M|$. It is defined for bolometer mounts as the ratio of the substituted bias power to the RF power incident upon the bolometer mount [29]. For bolometer-mount/directional-coupler combinations, the calibration factor is defined as the ratio of the substituted bias power in the bolometer element on the side arm of the directional coupler to the RF power incident on a nonreflecting load connected to the main arm of the directional coupler [29].

It should be emphasized that any calibration of a bolometer mount or a bolometer-mount/directional-coupler combination is not necessarily invariant with time or environmental conditions. The user should recalibrate the instruments periodically, as may be indicated by the stability of prior calibration.

4.2.1 Measuring Effective Efficiency of Bolometer Mounts: The effective efficiency of a bolometer mount can be measured by several methods.

a) The simplest method is the direct comparison of the efficiency of the bolometer mount with that of a mount previously calibrated by a standards laboratory. By this method, the effective efficiency and the calibration factor may be obtained [29].

b) The reflectometer method [30] requires a calibrated standard power meter to calibrate the reflectometer. This technique yields the magnitude of the reflection coefficient and the effective efficiency of the bolometer mount.

c) The impedance technique [27], [31], [32] is used to determine the efficiency and the reflection coefficient of the bolometer mount. One drawback to the impedance method is that it may only be used to calibrate bolometer mounts containing barretter elements. Effective efficiency, generally a more useful quantity, may then be formed from knowledge of the substitution error.

d) Specially constructed calorimeters are used at the National Bureau of Standards to determine the effective efficiency of bolometer mounts [28], [33].

The effective efficiency is almost independent of power level between 10 μW and 10 mW. Although it includes the substitution error, it does not include the mismatch error, which must be accounted for separately.

4.2.2 Transfer of Calibration between Mounts with Different Input Waveguides: The transfer of an efficiency calibration from a standard bolometer mount to other bolometer mounts usually implies that the standard and unknown mounts have the same waveguide connectors, i.e., both have rectangular waveguide inputs, or both have the same type of coaxial connectors. A method for transferring the calibration from a rectangular waveguide bolometer mount to a coaxial bolometer mount has been developed [34]. The method can be extended to transfer calibrations in other waveguide types as well. It entails two measurements by standard calibration techniques [30]. The geometric mean of these two measurements plus the known efficiency of the standard waveguide mount give the efficiency of the coaxial bolometer mount, while the quotient of the two measurements yields an estimate of the adaptor losses or efficiency. The limits of uncertainty of a calibration transfer by this method are given as 1.0 to 1.4 percent.

4.2.3 Measurement of the Calibration Factor: The calibration factor of a bolometer mount may be measured by a comparison measurement with a calibrated bolometer mount, the reflectometer method, or the impedance technique.

The calibration factor includes the effects of the mismatch loss when a nonreflecting generator is used, as well as the efficiency and substitution errors, and it is used as a correction factor to obtain accurate power measurements.

4.2.4 Efficiency Errors in Calorimeters: Efficiency and substitution errors are also present in calorimetric power meters. Dissipation in the waveguide not measured by the calorimeter causes the efficiency error. The substitution error, as in the case of bolometers, is the result of the difference in the heating effects of the dc or LF and the RF power.

For specific calorimeters these errors have been evaluated analytically or experimentally, or both. These errors in commercial calorimeters are often grouped into one correction factor, and a formula for calculating the RF power in terms of a measured dc or LF voltage will include this factor. The user must ascertain that the calibration will not change with time, temperature, or other environmental factors.

4.2.5 Calibration Services: The Electronic Calibration Center of the National Bureau of Standards, Radio Standards Laboratory, Boulder, Colo., provides calibration services for bolometer mounts, bolometer-mount/directional-coupler combinations, RF calorimeters, and RF power meters. These services include measurement of efficiency, effective efficiency, or calibration factor for bolometer mounts or their combinations, and of power versus output voltage reading for calorimeters and other power meters. Table I lists NBS calibration services currently available in this area and includes the limits of un-

TABLE I

NBS Calibration Services Available in CW RF and CW Microwave Power

Method of Calibration	Item	Quantity Measured	Transmission Line	Frequency	Power Levels	Estimated Limits of Uncertainty
1) Direct comparison with transfer standard	Bolometer mounts	Effective efficiency or calibration factor	Coaxial	10, 30, 100, 200, 300, 400, 500, 700, 1000 MHz	1–10 mW	±1 percent
			Coaxial	1.3, 2, 2.2, 3, 3.5, 4 GHz		±1.5 percent
			WR 90	8.2–12.4 GHz		±1 percent
	Bolometer mount-directional coupler combination	Calibration factor	Coaxial	10, 30, 100, 200, 300, 400, 500, 700, 1000 MHz 1.3, 2, 2.2, 3, 4 GHz	1–10 mW in bolometer element	±2 percent
			WR 284 WR 187 WR 137	2.6–3.95 GHz 3.95–5.85 GHz 5.85–8.2 GHz		±1.5 percent
			WR 112 WR 90	7.05–10 GHz 8.2–12.4 GHz		±1 percent
			WR 62 WR 42	12.4–18 GHz 18–26.5 GHz		±1 to ±1.5 percent
2) Reflectometer method of calibration	Bolometer mount	Effective efficiency or calibration factor	WR 284 WR 187 WR 137	2.6–3.95 GHz 3.95–5.85 GHz 5.85–8.2 GHz	1–10 mW	±1.5 percent
			WR 112 WR 90 WR 62 WR 42	7.05–10 GHz 8.2–12.4 GHz 12.4–18 GHz 18–26.5 GHz		±1 percent
3) Impedance method of calibration	Bolometer mount (barretter elements only)	Efficiency or effective efficiency or calibration factor	WR 284 WR 187 WR 137	2.6–3.95 GHz 3.95–5.85 GHz 5.85–8.2 GHz	1–10 mW	±1 percent
4) Power versus output voltage	RF calorimeters and RF power meters	Power	Coaxial	10 MHz	0.001–200 W	±2 percent
				30 MHz	0.001–150 W	±2 percent
				100, 200, 300, 400, 500 MHz	0.001–100 W	±2 percent
				700, 1000 MHz	0.001–10 W	±2 percent
				1.3 GHz	0.001–50 W	±2 percent
				2, 2.2, 3, 4 GHz	0.001–1 W	±2 percent
5) Adaptor method of calibration	Thermistor mount, 200-ohm (Type *N* male connector)	Effective efficiency	Coaxial	4–7.05 GHz	10 mW	±2 percent
				7.05–10 GHz		±1.5 percent
	Thermistor mount, 200-ohm (14-mm precision connector)	Effective efficiency	Coaxial	4–7.05 GHz	10 mW	±1 percent
				7.05–10 GHz		1.5 percent

TABLE II

COMPARISON OF POWER METERS

Method	Typical Power Level	Frequency Range	Estimated Accuracy	Incident Power Absorbed	Time Constant	Remarks
Thermal Devices						
1) Static calorimeter						
a) Basic	100 μW–10 W	All	$\pm$0.5–$\pm$5 percent	100 percent	to 10^4 seconds	Fundamental type of measurement, power level must be maintained constant during a measurement, slow response time.
b) Substitution	100 μW–1 kW	All	$\pm$0.5–$\pm$3 percent	100 percent	to 10^2 seconds	
c) Twin	100 μW–10 mW	All	$\pm$0.2–$\pm$3 percent	100 percent	0.1–10^3 seconds	
2) Flow calorimeter						
a) Basic	1–10 W	All	$\pm$2–$\pm$5 percent	100 percent	1–10^2 seconds	Response time depends on the flow rate, complicated construction.
b) Substitution	10 mW–100 W	All	$\pm$1–$\pm$5 percent	100 percent	1–10^2 seconds	
3) Bolometers						
a) Barretter	10^{-6}–10^{-2} W	10 MHz–40 GHz	$\pm$1–$\pm$5 percent	100 percent	10^{-3} second	Easily burned-out, needs appropriate instrumentation.
b) Thermistor	10^{-6}–10^{-2} W	10 MHz–40 GHz	$\pm$1–$\pm$5 percent	100 percent	1 second	Rugged, difficult to burn-out, temperature-sensitive.
c) Film	10^{-3}–10^{-1} W	0.5 GHz–10 GHz	$\pm$2–$\pm$5 percent	100 percent	5–10 seconds	
d) Waveguide wall	10^{-3}–1 W	10 GHz–40 GHz	$\pm$2–$\pm$5 percent	3–4 percent	10 seconds	Can be used for a feed-through power meter.
4) Thermocouples						
a) Directly heated	10^{-3}–10^{-1} W	to 18 GHz	$\pm$2 percent	100 percent	0.1 second	Simple instrumentation.
b) Indirectly heated thermoelement	10^{-3}–10^{-1} W	to 40 GHz	$\pm$2 percent	100 percent	1–5 seconds	
c) Vacuum thermoelement	-5 mW	to 1 GHz	$\pm$1 percent	100 percent	0.5 second	
Mechanical Devices						
1) Radiation pressure	10–50 mW	3–40 GHz	$\pm$5 percent	Fraction	30 seconds	Absolute calibration possible, vibration-sensitive.
2) Torque vane	10–200 W	to 10 GHz	$\pm$2–$\pm$3 percent	2.5 percent	2 seconds	Absolute calibration possible, vibration-sensitive.
3) Vibration	10^{-5}–1 W	10–26.5 GHz	$\pm$3–$\pm$5 percent	100 percent	0.2–2 seconds	Linear detector, CW power must be chopped at the mechanical system resonant frequency.
Electronic Devices						
1) Crystal diode	10^{-6}–10^{-3} W	to 40 GHz	$\pm$5–$\pm$20 percent	100 percent	10^{-6} second	Useful only as level indicator.
2) Hall effect	10^{-3}–1 W	to 10 GHz	—	12–15 percent	—	Measures Poynting vector directly, very small effect at microwave frequency, still in experimental stage of development.

TABLE III

COMPARISON OF CALIBRATION METHODS

Method of Calibration	Frequency Range	Quantity Measured	Estimated Accuracy of Method	Remarks
Basic Methods				
Calorimeter	dc–40 GHz and higher	Effective efficiency	$\pm$0.2–$\pm$0.5 percent	Complicated to construct and evaluate for errors; long time constants.
Impedance technique	2.6–40 GHz	Efficiency	$\pm$0.5 percent	Can be used only with elements whose reflection coefficients can be calculated on the basis of dc measurements, such as barretters.
Transfer Methods				
Reflectometer methods	2.6–40 GHz	Effective efficiency and calibration factor	$\pm$1–$\pm$2 percent	Requires a calibrated bolometer mount for initial calibration of the reflectometer.
Direct comparison methods	dc–40 GHz	Effective efficiency, calibration factor, and power	$\pm$1–$\pm$3 percent	Requires a calibrated bolometer mount for comparison and evaluation of minimum error.

certainty of the different types of measurements [35]. Some commercial laboratories also provide some similar calibration services.

4.3 Substitution Errors

4.3.1 Dc-RF Substitution Error [3], [36]: In bolometers the same dc and RF power is initially assumed to produce the same resistance change. However, there is a nonuniform RF current distribution in the bolometer that causes a nonuniform distribution of the RF heating. The substitution error will vanish if the resistance is independent of the power distribution. This criterion is not completely satisfied in practice, but it is closely approached by making the bolometer elements very small.

The dc-RF substitution error has been analyzed for barretter elements [36]. This analysis shows that for typical conditions the substitution error will be 1 percent or less, while experimental evidence at 10 GHz [32] indicates that it may be about 0.1 percent. For thermistor elements, the substitution error is presumed to be quite small because of the small size of the thermistor bead [3], [4]. For both thermistors and barretters, it is taken into account in a bolometer unit effective efficiency calibration.

For bolometer bridges that supply simultaneous dc and AF bias power, there is an additional small substitution error [37]. The AF resistance of the bolometer is different from the dc, and it varies over the audio cycle. Power-meter manufacturers reduce or eliminate this error by choice of audio frequency, and by meter calibration.

4.3.2 Substitution Error in Dual-Element Bolometers [25], [26]: A dual-element error occurs in coaxial bolometer power meters that use a pair of bolometer elements. These elements are connected in series for the dc bias current and in parallel for RF voltage.

The series connection for dc bias assures equal current in the two elements. If their resistances are unequal, the one with the larger resistance dissipates more dc power. The parallel connection for RF causes the element with smaller resistance to dissipate more RF power. If the ohms-per-milliwatt coefficients of the two elements are not identical, an error results in the indicated power. This error is given by [26]

$$e = \left[\frac{1}{\gamma_b} - \frac{1}{\gamma_a}\right](r_{a1} - r_{b1}),$$

where γ_a and γ_b are the ohms-per-milliwatt coefficients of the two bolometer elements, and r_{a1} and r_{b1} are the respective values of the two elements when the bridge is balanced by dc or LF only. For thermistor mounts, this error increases with increasing RF power, and above 10 mW increases rapidly. For barretter mounts, the error decreases with increasing RF power. It is therefore advisable to limit the RF dissipation in the dual thermistor mount to about 25 percent of the total power (RF + dc bias power). At 10-mW RF the error ranges between 0.1 and 1.0 percent [26].

5. Summary

A survey has been made of some of the types of RF power meters. These are summarized in Table II. The various methods of measuring power and of calibrating power meters have been discussed. The latter are summarized in Table III. Finally, several important sources of error in bolometric power measurements have been mentioned briefly. These were the mismatch errors, efficiency errors, and substitution errors.

References

[1] R. A. Schrack, "Radio frequency power measurements," NBS Circular 536, March 16, 1953.

[2] *The Technique of Microwave Measurements*, C. G. Montgomery, Ed., M.I.T. Radiation Lab. Ser., vol. 11. New York: McGraw-Hill, 1947.

[3] E. L. Ginzton, *Microwave Measurements*. New York: McGraw-Hill, 1957.

[4] *Handbook of Microwave Measurements*, M. Sucher and J. Fox, Eds. New York: Wiley, 1963, ch. 3.

[5] A. F. Harvey, *Microwave Engineering*. New York: Academic, 1963.

[6] J. Collard, "The enthrakometer, an instrument for the measurement of power in rectangular wave guides," *J. IEE (London)*, vol. 93, pt. IIIA, no. 9, pp. 1399–1402, 1946.

[7] L. E. Norton, "Broad-band power-measuring methods at microwave frequencies," *Proc. IRE*, vol. 37, pp. 759–766, July 1949.

[8] G. F. Engen, "A self-balancing direct-current bridge for accurate bolometric power measurements," *J. Res. NBS*, vol. 59, pp. 101–105, August 1957.

[9] I. Lemco and B. Rogal, "Resistive film milliwattmeters for the frequency bands 8.2–12.4 Gc/s, 12.4–18 Gc/s, and 26.5–40 Gc/s," *Proc. IEE (London)*, vol. 107, pt. B, pp. 427–430, September 1960.

[10] P. A. Hudson, "A precision RF power transfer standard," *IRE Trans. on Instrumentation*, vol. I-9, pp. 280–283, September 1960.

[11] A. L. Cullen, "Absolute power measurement at microwave frequencies," *Proc. IEE (London)*, vol. 99, pt. IV, pp. 100–110, April 1952.

[12] ——, "A general method for the absolute measurement of microwave power," *Proc. IEE (London)*, vol. 99, pt. IV, pp. 112–120, April 1952.

[13] A. L. Cullen and L. M. Stephenson, "A torque-operated wattmeter for 3 cm microwaves," *Proc. IEE (London)*, vol. 99, pt. IV, pp. 294–301, July 1952.

[14] A. L. Cullen, B. Rogal, and S. Okamura, "A wide-band double-vane torque-operated wattmeter for 3-cm microwaves," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-6, pp. 133–136, April 1958.

[15] J. F. Rando, "Power measurements by means of thermally excited mechanical vibrations," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-15, pp. 59–61, March–June 1966.

[16] H. E. M. Barlow and L. M. Stephenson, "The Hall effect and its application to power measurement at microwave frequencies," *Proc. IEE (London)*, vol. 103, pt. IIIB, pp. 110–112, January 1956.

[17] H. E. M. Barlow and S. Katoaka, "The Hall effect and its application to power measurement at 10 Gc/s," *Proc. IEE (London)*, vol. 105, pt. B, pp. 53–60, January 1958.

[18] R. A. Leavenworth, J. Maczuk, and R. F. Schwartz, "The application of Hall-effect devices to the measurement of microwave power density," *IRE Trans. on Instrumentation*, vol. I-11, pp. 264–269, December 1962.

[19] D. M. Kerns and R. W. Beatty, *Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis*. New York: Pergamon, 1967.

[20] R. W. Beatty and A. C. Macpherson, "Mismatch errors in microwave power measurements," *Proc. IRE*, vol. 41, pp. 1112–1119, September 1953.

[21] "Microwave power measurement," Microwave Lab., Hewlett-Packard Co., Palo Alto, Calif., Appl. Note 64, October 1965.

[22] G. F. Engen, "Amplitude stabilization of a microwave signal source," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-6, pp. 202–206, April 1958.

[23] ——, "A variable impedance power meter and adjustable reflection coefficient standard," *J. Res. NBS*, vol. 68C, pp. 7–24, January–March 1964.
[24] G. F. Engen and R. W. Beatty, "Microwave reflectometer techniques," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-7, pp. 351–355, July 1959.
[25] I. A. Harris, "A coaxial film bolometer for the measurement of power in the UHF band," *Proc. IEE (London)*, vol. 107, pt. B, pp. 67–72, January 1960.
[26] G. F. Engen, "A dc-RF substitution error in dual-element bolometer mounts," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-13, pp. 58–64, June–September 1964.
[27] D. M. Kerns, "Determination of efficiency of microwave bolometer mounts from impedance data," *J. Res. NBS*, vol. 42, pp. 579–585, June 1949.
[28] A. C. Macpherson and D. M. Kerns, "A microwave microcalorimeter," *Rev. Sci. Instr.*, vol. 26, pp. 27–33, January 1955.
[29] R. F. Desch and R. E. Larson, "Bolometric microwave power calibration techniques at the National Bureau of Standards," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-12, pp. 29–33, June 1963.
[30] G. F. Engen, "A transfer instrument for the intercomparison of microwave power meters," *IRE Trans. on Instrumentation*, vol. I-9, pp. 202–208, September 1960.
[31] R. W. Beatty and F. Reggia, "An improved method of measuring efficiencies of ultra-high-frequency and microwave bolometer mounts," *J. Res. NBS*, vol. 54, pp. 321–327, June 1955.
[32] G. F. Engen, "A bolometer mount efficiency measurement technique," *J. Res. NBS*, vol. 65C, pp. 113–124, April–June 1960.
[33] ——, "A refined *X*-band microwave microcalorimeter," *J. Res. NBS*,vol. 63C, pp. 77–82, July–September 1959.
[34] ——, "Coaxial power meter calibration using a waveguide standard," *J. Res. NBS*, vol. 70C, pp. 127–138, April–June 1966.
[35] "Calibration and test services of the National Bureau of Standards," NBS Misc. Publ. 250, pp. 35–37 and 44–45, 1965.
[36] H. J. Carlin and M. Sucher, "Accuracy of bolometric power measurements," *Proc. IRE*, vol. 40, pp. 1042–1048, September 1952.
[37] S. J. Raff and G. U. Sorger, "A subtle error in RF power measurements," *IRE Trans. on Instrumentation*, vol. I-9, pp. 284–291, September 1960.

Bibliography

Books and Handbooks

[38] M. Wind and H. Rapaport, *Handbook of Microwave Measurements*, 2nd ed. Brooklyn, N. Y.: Polytechnic Institute of Brooklyn, 1955, ch. 4.
[39] Staff, Hewlett-Packard Co., *Microwave Theory and Measurements*. Englewood Cliffs, N. J.: Prentice-Hall, 1962.
[40] A. L. Lance, *Introduction to Microwave Theory and Measurements*. New York: McGraw-Hill, 1964.
[41] *PRD Reports, Vols. 1–7*, W. A. Weissman, Ed., 1st ed. Brooklyn, N. Y.: PRD Electronics, January 1964.

Survey Papers

[42] G. F. Engen, "Recent developments in the field of microwave power measurements at the National Bureau of Standards," *IRE Trans. on Instrumentation*, vol. I-7, pp. 304–306, December 1958.
[43] K. Morita, "Report of advances in microwave theory and techniques in Japan—1959," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-8, pp. 395–397, July 1960.
[44] G. Goudet, "Report of advances in microwave theory and techniques in Western Europe—1959." *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-8, pp. 387–394, July 1960.
[45] J. Brown, "Report of advances in microwave theory and techniques in Great Britain—1959," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-8, pp. 382–386, July 1960.
[46] G. F. Engen, "Radiofrequency and microwave power measurements," *J. Res. NBS*, vol. 64D, pp. 596–597, November–December 1960.
[47] A. E. Karbowiak, "Microwaves—A review of progress in Great Britain during 1960," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-9, pp. 374–384, September 1961.
[48] A. L. Cullen, "Microwave power measurement," *Proc. IEE (London)*, vol. 109, pt. B, suppl. 23, pp. 724–733, May 1962.
[49] R. E. Larson, "Microwave measurements in the NBS Electronic Calibration Center," *Proc. IEE (London)*, vol. 109, pt. B, suppl. 23, pp. 644–650, May 1962.
[50] G. F. Engen, "A survey of microwave power measurement techniques employed at the National Bureau of Standards," *Proc. IEE (London)*, vol. 109, pt. B, suppl. 23, pp. 734–739, May 1962.
[51] A. Serchuk, "Microwave calorimeters, a product survey," *Microwaves*, vol. 2, pp. 48–51, September 1963.
[52] R. L. Conhaim, "R.F. power output measurements," *Electronics World*, vol. 73, pp. 53–56, 100, October 1963.
[53] A. Serchuk, "Bolometers, a product survey," *Microwaves*, vol. 2, pp. 52–55, November 1963.
[54] R. W. Beatty, "Microwave standards and measurements, a progress review 1960 to 1963," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-12, pp. 134–138, December 1963.
[55] G. F. Engen and N. T. Larsen, "RF and microwave power measurements," *J. Res. NBS (Radio Science)*, vol. 68D, pp. 527–528, May 1964.
[56] Staff Report, "The NBS and measurement standards," *Electro-Tech.*, vol. 74, pp. 91–99, October 1964.
[57] M. Magid, "UHF measurements," *Electro-Tech.*, vol. 74, pp. 67–88, October 1964.
[58] J. Macrie, "Microwave power meters: a user's guide," *Microwaves*, vol. 4, pp. 24–35, January 1965.
[59] M. Magid, "Accurate microwave component testing," *Electro-Tech.*, vol. 76, pp. 42–49, October 1965.
[60] R. W. Beatty, "The system of electromagnetic quantities above 1 GHz," *Metrologia*, vol. 2, pp. 46-54, January 1966.
[61] M. C. Selby, "The system of electromagnetic quantities at 30 kHz to 1 GHz," *Metrologia*, vol. 2, pp. 35–45, January 1966.
[62] R. F. Clark, A. Jurkus, and C. F. Patterson, "Calibration facility for RF quantities offers central tie-in for testing labs," *Canad. Electronics Engrg.*, pp. 40–45, February 1966.
[63] R. F. Schwartz, "Precision microwave power measurements, a survey," *Electronic Ind.*, vol. 25, pp. 88–96, June 1966.
[64] R. W. Beatty, "Microwave standards and measurements in the U. S. A., 1963–1966," *Prog. Radio Sci. (URSI)*, February 1967.
[65] I. A. Harris, "Radio measurements and standards at frequencies up to 1 GHz, a progress review 1963–1966," *Prog. Radio Sci. (URSI)*, February 1967.

Power Meters

[66] B. Bleaney, "Radio frequency measurement by bolometer lamps of centimetre wave lengths," *J. IEE (London)*, vol. 93, pt. IIIA, no. 9, pp. 1377–1382, 1946.
[67] D. M. Kerns, "Analysis of direct-current bolometer bridge," *J. Res. NBS*, vol. 43, pp. 581–589, December 1949.
[68] H. A. Thomas, "Microwave power measurements employing electron beam techniques," *Proc. IRE*, vol. 45, pp. 205–211, February 1957.
[69] M. Sucher and H. J. Carlin, "Broad-band calorimeters for the measurement of low and medium level microwave power. I. Analysis and design," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-6, pp. 188–194, April 1958.
[70] A. V. James and L. O. Sweet, "Broad-band calorimeters for the measurement of low and medium level microwave power. II. Construction and performance," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-6, pp. 195–202, April 1958.
[71] P. A. Hudson and C. M. Allred, "A dry, static calorimeter for RF power measurement," *IRE Trans. on Instrumentation*, vol. I-7, pp. 292–296, December 1958.
[72] D. L. Birx and N. Fuschillo, "The theory of low-temperature bolometer detectors applied to the measurement of low-level RF power," *IRE Trans. on Instrumentation*, vol. I-7, pp. 310–315, December 1958.
[73] E. E. Aslan, "Temperature-compensated microwatt power meter," *IRE Trans. on Instrumentation*, vol. I-9, pp. 291–297, September 1960.
[74] J. A. Lane, "The design and performance of transverse film bolometers in rectangular waveguides," *Proc. IEE (London)*, vol. 108, pt. B, pp. 133–135, January 1961.
[75] J. P. Vinding, "An accurate calorimeter for high microwave power,"

Microwave J., vol. 4, pp. 41–46, January 1961.
[76] D. Woods, "A coaxial millivoltmeter/milliwattmeter for frequencies up to 1 Gc/s," *Proc. IEE* (*London*), vol. 109, pt. B, suppl. 23, pp. 750–756, May 1962.
[77] H. E. M. Barlow, "A new approach to the measurement of H.F. power by Hall effect in a semiconductor," *Proc. IEE* (*London*), vol. 109, pt. B, suppl. 23, pp. 746–749, May 1962.
[78] E. Acs, "A proposed method of microwave power measurement," *Proc. IEE* (*London*), vol. 109, pt. B, suppl. 23, pp. 744–745, May 1962.
[79] M. M. Brady, "In-line waveguide calorimeter for high-power measurement," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-10, pp. 359–366, September 1962.
[80] H. M. Barlow, "Microwave power measurements," *IRE Trans. on Instrumentation*, vol. I-11, pp. 257–263, December 1962.
[81] M. M. Brady, "Correction to 'In-line waveguide calorimeter for high-power measurements'—Accounting for transverse waveguide wall currents," *IEEE Trans. on Microwave Theory and Techniques* (*Correspondence*), vol. MTT-11, p. 152, March 1963.
[82] S. Casper, "Bolometer characteristics," *Microwaves*, vol. 1, pp. 56–63, November 1963.
[83] K. A. Yakovlev, "A method for measuring power by means of multigrid tubes," *Izmeritel'naya Tekhnika*, pp. 34–37, May 1965.
[84] R. F. Clark, "A coaxial calorimeter for use as a microwave power standard," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-14, pp. 59–63, March/June 1965.
[85] E. W. Helmholz and G. Bittner, "Hochfrequenz-Leistungsmessung mit Soffittenlampen-Eigenschaften und Auswahl der Lampen," *Phys.-Tech. Bundesanst. Mitt.*, vol. 75, pp. 343–347, August 1965.
[86] E. E. Aslan, "A single bead broadband coaxial thermistor mount," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-14, pp. 102–106, September 1965.
[87] V. I. Pronenko, "Wide-band calorimetric U.H.F. power meters," *Izmeritel'naya Tekhnika*, pp. 44–47, September 1965.
[88] L. Leysk, "Calorimetry the simple way," Bird Electronic Co., Cleveland, Ohio, Engrg. Essay 1, 1966.
[89] H. E. M. Barlow and E. S. Tay, "Power measurement at 10 Mc/s by absorption of the angular momentum of a rotating field," *Proc. IEE* (*London*), vol. 113, pp. 1253–1255, July 1966.
[90] H. E. M. Barlow and A. E. Williams, "Power measurements at 10 Gc/s by reversal of the angular momentum of a circularly polarised field," *Proc. IEE* (*London*), vol. 113, pp. 1248–1252, July 1966.

Measurement Techniques

[91] W. C. Reisener and D. L. Birx, "A new precision low-level bolometer bridge," *Proc. IRE*, vol. 50, pp. 39–42, January 1962.
[92] J. F. Warren and M. H. Wiese, "The measurement of radio frequency and microwave power in an industrial primary standards metrology laboratory," *Proc. ISA Conf.*, paper 47.5.63, 1963.
[93] J. J. Rooney, "How to measure high frequency power," *EEE—Circuit Design Engrg.*, vol. 12, pp. 82–84, May 1964.
[94] V. G. Orlov, "Absolute calibration of pondermotive force wattmeters with losses," *Izmeritel'naya Tekhnika*, pp. 46–48, May 1964.
[95] V. I. Pronenko and V. D. Frumkin, "Testing of power meters," *Izmeritel'naya Tekhnika*, pp. 43–45, September 1964.
[96] M. C. Selby, "Single-bolometer series-parallel circuit for accurate voltage and power measurement," *Proc. ISA Conf.*, paper 21.6-1-64, October 1964.
[97] B. P. Hand, "The rapid measurement of bolometer mount efficiency," *Proc. ISA Conf.*, paper 21.6-2-64, October 1964.
[98] A. S. Elizarev, "Analysis of the operation of automatic thermistor bridges," *Izmeritel'naya Tekhnika*, pp. 49–52, November 1964.
[99] L. M. Zaks, V. M. Petrov, and E. N. Belikov, "Self-balancing dc thermistor bridge type M4-3 for measuring UHF power," *Izmeritel'naya Tekhnika*, pp. 41–43, June 1965.
[100] J. J. Macrie, "Efficiency: The missing link in bolometer power measurements," *Microwave J.*, vol. 8, pp. 26–32, June 1965.
[101] L. M. Zaks, E. N. Belikov, S. V. Rypalev, and V. M. Petrov, "Thermistor bridge with an automatic digital display and zero correaction," *Izmeritel'naya Tekhnika*, pp. 41–43, September 1965.
[102] T. Tamaru, "A note on bolometer mount efficiency measurement technique by impedance method in Japan," *IEEE Trans. on Microwave Theory and Techniques* (*Correspondence*), vol. MTT-14, p. 437, September 1966.
[103] G. U. Sorger and B. O. Weinschel, "Coaxial transfer standards for microwave power measurements," *Electronic Instr. Digest*, vol. 2, pp. 6–12, November 1966.

Error Analyses

[104] S. Omori and K. Sakurai, "A new estimating method of equivalence error in microwave microcalorimeter," *IRE Trans. on Instrumentation*, vol. I-7, pp. 307–309, December 1958.
[105] B. P. Hand, "Microwave mismatch error analysis," Hewlett-Packard Co., Palo Alto, Calif., Appl. Note 56, October 1965.

Miscellaneous

[106] I. K. Munson, "Microwave power stabilizer," *Rev. Sci. Instr.*, vol. 21, pp. 622–624, July 1950.
[107] G. D. Monteath, "Coupled transmission lines as symmetrical directional couplers," *Proc. IEE* (*London*), vol. 102, pt. B, pp. 383-392, May 1955.
[108] G. F. Engen and R. W. Beatty, "Microwave attenuation measurements with accuracies from 0.0001 to 0.06 decibel over a range of 0.01 to 50 decibels," *J. Res. NBS*, vol. 64C, pp. 139–145, April–June 1960.
[109] G. F. Engen, "A method of improving isolation in multichannel waveguide systems," *IRE Trans. on Microwave Theory and Techniques* (*Correspondence*), vol. MTT-8, pp. 460–461, July 1960.
[110] W. J. Anson, "A guide to the use of the modified reflectometer technique of VSWR measurement," *J. Res. NBS*, vol. 65C, pp. 217–223, October–December 1961.
[111] D. Woods, "Improvements in precision coaxial resistor design," *IRE Trans. on Instrumentation*, vol. I-11, pp. 305–309, December 1962.
[112] V. I. Pronenko and V. D. Frumkin, "Uniformity and accuracy in measuring power at high and ultra-high frequencies," *Izmeritel'naya Tekhnika*, pp. 51–52, April 1964.
[113] P. A. Hudson, "A high directivity, broadband coaxial coupler," *IEEE Trans. on Microwave Theory and Techniques* (*Correspondence*), vol. MTT-14, pp. 293–294, June 1966.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967
pp. 837-850
THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

ERRATA

in

RADIO FREQUENCY POWER MEASUREMENTS
by Anne Y. Rumfelt and Lyman B. Elwell

Proc. IEEE, Vol. 55, No. 6, pp. 837-850, June 1967.

page 842, column 1, paragraph 2, lines 9-10 should have read: "the uncertainty in the comparison measurement is about ± 8 percent."

page 842, column 2, Figure 12. Γ_2 should have been shown as :

$$\Gamma_2 = a_2/b_2$$

page 842, column 2, equation (11) should have read:

$$b_2 = \Gamma_{Ge}\, a_2 + b_{Ge} .$$

page 842, column 2, equation (12) should have read:

$$\Gamma_{Ge} = S_{22} - (S_{21}\, S_{32})/S_{31} .$$

page 843, column 1, paragraph 1, lines 13-15, should have read: "the uncertainty in the direct measurement (Figure 9) is 1 percent for a Γ_M equal to $0.2 (\sigma_M = 1.5)$."

page 843, column 1, paragraph 2, lines 5-6, should have read:

$$\Gamma_{2i} = \Gamma_{Ge} = S_{22} - (S_{21}\, S_{32})/S_{31} .$$

page 845, column 1, Section 4.2, paragraph 2, line 3 should have read: "magnitude of the mount reflection coefficient $|\Gamma_M|$.

page 850, add to the Bibliography Section, under Miscellaneous

[114] I. A. Harris, "The Theory and Design of Coaxial Resistor Mounts for the Frequency Band 0-4000 Mc/s," Proc. I. E. E. (London), Vol. 103, Part C., No. 3, pp. 1-10, March 1956.

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 65C, No. 2, April–June 1961

A Bolometer Mount Efficiency Measurement Technique*

G. F. Engen

(December 14, 1960)

In the measurement of microwave power by means of the bolometric technique, the efficiency of the bolometer mount must be measured and applied as a correction in order to meet many of the accuracy requirements of today's technology.

An impedance technique of determining this efficiency was proposed over ten years ago, but has found but little use to date because of the rather severe performance requirements imposed on the attendant instrumentation. This paper describes an improved method of implementing the technique which is based on the reflectometer concept.

A particularly attractive feature of the new method is its substantial independence of connector discontinuity which has been a particularly troublesome source of error in coaxial systems. In further contrast with some of the earlier proposals for implementing this impedance technique, the new method is readily applicable to both matched and unmatched mounts and does not require mathematical approximations (in the first order theory).

A comprehensive error analysis indicates that an accuracy of 0.5 percent is possible in the existing state of the art.

1. Background

The "impedance" method of measuring bolometer mount efficiency devised by Kerns [1] [1] is one of the few basic techniques developed thus far for determining this parameter. As originally outlined, however, the accuracy which could be achieved was rather severely limited by the state of the impedance measuring art, and this led in turn to the development of a number of modifications of the technique with the objective of reducing the overall error. Beatty [2], for example, proposed a modification based on certain mathematical approximations and restrictions in generality which provided improved accuracy, but the associated operating procedures proved to be nonetheless time consuming and exacting, and the accessory instrumentation was never developed or refined to the point where one was, on a routine basis, able to place a great deal of confidence in the results. Additional refinements or modifications of the technique have also been suggested by Weinschel [3], Ginzton [4], Lane [5], and perhaps others, but it is probably safe to say that none of these proposals has, as yet, come into widespread use.

At the Boulder Laboratories of the National Bureau of Standards another variation of the impedance method has been developed which provides improved accuracy and simplified operational procedures but, unlike the earlier modifications, requires neither mathematical approximations (in the first order theory), nor restrictions in the generality of the method. A particularly attractive feature of this new version is its substantial independence of connector discontinuity, which has been an especially troublesome source of error in coaxial systems. This technique constitutes an improved method of making the measurements implicit in the "impedance" method formulated by Kerns, and the subsequent discussion and error analysis will be limited to the procedures to be described. For a discussion of the more basic postulates upon which the procedure is based the reader is referred to the original paper [1]. In particular it should be noted that the technique is, in its present form, applicable to barretter but not to thermistor type bolometers.[2]

2. Introduction

It was shown by Kerns [1] that the efficiency of a bolometer mount may, under suitable conditions, be determined from three impedance measurements at the bolometer mount input terminals corresponding to three different values of bolometer resistance. If one of these resistance values is chosen to coincide with the value of resistance for which the efficiency is desired, this result may be expressed in terms of the input reflection coefficients thus [2]:

$$\eta = K \frac{|(\Gamma_3 - \Gamma_2)(\Gamma_1 - \Gamma_2)|}{|\Gamma_3 - \Gamma_1|(1 - |\Gamma_2|^2)} \tag{1}$$

where the Γ_1, Γ_2, Γ_3 are the reflection coefficients corresponding to bolometer resistances R_1, R_2, R_3, respectively, η is the efficiency when the bolometer

*A preliminary report of this technique was given at a joint meeting of the International Scientific Radio Union and the Institute of Radio Engineers, April 26, 1958, Washington, D.C.

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

[2] In order to apply the technique to a given bolometer mount, it is a *sufficient* (but not necessary) condition that it be possible to determine the microwave impedance of the bolometer element from measurements of its d-c resistance. A somewhat different sufficient condition is that the resistive component of the microwave impedance be related to the d-c value by means of a real proportionality factor (whose magnitude need not be known), and that the reactive component be constant or independent of the d-c resistance. Necessary conditions have not been determined, but it has not been possible, thus far, to relate the microwave impedance to the d-c resistance in an adequate fashion in order to utilize this technique with thermistors.

resistance has the value R_2 (typically 200 ohms), and $K=\left|\frac{2R_2(R_3-R_1)}{(R_2-R_1)(R_3-R_2)}\right|$. For the special case of a "matched" mount ($\Gamma_2=0$) the expression becomes

$$\eta=K\left|\frac{\Gamma_3\Gamma_1}{\Gamma_3-\Gamma_1}\right|. \tag{2}$$

The determination of the factor K, while presenting a number of practical problems, does not pose the difficulties encountered in measuring the factor containing the Γ's. It will be noted that this latter factor is characterized by differences or changes in the reflection coefficient values as the bolometer resistance is varied. Because the resistance variation which can be achieved with the available barretter elements is at best rather limited, the corresponding changes in the reflection coefficient are not large, and if the individual Γ's are measured (by means of a slotted line for example) a rather severe requirement on the accessory instrumentation results. Beatty [2] has shown that an error of ± 1 percent in the VSWR measurements may lead to an error as large as ± 6 percent in the determined efficiency.

More recently, the reflectometer technique has found increased usage in the measurement of reflection coefficient magnitudes. This device, shown in figure 1, ideally yields a response of the form:

$$\left|\frac{b_3}{b_4}\right|=k|\Gamma| \tag{3}$$

where b_3 and b_4 are the wave amplitudes of the signals at the respective detectors, Γ is the reflection coefficient of the bolometer mount or other load terminating arm 2, and k is a (real) constant whose value may be determined by observing the detector outputs with arm 2 terminated by a load of known reflection—a fixed short for example.

While the usual reflectometer fails to produce this type of response because of imperfect directivity and other deviations from ideal behavior of the directional couplers employed, it is possible to compensate for these imperfections at a single frequency by means of auxiliary tuners [6], as shown in figure 2. Procedures for the adjustment of these tuners, such that the response of eq (3) may be realized, were given in the cited reference [6]. In essence the procedure is to adjust the tuning transformer T_x such that the directivity of the associated directional coupler becomes infinite, while T_y is adjusted

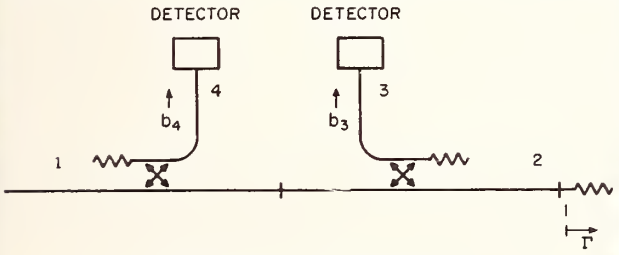


FIGURE 1. *Ideal reflectometer.*

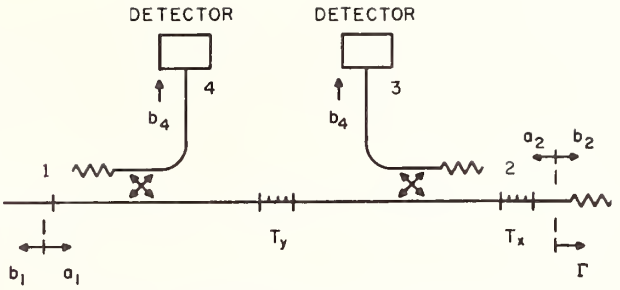


FIGURE 2. *Reflectometer with auxiliary tuners.*

such that the reflection coefficient of the equivalent generator, at the terminals where the bolometer mount is connected, vanishes.

The reflectometer does not, in this form, provide a determination of the argument or phase angle of the reflection coefficient, however Beatty [2] has noted that for matched mounts of high efficiency the reflection coefficient vectors Γ_1 and Γ_3 are nearly colinear so that to a good approximation $|\Gamma_3-\Gamma_1|=|\Gamma_3|\pm|\Gamma_1|$. (The proof of a more general counterpart of this statement will be given later in this paper.) To the extent of the validity of this approximation, the cited reflectometer technique may thus be employed to determine the efficiency of matched bolometer mounts by means of eq (2). The application of the reflectometer method to the more general problem of unmatched mounts is, however, by no means obvious.

Within the limitations discussed, the described method is practical in rectangular waveguide systems. In coaxial systems, however, the procedure leaves a great deal to be desired, and even in rectangular waveguide, as will be subsequently demonstrated, the procedure, as given, is unnecessarily complex for the application.

The problem in coaxial measurements centers around the impedance discontinuity which is part of the coaxial connector. In order to provide a useful result, an overall accuracy in the efficiency determination of 1 percent would appear to be a reasonable goal, and since this figure will include contributions from a number of different sources it is desirable to reduce the individual contributions to the order of 0.1 percent. Applying this criterion to the amount of mismatch permissible in the equivalent generator impedance implies a generator match with a VSWR of 1.002 or less. While this value may be realized in rectangular waveguide systems by means of sufficiently refined techniques, the prospect of achieving such a result in a coaxial system, where connector discontinuities with VSWR's of the order of 1.1 are typical, is certainly far from encouraging.

It may be noted, in passing, that this problem has appeared so regularly in attempts to refine this technique for use with coaxial type bolometer mounts that one is led to wonder if any significant improvement in the technique is possible with the existing coaxial connectors. For example in Beatty's modification, which was based on a variation of the slotted line technique, the measured efficiency included the

losses in that portion of the slotted section between the probe position and bolometer mount, leaving one with the awkward additional problem of measuring and applying as a correction the losses in this portion of the slotted line.

3. General Theory

The contributions of the new method to be described in the following paragraphs include an "exact" procedure for determining the vector difference between two reflection coefficient values, and a tuning procedure which virtually eliminates the potential error from connector discontinuity.

The magnitude of the vector difference between two reflection coefficient values may be determined with the help of the circuit arrangement shown in figure 3. For convenience an ideal reflectometer will be assumed. Terms of the form $|\Gamma_3-\Gamma_1|$ may be determined as follows: The barretter resistance is first adjusted such that the bolometer mount has the input reflection coefficient Γ_1. The signal at the detector 3 is then "nulled" or balanced out by means of the auxiliary signal channel. Under these circumstances the signal provided by the auxiliary channel is proportional to $-\Gamma_1$. If the barretter resistance is now adjusted to produce Γ_3, and if suitable precautions (to be described later) have been taken to isolate the two channels, the detector 3 signal will now be proportional to $|\Gamma_3-\Gamma_1|$ as required.

In order to illustrate the improved tuning procedure it will prove helpful to consider a special case. If a perfectly matched bolometer mount is assumed, the tuning of the reflectometer may be effected as follows. First, with the bolometer mount connected and the bolometer resistance adjusted to the value corresponding to the perfect match, the transformer T_x is adjusted for a null in arm 3, corresponding to infinite directivity for the associated coupler. Next the bolometer mount is replaced by a sliding short and the transformer T_y adjusted such that the ratio $\left|\frac{b_3}{b_4}\right|$ is constant for all positions of the short. As was explained in the reference [6], this adjustment produces an equivalent generator match (provided that the previous adjustment has been correctly made), and having completed these two adjustments, the measurement of the mount efficiency may be carried out using eq (2) as already described. (In practice additional tuning elements, whose use will be described later, are required to isolate the auxiliary channel.)

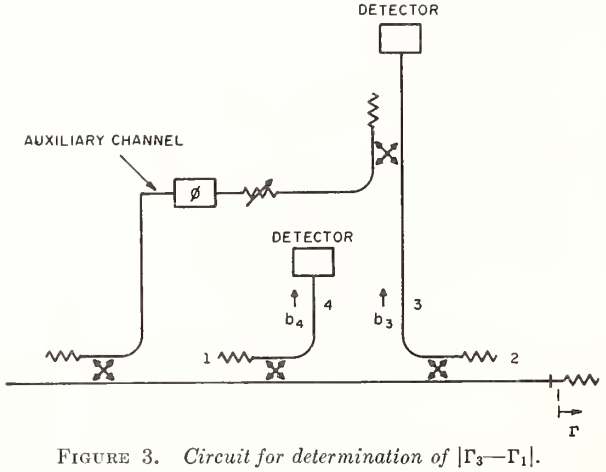


FIGURE 3. *Circuit for determination of* $|\Gamma_3-\Gamma_1|$.

It should be noted that to the extent that the (coaxial) connectors may be considered dissipation free, the attendant impedance discontinuity has no effect upon the sliding short adjustment since an ideal sliding short preceeded by a lossless discontinuity still presents a reflection coefficient of unit magnitude and (except for trivial cases) of variable phase angle. The place where the connector discontinuity is potentially important is in the initially assumed impedance match for the bolometer mount. But, as will be proved, *the assumed impedance match is entirely unnecessary.* To be sure, if the mount is matched, the measurement may be carried out as described; while if the mount is not matched, the identical series of measurements and operations described above will yield the factor

$$\frac{|(\Gamma_3-\Gamma_2)(\Gamma_1-\Gamma_2)|}{|\Gamma_3-\Gamma_1|(1-|\Gamma_2|^2)}$$

as required in the more general expression for mount efficiency.

In other words, the reflectometer adjustment and measurement proceeds on the basis of an assumed perfect impedance match for the bolometer mount (but employing only the specific procedures mentioned). It then develops that the method gives the correct value of efficiency regardless of whether the mount is matched or not! The proof for these assertions will be found in the following paragraphs.

It is a general property of a four-arm junction (of which the reflectometer of figure 2 may be regarded as a special case) that the ratio of the emergent wave amplitudes in arms 3 and 4 may be written in the form:

$$\frac{b_3}{b_4}=\frac{A\Gamma+B}{C\Gamma+D} \tag{4}$$

where the A, B, C, D are functions of the parameters of the four-arm junction and the detectors terminating arms 3 and 4, and Γ is the reflection coefficient of the load terminating arm 2.

If the reflection coefficient of the bolometer mount is designated by Γ_2 when the bolometer is at its nominal operating value R_2 (the value at which the mount efficiency is to be determined) and the junction has been adjusted (by means of T_x) such that the signal b_3 vanishes when arm 2 is terminated by this load (Γ_2), then $B=-A\Gamma_2$. A solution of eq (4), subject to the condition that the ratio $\left|\frac{b_3}{b_4}\right|$ remain constant (as provided by adjustment of T_y) while the phase of Γ varies, yields $\frac{B}{A}=\left(\frac{C}{D}\right)^* |\Gamma|^2$ where the asterisk (*) denotes the complex conjugate.

(A second solution, $A/C=B/D$ is trivial since it gives for $\frac{b_3}{b_4}$ a value which is independent of Γ.) For a sliding short the above result becomes $\frac{B}{A}=\left(\frac{C}{D}\right)^*$. Substituting the first and last of these results in eq (4) gives:

$$\frac{b_3}{b_4}=\frac{A(\Gamma-\Gamma_2)}{D(1-\Gamma\Gamma_2^*)}. \tag{5}$$

Substituting Γ_1 and Γ_3 for Γ and taking the absolute value of the ratio of the product to the difference gives:

$$\frac{\left|\frac{A}{D}\right|^2\left|\frac{\Gamma_3-\Gamma_2}{1-\Gamma_3\Gamma_2^*}\right|\cdot\left|\frac{\Gamma_1-\Gamma_2}{1-\Gamma_1\Gamma_2^*}\right|}{\left|\frac{A}{D}\right|\cdot\left|\frac{\Gamma_3-\Gamma_2}{1-\Gamma_3\Gamma_2^*}-\frac{\Gamma_1-\Gamma_2}{1-\Gamma_1\Gamma_2^*}\right|}=\left|\frac{A}{D}\right|\frac{|\Gamma_3-\Gamma_2||\Gamma_1-\Gamma_2|}{|\Gamma_3-\Gamma_1|(1-|\Gamma_2|^2)} \tag{6}$$

The term on the right is just the one required in the general expression for mount efficiency, while the factor $\left|\frac{A}{D}\right|$ may be determined by observing the system response with the bolometer mount replaced by a short or other termination for which $|\Gamma|=1$. This may be demonstrated by letting $\Gamma=e^{i\theta}$. Then:

$$\left|\frac{b_3}{b_4}\right|=\left|\frac{A}{D}\right|\cdot\left|\frac{e^{i\theta}-\Gamma_2}{1-e^{i\theta}\Gamma_2^*}\right|=\left|\frac{A}{D}\right|\cdot|e^{i\theta}|\cdot\left|\frac{1-\Gamma_2e^{-i\theta}}{1-(\Gamma_2e^{-i\theta})^*}\right|=\left|\frac{A}{D}\right| \tag{7}$$

as required.

Comparison of eqs (5) and (6) with (3) and (2) respectively, indicates that the measurement may be effected as outlined with $\left|\frac{A}{D}\right|$ taking the part of k, and $\frac{\Gamma-\Gamma_2}{1-\Gamma\Gamma_2^*}$ being the more general counterpart of Γ.

4. Circuit Arrangement and Tuning Procedure

A practical waveguide system utilizing these concepts is shown in figure 4. For convenience the directional couplers will be designated by P, Q, R, and S. The auxiliary arm between couplers R and S contains an on-off switch, phase shifter, and attenuator as required to balance out the signals from coupler P. As a matter of operating convenience it is generally helpful (but not essential) to adjust the system in such a way that the signal at detector 4 is nominally constant or independent of Γ. This is achieved by an adjustment of T_z to be described.

In order to obtain the vector difference between two reflection coefficients by the procedure described earlier, it is important to investigate the possible effect of variations in Γ upon the signal delivered to the detector 3 via the auxiliary arm. It is intuitively evident (and proof will be omitted) that the desired operation will be achieved if the following criteria are satisfied.

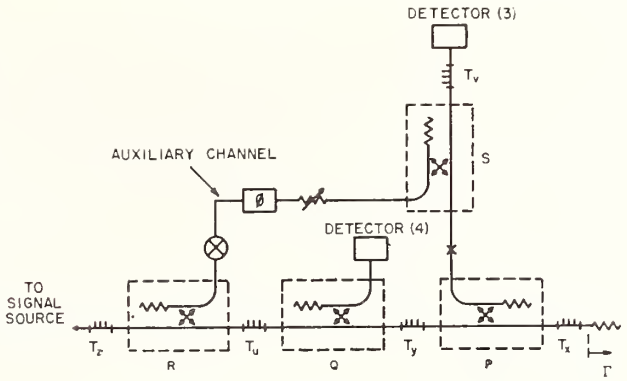


FIGURE 4. *Waveguide system for measuring bolometer mount efficiency.*

1. The signal coupled into the sidearm of coupler R is independent of Γ, or more specifically since it is the ratio of the signal in detector 3 to detector 4 which is observed, this ratio must be independent of Γ.

2. The impedance "looking into" coupler S at its sidearm is independent of Γ.

As a consequence of criterion 2, it will also be true that the impedance looking into this coupler and detector from the main arm will be independent of the operation of the phase shifter, attenuator, or switch in the auxiliary arm.

The adjustment of the tuning transformers to achieve these conditions may be done by the following steps:

(1) Directional coupler P and tuner T_y are temporarily removed from the system. In their place a sliding short is connected to coupler Q, and an auxiliary detector connected at the point X. The microwave signal is then delivered to the detector 3 via the auxiliary arm, and tuner T_v is adjusted for a null in the auxiliary detector connected at point X. This satisfies criterion 2 above.

(2) Tuner T_u is now adjusted such that the ratio $\left|\frac{b_3}{b_4}\right|$ is constant for all positions of the sliding short, and T_z is similarly adjusted such that $|b_4|$ remains constant as the position of the short is varied. If, as is often the case, the type of instrumentation employed to measure $|b_4|$ and $\left|\frac{b_3}{b_4}\right|$ is of such a nature that the measurement of $\left|\frac{b_3}{b_4}\right|$ is awkward in the presence of large variations in $|b_4|$, T_z may be first adjusted such that $|b_3|$ remains constant as the sliding short is varied, and T_u then adjusted for a constant value of $|b_4|$. Ideally, the two procedures will yield identical results, while in practice, depending upon the degree of precision toward which one is working, it may prove desirable to check the result for conformity to the criteria stated earlier.

If attention is centered on the four-arm junction comprised of couplers R, Q, and tuner T_u, it is of interest to note that the adjustment of T_u to make

the ratio $\left|\frac{b_3}{b_4}\right|$ a constant corresponds to the "trivial" solution $A/C=B/D$ rejected in the earlier discussion of a reflectometer. Although a sliding short has been specified, other types of variable loads, for example, a fixed short preceded by a variable attenuator, might also be employed in this particular adjustment. For practical purposes, the method of connecting the directional couplers will determine which of the solutions $A/C=B/D$ or $B/A=(C/D)^*$ is achieved by the sliding short adjustment, but should the question ever arise as to which of the two conditions has been realized by the given procedure it is sufficient to note that the former solution makes $\left|\frac{b_3}{b_4}\right|$ completely independent of Γ while this is not true of the latter one.

(3) The coupler P and tuner T_y are next replaced, the bolometer mount connected and T_z adjusted for a null at detector 3 with the bolometer resistance at its nominal value. The switch in the auxiliary arm is in the "off" position for this and the subsequent adjustments.

(4) The bolometer mount is then replaced by the sliding short and T_y adjusted for a constant ratio $\left|\frac{b_3}{b_4}\right|$ as the position of the short is varied. This completes the alinement procedure.

If the ratios $\left|\frac{b_3}{b_4}\right|$ are designated by subscripts 1, 3, and S, corresponding to values for the bolometer resistance of R_1, R_3, and the response when the bolometer mount is replaced by a short, the desired result may be expressed in terms of the measured values:

$$\frac{\left|\frac{b_3}{b_4}\right|_1 \cdot \left|\frac{b_3}{b_4}\right|_3}{\left|\left(\frac{b_3}{b_4}\right)_3 - \left(\frac{b_3}{b_4}\right)_1\right| \cdot \left|\frac{b_3}{b_4}\right|_S} = \frac{|(\Gamma_3-\Gamma_2)(\Gamma_1-\Gamma_2)|}{|\Gamma_3-\Gamma_1|(1-|\Gamma_2|^2)} \qquad (8)$$

where the first factor in the denominator of the left side is determined in the manner described at the beginning of section 3.

5. Practical Considerations

By means of the techniques described in the preceding sections, the microwave impedance measurements required for a determination of mount efficiency have been reduced to a set of tuning adjustments and the measurement or determination of the ratios $\left|\frac{b_3}{b_4}\right|_1$, $\left|\frac{b_3}{b_4}\right|_3$, etc., for which a variety of techniques are available. Depending upon the type of instrumentation employed, the ratio $\left|\frac{b_3}{b_4}\right|$ may be indicated directly, or the terms $|b_3|$ and $|b_4|$ may be obtained individually. If the adjustment of T_z has been made with sufficient care, the signal $|b_4|$ will depend only upon the output level of the signal source; and if the latter is sufficiently stable, $|b_4|$ will be constant and thus may be cancelled out of eq (7), leaving only the values of $|b_3|$ to be determined. Indeed, as discussed in the reference [6], if a sufficiently stable signal source is used the directional coupler Q, detector 4 and tuner T_u may be omitted from the system. Under these conditions, step (2) of the tuning procedure described above consists of adjusting T_z such that $|b_3|$ is constant, while the other steps follow as given with $\left|\frac{b_3}{b_4}\right|$ replaced by $|b_3|$.

The detection techniques which may be employed include audio modulation and detection, heterodyne detection, and bolometric power detection. Although it is not within the scope of this paper to examine each of these methods in detail, it may prove useful to call attention to some of their more prominent features.

Perhaps the most convenient, but one of the least accurate, in the existing state of the art, of the cited methods is that of conventional audio modulation and detection. The detection equipment may take the form of a ratio-meter with barretter detectors such that $\left|\frac{b_3}{b_4}\right|$ is indicated directly. Barretter detectors are specified in order to ensure that the detector impedance will be at least nominally constant as required above. Alternatively, a standing-wave type amplifier may be used, and the signal $|b_4|$ applied to the automatic gain control channel (if available). It should be noted that such an instrument is actually a ratio-meter although the permissible excursion of the input signal applied to the AGC channel is usually quite limited. In practice these excursions may be held to a small value by adjustment of T_z, and once again, if the signal source is sufficiently stable and the adjustment of T_z is made with sufficient care, it is only necessary to measure $|b_3|$, and the AGC channel is not required. The chief problem with conventional audio techniques appears to be in the difficulty of accurately measuring changes in the level of an audio signal. In addition, because the measurements must be made over a nominal 20 db dynamic range, the deviations of the barretter detector from true square law response would be ultimately a source for concern.

Another objection to the conventional audio technique is that the barretter resistance (in the mount to be measured) tends to follow the modulation envelope, in the same manner as occurs in the detectors. In order to avoid this difficulty it is necessary to hold the microwave power dissipated in the barretter to a small fraction, nominally 1 percent or so, of the total bias power. In addition, the resistance excursions which may be realized with the commercially available barretters are such that the reflected signal is down by another nominal 20 db, and thus even if the modification (to be described below) which permits tight coupling to this reflected signal is employed, the signal power available is only of the order of a few microwatts, which taxes the ability of the best audio detection equipment. In spite of these objections, however,

the audio technique does provide a convenient method of developing a familiarity with the overall system behavior, and as this is being written, some of these objections are being overcome [7]. When used in conjunction with some of the additional refinements to be discussed below, the conventional audio technique would probably permit an efficiency determination to an accuracy of a few percent or better.

The IF substitution technique of attenuation measurement is immediately applicable to the measurement of relative levels of $|b_3|$, while a possible scheme for the measurement of $\left|\frac{b_3}{b_4}\right|$ is shown in figure 5. In figure 5, the 30 Mc/s signal derived from b_4 is adjusted by means of the piston attenuator to equal in amplitude that obtained from b_3, while phase balance is obtained from the phase shifter as shown. The changes in $|b_3|$ with respect to $|b_4|$ may thus be read directly from the piston attenuator. The measurement procedure thus consists of a series of nulling adjustments and is independent of fluctuations in the amplitude of the signal source, or the gain of the 30 Mc/s null detector. For these reasons the scheme appears attractive (on paper at least) for the proposed application.

A preliminary investigation of this technique has shown, however, the existence of a number of attendant practical problems. First, the 30 Mc/s circuits are frequency sensitive, requiring high stability of the intermediate frequency. Second, for proper operation the conversion efficiency of the crystal mixer should be independent of the changes in level of the local oscillator signal which inevitably accompany the operation of the phase shifter, but the conversion gain proved to be more sensitive to the local oscillator amplitude than anticipated. Third, the fact that the null is phase sensitive complicates the alignment procedure somewhat, particularly the adjustment of T_y. Undoubtedly, a more careful examination of the technique will uncover additional problems; this technique has not been investigated further because the method described below proved to be more convenient with the auxiliary apparatus on hand.

The method referred to above consists of power detection by the bolometric technique. These techniques have been refined to the point [8] where accuracies of 0.1 to 0.2 percent may be realized at the required power levels. (It will be noted that questions as to bolometer mount efficiency of the detectors, etc., are unimportant here, since only power *ratios* are required.) It will prove useful to consider certain aspects of this procedure in detail since they are also applicable to the techniques described earlier.

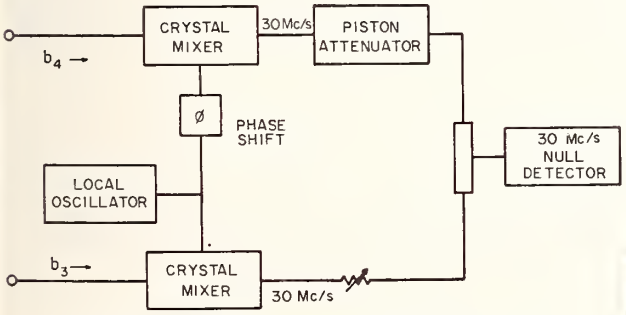


FIGURE 5. *Possible heterodyne detection system.*

The techniques by which the power is measured were discussed in detail in the cited reference [8] and need not be described here. In order to increase the signal at the detectors to the level where power detection is feasible, the configuration shown in figure 6 is employed. With the exception of the reversal of the connections to the coupler P, the system and tuning procedure are the same as that given for figure 4. A convenient choice of coupling ratio for the couplers P, Q, R, S, is 10 db. Under these conditions, approximately 80 percent of the power reflected from the termination is delivered to the detector 3. A further increase in this figure may be realized by other choices of coupling ratios, for example a figure of 90 percent may be realized by changing R and S to 3 and 20 db respectively, while a value of 98 percent would obtain with the values 20, 20, 3, and 20 db for P, Q, R, and S, respectively. These two alternatives would require an additional 3 and 10 db of power from the generator, however, and thus the first set of values appears to be a reasonable compromise when the available power is a potential problem.

The power delivered to the detector 4 is stabilized or held constant by means of techniques developed in this laboratory, [9] and thus only the values of power at the detector 3 are required.

In order that the detector signal may be as large as possible, it is desirable to operate the barretter, in the mount to be calibrated, over as wide a resistance excursion as possible. The commercially available barretters are not all equally suitable in

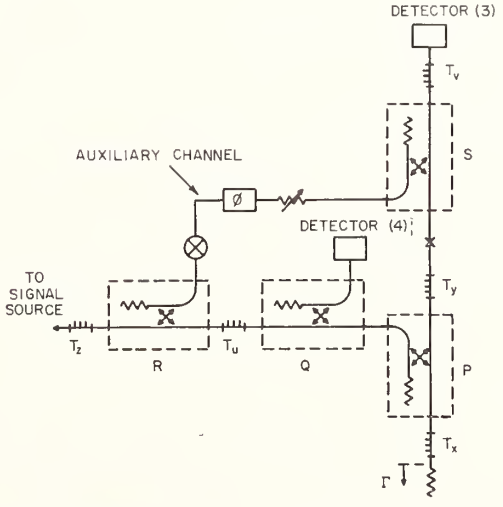


FIGURE 6. *Alternative waveguide system for measuring bolometer mount efficiency.*

this regard, but the type selected for this application may be operated over the range 140 to 260 ohms (nominal operating resistance 200 ohms). In operation the barretter forms one arm of a self-balancing bridge [8], another arm of which is varied to produce the required changes in resistance.

The bolometer mount under measurement is provided with a nominal 10 mw of microwave power, resulting in reflected signals of approximately 310 and 170 μw at the resistance values of 140 and 260 ohms. A practical problem exists with reference to the measurement at 140 ohms, since the 10 mw of microwave power alone, without any additional d-c power, will bias the element at perhaps 180 ohms. This difficulty is avoided by an indirect measurement technique. In order to obtain the measurement corresponding to the 140-ohm resistance value the microwave power is first reduced to such a value that the element may be operated at 140 ohms. The signal in the detector 3 is then nulled or balanced out by means of the auxiliary signal channel. Assuming linearity of the system, this null will be independent of power level. The bolometer resistance is then returned to 200 ohms, and the microwave power increased to its nominal 10 mw value. Under these conditions, the only signal delivered to detector 3 will come from the auxiliary channel and will be equal in amplitude (but opposite in phase) to the signal which would be obtained from the bolometer mount were it possible to reduce its resistance to the 140 ohm value in the presence of the 10 mw signal. This signal is then measured and used in the equations where the 140 ohm value is required.

With the bolometer mount replaced by a short, the reflected signal has, of course, a power level equal to that in the incident wave, nominally 10 mw. The procedure thus requires measurements over the nominal 20 db range between the 10mw and 100 μw power levels. In addition to the measurement of these signal levels, it is also important to be able to recognize the presence of signal levels in the region 60 to 80 db below these values during the course of the tuning operations. For this purpose it has proved convenient to split off a portion of the signal from arm 3 by means of a directional coupler and employ an auxiliary heterodyne detector. A waveguide switch in the secondary arm of this coupler is kept in the nontransmitting position when the adjustment of tuner T_v is made and when the signal amplitudes are being measured by the bolometric detector. In other words, the switch is in the transmitting position only when the high sensitivity for tuning adjustments is required. This isolates the remainder of the system from possible variations in the impedance of this heterodyne detector.

Although little has been said concerning the details of the power measurement, this aspect of the procedure is not without its share of practical problems. In order to make power measurements to an accuracy of a few tenths of a percent at the 100μw level, temperature control of the bolometer mount to about 0.001 ° C is required. A second problem concerns the barretter element itself. In practice it is frequently observed that even though the ambient temperature is held constant, the bias power required to maintain a given operating resistance may still be subject to considerable variation or drift. Certain types of elements have been observed to be much worse than others in this respect, but even among the better ones it is usually necessary to apply the bias power continuously for several days or even a week before stable operation is achieved.

Although the foregoing discussion has been concerned primarily with the barretter element, the measurement also imposes stringent stability requirements on the associated apparatus as well. This will be easily recognized when attention is called to the fact that the bolometric measurement is in reality a differential power measurement. Thus if it is required to measure a one percent change to an accuracy of a part in 10^3, a stability or resolution in the total d-c bias power of a part in 10^5 is implied.

6. Approximate Methods

The foregoing procedure is "exact" in the sense that the only mathematical approximations which have been introduced thus far have been in assuming that eqs (1) and (4) adequately describe the bolometer mount and reflectometer behavior. In a large percentage of the cases of practical interest a considerable simplification in the technique may be effected with little loss in accuracy by the introduction of suitable approximations. As a matter of fact, the accuracy achieved in practice is actually improved in certain instances by one of these approximations which will be described.

The first of these approximations relates to the substitution of $||\Gamma_3|\pm|\Gamma_1||$ for $|\Gamma_3-\Gamma_1|$ in eq (2), or more generally, the substitution of

$$\left|\frac{\Gamma_3-\Gamma_2}{1-\Gamma_3\Gamma_2^*}\right|\pm\left|\frac{\Gamma_1-\Gamma_2}{1-\Gamma_1\Gamma_2^*}\right| \text{ for } \left|\frac{\Gamma_3-\Gamma_2}{1-\Gamma_3\Gamma_2^*}-\frac{\Gamma_1-\Gamma_2}{1-\Gamma_1\Gamma_2^*}\right|$$

in eq (6). (The minus sign is used if both R_1 and R_3 are either greater than or less than R_2, while the positive sign is used when, as in the following example, one resistance value is above and the other below R_2.)

In order to examine the validity of this approximation it is convenient to obtain a general expression for this error in terms of the magnitude and arguments of the complex numbers Γ_1, Γ_3, or their more general counterparts $\frac{\Gamma_1-\Gamma_2}{1-\Gamma_1\Gamma_2^*}$, etc. Let $A=|A|$ and $B=|B|e^{i(\pi-\phi)}$ represent two complex numbers. Then for small values of ϕ, the error in replacing the vector difference by the sum of the magnitudes is given by the expression:

$$\text{Error}=\frac{|AB|}{(|A|+|B|)^2}(1-\cos\phi), \qquad (9)$$

The input reflection coefficient of the bolometer mount is given by the equation [2]:

$$\Gamma = S_{11} + \frac{S_{12}^2 \Gamma_l}{1 - S_{22}\Gamma_l} \tag{10}$$

where Γ_l is the load reflection coefficient, and the $S_{m,n}$ are the scattering coefficients of the bolometer mount (see fig. 7).

A substantial simplification in the analysis results from the fact that it is permissible to choose the impedance normalizing parameters in such a way that $\Gamma_l = 0$ when the bolometer is at its nominal operating value R_2. For any other value of R, Γ_l has the value

$$\Gamma_l = \frac{R - R_2}{R + R_2}.$$

Utilizing these results, the term

$$\frac{\Gamma_1 - \Gamma_2}{1 - \Gamma_1 \Gamma_2^*}$$

becomes:

$$\frac{\Gamma_1 - \Gamma_2}{1 - \Gamma_1 \Gamma_2^*} = \frac{\dfrac{S_{12}^2 \Gamma_{l1}}{1 - |S_{11}|^2}}{1 - \left(S_{22} + \dfrac{S_{11}^* S_{12}^2}{1 - |S_{11}|^2}\right)\Gamma_{l1}}. \tag{11}$$

Let Ψ_{12} denote the argument of S_{12} and let

$$F = \left(S_{22} + \frac{S_{11}^* S_{12}^2}{1 - |S_{11}|^2}\right); \tag{12}$$

then, since the efficiency η of the bolometer mount when terminated in R_2 is given by the expression [2]:

$$\eta = \frac{|S_{12}|^2}{1 - |S_{11}|^2}, \tag{13}$$

eq (11) may be written:

$$\frac{\Gamma_1 - \Gamma_2}{1 - \Gamma_1 \Gamma_2^*} = \frac{\eta \Gamma_{l1} e^{j2\psi_{12}}}{1 - F\Gamma_{l1}}, \tag{14}$$

and

$$\arg \frac{\Gamma_1 - \Gamma_2}{1 - \Gamma_1 \Gamma_2^*} = 2\psi_{12} - \tan^{-1} \frac{-\mathrm{Im}\,(F)\Gamma_{l1}}{1 - \mathrm{Re}\,(F)\Gamma_{l1}}. \tag{15}$$

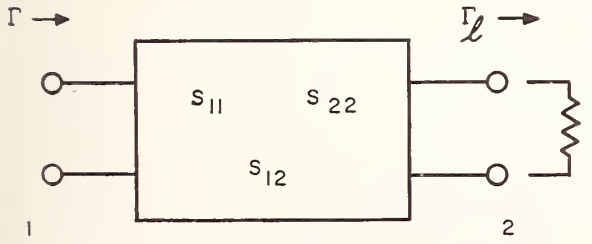


FIGURE 7. *Scattering coefficient representation of bolometer mount.*

It is shown in the appendix that $|F|$ satisfies the inequality

$$|F| \leq 1 - \eta. \tag{16}$$

In a large percentage of the cases of practical interest the efficiency will be in excess of 90 percent, while $|\Gamma_l|$ is ordinarily limited to the range 0.1 to 0.2. Substituting these results in eq (9) yields the approximate result:

$$\mathrm{Error}_{(\max)} \approx \frac{1}{2}|\Gamma_{l1}\Gamma_{l3}|(1-\eta)^2. \tag{17}$$

This is the approximate maximum limit of error which obtains when arg $F = \pm\frac{\pi}{2}$. For other values of arg F the error will be smaller, becoming zero for arg $F = 0$ or π.

In a typical case where the bolometer is operated at the resistance values 140–200–260 ohms and for a mount efficiency of 90 percent the error is approximately ≤ 0.012 percent. The uncertainty introduced by replacing the vector difference by the sum of the magnitudes in this example is thus wholly negligible. Under these conditions it will thus be recognized that the auxiliary waveguide channel, which was introduced to permit a determination of this vector difference, contributes virtually nothing to the accuracy of the final result as far as this aspect of the operation is concerned.

The auxiliary loop is still a practical necessity, however, where power detection is employed because of the problem with the 140-ohm measurement as outlined above. In addition, this auxiliary loop may contribute to the accuracy of the result as follows:

Let the resistance excursion of the bolometer element be chosen in such a way that $\Gamma_{l1} = -\Gamma_{l3}$; then it can be shown that:[3]

$$\frac{1}{4}\left|\frac{\Gamma_3 - \Gamma_2}{1 - \Gamma_3\Gamma_2^*} - \frac{\Gamma_1 - \Gamma_2}{1 - \Gamma_1\Gamma_2^*}\right| \approx \left|\frac{(\Gamma_3 - \Gamma_2)(\Gamma_1 - \Gamma_2)}{(\Gamma_3 - \Gamma_1)(1 - |\Gamma_2|^2)}\right| \tag{18}$$

with an error

$$\mathrm{Error} \leq \Gamma_{l1}^2(1-\eta)^2 \tag{19}$$

which is approximately twice the value obtained in eq (17).

One of the substantial sources of error attending this technique occurs in practice in the measurement of the signal amplitudes (corresponding to the various resistance values) at the detector 3. As outlined above, a total of four measurements are required, while use of the first approximation described reduces this number to three. Use of the approximation (18) however, reduces the number of required measurements to two, the vector difference and the measurement with the bolometer mount replaced by the

[3] The proof may be effected by substituting eq (10), (12), and (16) in (18) and simplifying the resultant expressions.

short. In addition the power level which obtains at detector 3 in measuring this vector difference is four times that obtained from Γ_{l1} alone which reduces the dynamic range and improves the accuracy with which these measurements can be made. In the existing state of the art this latter technique is the most accurate, for mounts of high efficiency, of those described.

7. Measurement of Attenuation

The foregoing presentation has been directed primarily at the problem of measuring bolometer mount efficiencies. The efficiency, however, determines the dissipative component of attenuation, and a modification of the techniques described should also prove useful in other applications, such as measuring the efficiency of waveguide to coax adaptors, short sections of waveguide, etc. A complete discussion of these potential applications and the requisite modifications in the method is not, however, within the scope of this paper.

8. Error Analysis

In any technique or method with potential applications in the field of standards work, a thorough investigation of the attendant sources of error is an important part of the description. The sources of error to be discussed are the following: (1) Error due to misadjustment of tuning transformers, (2) error due to dissipation in reference short, (3) error due to uncertainty in the values of bolometer resistance, and (4) error in the measurement of the signal amplitudes. These will be treated in the order listed.

The error analysis or study which has been made in connection with this technique, is, however, too lengthy to be presented in its entirety. For this reason the discussion to follow will be largely of a summary nature, and devoted primarily to the waveguide system of figure 6. The extension of these results to other variations of the method, however, should prove straightforward. The procedure to be followed in the analysis is that of assuming ideal operation of the system except for the source of error being considered. This technique yields the first order correction terms to the ideal theory for individual sources of error.

8.1. Error Due to Misadjustment of Tuning Transformers

a. Misadjustment of Tuner T_v

The adjustment of T_v and T_u is important in obtaining the vector difference $(\Gamma_1-\Gamma_3)$ etc. As a consequence of the adjustment of tuner T_u the signals delivered to detector 3 via arms 1 and 3 of directional coupler S (see fig. 6) may be considered as originating from two independent sources (of constant amplitude and frequency). If tuner T_v is not properly adjusted, the signal delivered via arm 3 will vary with changes in Γ (as produced by changes in bolometer resistance). If the "null" observed at X in step 1 of the tuning procedure is down from the signal at detector 3 by at least 60 db, then the voltage amplitude of the signal emerging from arm 1 will be down by a factor of 10^3. During the course of the measurement, Γ varies between the approximate limits of $\pm 1/7$, and the maximum anticipated variation in the signal delivered to detector 3 via the auxiliary channel due to this variation is approximately 3 parts in 10^4. Finally, this is approximately one-half the total signal being measured, so the maximum anticipated error is one-half this value and may be neglected.

b. Misadjustment of Tuner T_u

The problem once again is one of attempting to insure that the signal delivered to detector 3 via the auxiliary loop is independent of Γ. This time the possible interaction is through coupler R instead of coupler S. In practice the variation in reflection coefficient observed "looking in" arm 3 of coupler P during the course of the measurement will be approximately ± 0.015. If with the sliding short connected to the output of coupler Q (tuning operation 2), T_u is adjusted so that the maximum variations in $\left|\frac{b_3}{b_4}\right|$ are reduced to 1 percent (0.1 percent may be realized in practice without too much difficulty), the maximum variation in the auxiliary channel signal will be of the order of a part or two in 10^4 and again may be neglected.

c. Misadjustment of Tuner T_x

The function of tuner T_x is to adjust the four arm junction in such a way that it satisfies the condition $B=-A\Gamma_2$. The failure to exactly realize this adjustment may be accounted for by letting $B=-A(\Gamma_2+\delta)$. Substituting this result in the equation for determining the bolometer mount efficiency yields two approximate upper limits for the error due to this source,

$$E\leq\frac{1}{1-|\Gamma_2|^2}\left[\frac{1}{\eta}\frac{|\Gamma_{l1}+\Gamma_{l3}|}{|\Gamma_{l1}\,\Gamma_{l3}|}+2|\Gamma_2|+4\left(\frac{1}{\eta}-1\right)\right]|\delta| \quad (20)$$

and

$$E\leq\frac{1}{1-|\Gamma_2|^2}\left[\frac{1}{\eta}\frac{|\Gamma_{l1}+\Gamma_{l3}|}{|\Gamma_{l1}\,\Gamma_{l3}|}+4|\Gamma_2|+2\left(\frac{1}{\eta}-1\right)\right]|\delta|. \quad (21)$$

The choice of which expression to use obviously depends upon the relative magnitudes of Γ_2 and $\left(\frac{1}{\eta}-1\right)$. It will be noted that the error can be reduced by choosing the resistance variation such that $\Gamma_{l1}=-\Gamma_{l3}$, and if the VSWR (σ) of the mount is less than 1.02, the coefficient of $|\delta|$ has the approximate limit of 0.2 for an efficiency greater than 90 percent. On the other hand, for a resistance excursion of 140 to 200 to 260 ohms and $\sigma=1.5$ as in an earlier example, this coefficient will have the approximate limit 2.8. The desirability of choosing resistance values such that $\Gamma_{l1}=-\Gamma_{l3}$ is readily apparent.

If the approximate method implied by eq (18) is employed, the approximate limit of error is given by the expression

$$E \leq 2\left[\frac{|\Gamma_2|}{1-|\Gamma_2|^2}+\eta(1-\eta)\Gamma_{l1}^2\right]\cdot|\delta|. \qquad (22)$$

In this expression the coefficient of $|\delta|$ has the approximate values of 0.06 and 0.5 for VSWR's of 1.02 and 1.5 respectively. The use of this approximation thus reduces the error.

If the tuner T_x has been adjusted such that the signal level at detector 3 is down by 60 db from the level which exists with the shorting plate connected in place of the bolometer element, $|\delta|$ will have the nominal value of 10^{-3}, while an 80 db "null" will yield a value for $|\delta|$ of 10^{-4}. The error due to this source can thus usually be held to 0.1 percent or less.

d. Misadjustment of Tuner T_y

The function of tuner T_y is to adjust the junction such that it satisfies the condition $D/C=(A/B)^*$. The failure to exactly realize this condition may be accounted for by letting $D/C=(A/B)^*+\epsilon$. Substituting this result into the equation for bolometer mount efficiency yields the approximate expression $E\approx|\epsilon|$ for the error due to this source. This error arises primarily in the determination of $|A/D|$ by means of the fixed short.

The failure to realize the proper adjustment of T_y can result from either failure to entirely eliminate the variations in $\left|\frac{b_3}{b_4}\right|$ as the sliding short is moved, or from deviations of the sliding short from the assumed ideal behavior. An analytic treatment of this problem yields for $|\epsilon|$ the approximate limit:

$$|\epsilon| \leq (1+2|\Gamma_2|)\left[\frac{1}{4}\frac{P_{max}-P_{min}}{P_{av}}+(1-|\Gamma_m|^2)|\Gamma_2|+(1-\eta_j)\right] \qquad (23)$$

where P_{max}, P_{min}, and P_{av} represent the maximum, minimum, and average values of the ratio $\left|\frac{b_3}{b_4}\right|^2$ as observed during the course of the sliding short adjustment, Γ_m is the reflection coefficient of the sliding "short", and η_j is the efficiency of the waveguide joint which connects the sliding short to the remainder of the waveguide system. In the above expression the waveguide in which the short slides has been assumed lossless, while in practice, when the correct adjustment has been realized, there will be a gradual and uniform decrease in the ratio $\left|\frac{b_3}{b_4}\right|$ as the sliding short is withdrawn.

The error due to η_j may be eliminated by including a precision waveguide section in arm 2, as shown in figure 8, into which the sliding short may be inserted for this tuning operation. Other implications of this arrangement will be discussed in subsequent paragraphs. It will be noted that the error due to the reflection coefficient of the sliding short (Γ_m) differing from unit magnitude decreases as the impedance match of the bolometer mount is improved. In practice the variations in $\left|\frac{b_3}{b_4}\right|^2$ can usually be reduced to the order of several parts in 10^3 such that the error from these sources can usually be held to 0.1 to 0.2 percent.

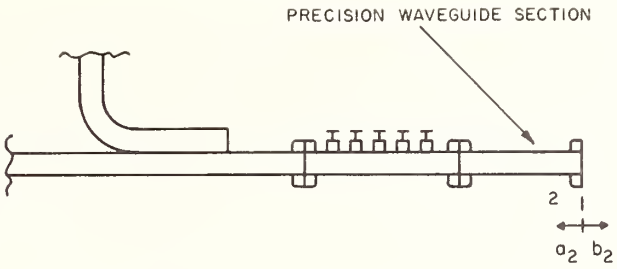


FIGURE 8. *Illustration of use of precision waveguide section.*

e. Error Due to Imperfect Nulling When Using Auxiliary Channel

If this "null" is measured in terms of the signal level which exists in the detector prior to the nulling procedure it will be readily recognized that a 60 db minimum corresponds to a possible difference between the two signals of a part in 10^3, etc. Since this signal is next combined with another one of nominally the same amplitude and phase the error contribution to the final measurement is approximately one-half this value.

8.2. Error Due to Dissipation in Reference Short

The error due to this source is closely connected with section (d) and is given by the approximate expression:

$$E \leq (1+2|\Gamma_2|)(1-|\Gamma_s|) \qquad (24)$$

where Γ_s is the reflection coefficient of the reference or standard short. In the present state of the art the highest reliable reflection coefficients are provided by "quarter wave" shorts, that is, a quarter wavelength section of waveguide terminated by a soldered shorting plate, or the entire structure may be electroformed in a single piece. The advantage of this construction is in the elimination of the longitudinal component of current flow at the plane of connection, thus tending to minimize flange losses. Measurements and calculations of the reflection coefficient magnitudes of such shorts [10] have yielded values in excess of 0.999.

If such a device is used in conjunction with the uniform waveguide section as in figure 8, the measurement will include, as part of the bolometer mount losses, the dissipation occurring at the bolometer mount input flange. If, on the other hand,

the variations in $\left|\frac{b_3}{b_4}\right|$ have not been completely eliminated, there is some advantage in taking the average of the results with a quarter wave and conventional "flat" short. This tends to cancel out the error due to imperfect tuning but introduces other uncertanties if there is an appreciable amount of flange loss [11]. The general features of this problem are discussed in greater detail in the cited reference.

8.3. Error Due to Uncertainty in the Values of Bolometer Resistance

The bolometer resistances enter the expression for bolometer mount efficiency through the factor

$$K=\left|\frac{2R_2(R_3-R_1)}{(R_2-R_1)(R_3-R_2)}\right|.$$

In practice the bolometer is biased by means of a self-balancing d-c bridge [8]. This device is capable of maintaining the bolometer resistance within 0.01 percent of the value called for by the bridge parameters. Use of NBS type resistors makes it possible to maintain the value of the reference arm within 0.01 percent, while an error in the upper or ratio arms cancels out since the R's may be multiplied by an arbitrary factor without changing the value of K. The R's may thus be effectively maintained within the limits ± 0.02 percent. However, because the expression for K involves their differences, the total error due to this source is of the order of 0.1 percent. In addition the measurement will be in error by the ratio of the bolometer lead resistance to R_2.

8.4. Error Due to Measurement of Signal Amplitudes

An approximate limit to the error introduced in the measurement of the signal amplitudes is the sum of the errors in the individual measurements. For the approximation method of eq (18), this error can be kept within 0.15 percent if power detection is employed, while for other variations of the procedure and/or other detection procedures the error will in general be larger.

The error due to instability in the frequency of the signal generator has not been considered explicitly but in general will manifest itself in an inability to realize the tuning conditions to the required degree of precision.

In summary, it will be recognized that the ultimate accuracy attainable will depend primarily upon which of the several alternate procedures is employed and upon the degree of refinement employed in the attendant adjustments and measurements. The foregoing results are somewhat arbitrarily summarized in the following table. The minimum values are intended to represent the approximate best values in the existing state of the art, while the larger values are intended to be typical of what may be achieved if the requirements are relaxed somewhat.

Source of error	Nominal error limits in percent	
	$\sigma<1.02$	$\sigma=1.5$
Adjustment of T_v	0. 00–0. 015	0. 00–0. 015
Adjustment of T_u	. 00– . 02	. 00– . 02
Adjustment of T_x	. 00– . 02	. 05– . 3
Adjustment of T_y	. 05– . 3	. 1 – . 5
Imperfect null in use of auxiliary arm	. 02– . 05	. 02– . 05
Reference short	. 05– . 3	. 05– . 3
Bolometer resistance	. 1 – . 15	. 1 – . 15
Measurement of signal amplitudes	. 15–1. 0	. 15–1. 0
Total	0. 4 –1. 8%	0. 5 –2. 3%

9. Experimental Results

An X-band waveguide setup employing the concepts developed in this paper is shown in figure 9. The efficiency values obtained utilizing this method have consistently shown agreement with microcalorimeteric determinations [12] to within one-half of one percent and usually within a few tenths of a percent. This is of considerable interest since the two methods should, in principle, differ by the amount of the substitution error, and suggests that the substitution error is typically at least an order of magnitude smaller than the previously estimated limits.

10. Appendix

The inequality

$$|F|\leq 1-\eta \tag{16}$$

can be proved in the following manner.

Let the terminals of the 2 arm junction represented in figure 7 be reversed such that the load is connected to arm 1. Then the input reflection coeffi-

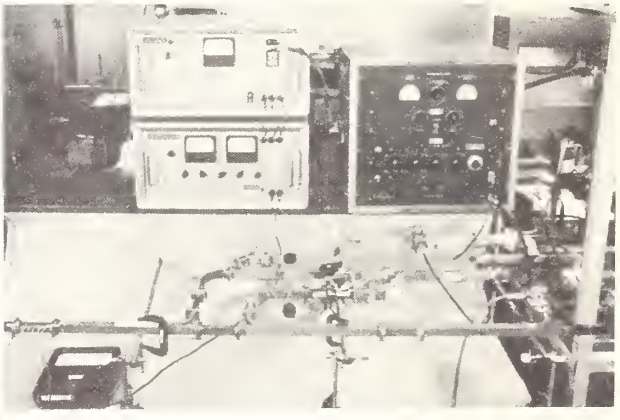


FIGURE 9. *Picture of waveguide system and accessory instrumentation employed in the measurement of bolometer mount efficiency.*

cient at arm 2 will be given by: (cf. eq 10)

$$\Gamma_{in}=S_{22}+\frac{S_{12}^2\Gamma_l}{1-S_{11}\Gamma_l}.$$

This can be written in the form,

$$\Gamma_{in}=\frac{\alpha\Gamma_l+\beta}{\gamma\Gamma_l+\delta} \tag{25}$$

where $\alpha=S_{12}^2-S_{11}S_{22}$

$\beta=S_{21}$

$\gamma=S_{11}$

$\delta=1.$ (26)

Equation (25) indicates that Γ_{in} is related to Γ_l by means of a linear fractional transformation. This transformation has the well known property [4] of mapping circles into circles with straight lines as limiting cases.

Let arm 1 be terminated by a sliding short. Then the locus of Γ_l is the unit circle centered at the origin. From the above property of the transformation it will be recognized that the locus of Γ_{in} is also a circle, and since a reflection coefficient magnitude can never exceed unity it is evident that this circle must lie completely within or on the unit circle as shown in figure 10. The distance to the center of this circle, r_c, and the radius, r, thus satisfy the inequality:

$$r_c\leq 1-r. \tag{27}$$

In terms of the constants of the transformation, r_c and r are given by

$$r_c=\frac{|\beta\delta^*-\alpha\gamma^*|}{|\delta|^2-|\gamma|^2} \tag{28}$$

and

$$r=\frac{|\alpha\delta-\beta\gamma|}{|\delta|^2-|\gamma|^2}. \tag{29}$$

Substituting eqs (26) into (28) and (29) and these in turn into (27) and recalling the definition of F, yields the result:

$$|F|=\left|S_{22}+\frac{S_{11}^*S_{12}^2}{1-|S_{11}|^2}\right|\leq 1-\eta$$

where $\eta=\frac{|S_{12}|^2}{1-|S_{11}|^2}$. See eq (13).

The author thanks R. W. Beatty and D. M. Kerns for their helpful suggestions in reviewing the manuscripts, and W. E. Case, M. E. Harvey, and J. E. Gilbert who provided experimental demonstrations of techniques described.

[4] See a text on complex variable theory.

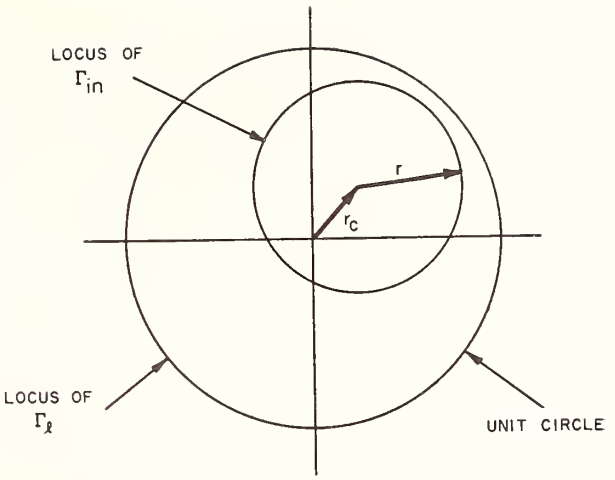


FIGURE 10. *Diagram giving locus of* Γ_l *and* Γ_{in}.

11. References

[1] D. M. Kerns, Determination of efficiency of microwave bolometer mounts from impedance data, J. Research NBS **42**, 579 (1949) RP1995.

[2] R. W. Beatty and F. Reggia, An improved method of measuring efficiencies of ultra-high-frequency and microwave bolometer mounts, J. Research NBS **54**, 321 (1955) RP2594.

[3] B. O. Weinschel, Phase engineering report No. 2 on standard signal generator output power, voltage, and attenuation calibration assembly, Contract AF 33(600)–25238, E.O. No. C–30047 SR–6J3 (May 1956).

[4] E. L. Ginzton, Microwave measurements, p. 185 (McGraw-Hill Book Co., Inc., New York, N.Y., 1957).

[5] J. A. Lane, Measurements of efficiency of bolometer and thermistor mounts by impedance methods, Proc. IEE, pt. D., No. 17, p. 485 (Sept. 1957).

[6] G. F. Engen and R. W. Beatty, Microwave reflectometer techniques, IRE Trans. on microwave theory and techniques **MTT–7**, No. 3, pp. 351–355 (1959).

[7] G. E. Schafer and R. R. Bowman, A modulated subcarrier technique ot measuring microwave attenuation, Spring Meeting, International Scientific Radio Union and Institute of Radio Engineers, Washington, D.C., May 3, 1960. (A formal report of this work is in preparation.)

[8] G. F. Engen, A self-balancing DC bridge for accurate bolometric power measurements, J. Research NBS **59**, 101 (1957) RP2776.

[9] G. F. Engen, Amplitude stabilization of a microwave signal source, IRE Trans. on microwave theory and techniques **MTT–6**, No. 2 (1958).

[10] R. W. Beatty and D. M. Kerns, Recently developed microwave impedance standards and methods of measurement, IRE Trans. on Instrumentation, **I–7**, No. 3 and 4 (1958).

[11] R. W. Beatty, G. F. Engen, and W. J. Anson, Measurement of reflections and losses of waveguide joints and connectors using microwave reflectometer techniques, IRE Trans. on Instrumentation, **I–9**, No. 2, 219–226 (Sept. 1960).

[12] G. F. Engen, A refined X-band microwave microcaolrimeter, J. Research NBS, **63C**, No. 1, 77 (1959).

(Paper 65C2–62)

A DC-RF Substitution Error in Dual-Element Bolometer Mounts

GLENN F. ENGEN

Summary—Most of the coaxial type bolometer mounts in current use employ a pair of bolometer elements which are connected in series for the dc or audio-frequency bias powers, and in parallel at radio or microwave frequencies.

In the frequency range where these techniques are most often employed, the dc-RF substitution error has been generally believed to be negligible. It is quite possible, however, for this to be true of the elements individually, and yet fail to be true of a pair of these elements as used in a typical coaxial mount. If only the *sum* of the resistances of the two elements is maintained at a constant value, and if the resistance division between the two elements changes with the application of RF power, an error is introduced which is given by the equation: $e=[(1/\gamma_b)-(1/\gamma_a)]\Delta r$, where γ_a and γ_b are the "ohms per milliwatt" coefficients of the two bolometer elements, and Δr is the shift in resistance division.

An experimental study indicates that, in the existing state-of-the-art, this error may be ignored in many applications but is large enough to be important in others, particularly at the higher power levels.

Background

IN THE ULTRA-HIGH frequency portion of the radio spectrum a bolometer mount design, for use with coaxial systems, has been widely adopted, which employs a pair of bolometer elements. As shown in Fig. 1, the configuration is such that the elements are operated in series for the dc or audio-frequency bias powers and in parallel at radio frequencies. This arrangement provides a better impedance match with existing bolometer elements and a simplification or avoidance of the problem of providing a dc return to the center conductor.

Associated with the bolometric technique are two sources of systematic error which are well recognized in the prior art: the dc-RF substitution error and bolometer mount efficiency. The first of these arises because the spatial distributions of current, power and resistance within the bolometer element are different for the dc and RF powers. Ideally, the bolometer resistance is a function of the total power absorbed and independent of its distribution. In practice, however, this condition is never completely satisfied, thus giving rise to the substitution error.

The second source of error is due to the fact that the bolometric technique, at best, measures only the power

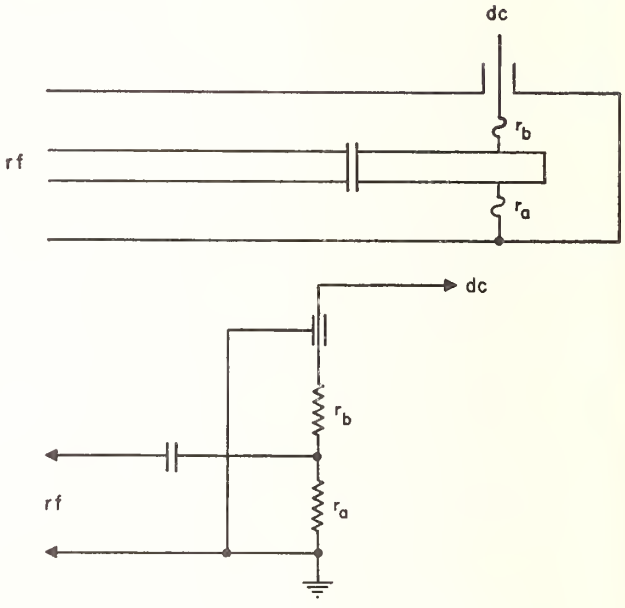


Fig. 1—Basic circuit, dual element bolometer mount.

absorbed by the bolometer. In practice one is usually interested in determining the power dissipated in the entire bolometer mount, and some of this dissipation inevitably occurs in the metal surfaces, dielectric supports, etc. The evaluation of these two phenomena has been the object of a sizable amount of research in recent years.[1-6]

Introduction

In the prior art, the substitution error has been associated with the differences in current distribution in the bolometer elements which are characteristic of the dc and RF powers due to frequency effects. However, the error from this source is generally believed to be negligi-

Manuscript received October 11, 1963; revised January 31, 1964. A preliminary account of this work was given at a 1962 Fall Meeting of URSI, Ottawa, Canada, and in the *NBS Tech. News Bull.*, April, 1963. More recently it has come to the author's attention that the existence of this phenomenon was previously noted by I. A. Harris, "A coaxial film bolometer for the measurement of power in the U.H.F. band," *J. IEE*, vol. 107, pp. 67-72; January, 1960.

The author is with the National Bureau of Standards, Boulder, Colo.

[1] H. J. Carlin and M. Sucher, "Accuracy of bolometric power measurements," PROC. IRE, vol. 40, pp. 1042-1048; September, 1952.

[2] A. C. MacPherson and D. M. Kerns, "A microwave microcalorimeter," *Rev. Sci. Instr.*, vol. 26, pp. 27-33; January, 1955.

[3] G. F. Engen, "A refined X-band microwave microcalorimeter," *J. Res. NBS*, vol. 63C, pp. 77-82; July-September, 1959.

[4] D. M. Kerns, "Determination of efficiency of microwave bolometer mounts from impedance data," *J. Res. NBS*, vol. 42, pp. 579-585; June, 1949.

[5] R. W. Beatty and F. Reggia, "An improved method of measuring efficiences of ultra-high-frequency and microwave bolometer mounts," *J. Res. NBS*, vol. 54, pp. 321-327; June, 1955.

[6] G. F. Engen, "A bolometer mount efficiency measurement technique," *J. Res. NBS*, vol. 65C, pp. 113-124; April-June, 1961.

ble in the frequency region where these dual element mounts are employed.

The substitution error, which is the subject of this paper, results from a different division of dc and RF powers between the two elements due to those circuit features which make possible the series-parallel operation. In the section to follow, an analytic treatment of this problem will be given. An intuitive understanding may be obtained from the following considerations.

In order for the measurement to be free of substitution errors it is *sufficient* that the bolometer resistance (or sum of bolometer resistances) be dependent only on the total power absorbed and independent of its distribution within (or among) the bolometer element(s). The realization of this condition for a pair of bolometers requires that either the "ohms-per-milliwatt" coefficients be identical and independent of power level, or at least that these characteristics complement one another in some predetermined manner. This is in addition to the usual criteria which must be satisfied by the elements individually.

Alternatively, the substitution error will also vanish provided that the dc and radio frequency power distributions are identical throughout the element(s). When a pair of elements are employed, however, it is possible to violate this criterion in a way which does not exist for a single element. This will become evident in the discussion which follows.

The frequency region where these techniques are usually employed also includes the transition zone where the use of circuit concepts must be either justified or replaced by the field equations. Although the resultant conclusions cannot be regarded as completely rigorous, a considerable amount of physical insight into the problem is provided by the application of low frequency circuit concepts to the bolometer mount.

The series connection for dc bias currents enforces the condition that the *currents* in the two elements be equal. If the resistance division is unequal, when energized with dc power only, the dc power division will be unequal with the *greater* power dissipated in the element of *larger* resistance.

On the other hand, the parallel connection for RF power requires that the RF *voltages* across the two elements be equal, and as a result the *larger* amount of RF power will be absorbed by the element of *smaller* resistance. The boundary conditions, which determine the power division ratio, are thus different for the two sources of power. For this reason the application of RF power is, in general, accompanied by a shift in resistance division. The dual element mount is thus subject to a source of substitution error which does not occur in single element mounts.

Theory

The basic equal-arm bridge circuit employed in bolometric measurements is illustrated in Fig. 2. The essential features include means for recognizing that the

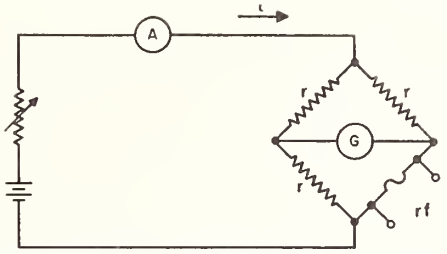


Fig. 2—Basic circuit of equal arm bolometer bridge.

bolometer is operating at the prescribed value of resistance and for determining the bias power absorbed. In the circuit shown this is given by $P=(r/4)i^2$. In practice the change or reduction in this bias power required to maintain bridge balance is taken as a measure of the applied RF power, that is:

$$P_{RF}' = \frac{r}{4}(i_1^2 - i_2^2) \tag{1}$$

where i_1 and i_2 are the balanced bridge currents without and with the RF power applied, and the prime is used to differentiate from the more exact expression which will be introduced later.

In order to describe the behavior of the dual element bolometer mount, where the resistances of the individual elements are subject to variation, it is necessary to introduce the bolometer resistance law. In a constant temperature environment this can be written in the approximate form:

$$P_2 = P_1 + \frac{1}{\gamma}(r_2 - r_1) \tag{2}$$

for each bolometer element, where P_2 and P_1 are the total (sum of dc and RF) powers required to bias the bolometer resistance at the values r_2 and r_1 respectively, and γ is the resistance-per-unit-power bolometer sensitivity. Eq. (2) includes the implicit assumption that the substitution procedure is satisfied by the bolometer elements individually, and that γ is contant or independent of power level.

Let P_2 include contributions from both RF and dc sources while P_1 includes only the latter. Then for each bolometer element:

$$P_{RF} = P_{1dc} - P_{2dc} + \frac{1}{\gamma}(r_2 - r_1)$$

$$= r_1 I_1^2 - r_2 I_2^2 + \frac{1}{\gamma}(r_2 - r_1) \tag{3}$$

where I_1 and I_2 are the bolometer currents. Note that when operated in a balanced bridge, $I_1=i_1/2$, etc.

Now if a pair of the elements (r_a and r_b) are connected in series as shown in Fig. 1, the total RF power will equal the sum of powers absorbed by each element:

$$P_{RF} = P_{aRF} + P_{bRF}$$
$$= r_{a1}I_{a1}^2 + r_{b1}I_{b1}^2 - r_{a2}I_{a2}^2 - r_{b2}I_{a2}^2$$
$$+ \frac{1}{\gamma_a}(r_{a2} - r_{a1}) + \frac{1}{\gamma_b}(r_{b2} - r_{b1})$$

where the subscripts a and b refer to bolometer elements a and b respectively. But $r_{a1}+r_{b1}=r_{a2}+r_{b2}=r$, if the total resistance is maintained at a constant value. In addition the series connection requires that $I_{a1}=I_{b1}=I_1 = i_1/2$, and $I_{a2}=I_{b2}=I_2=i_2/2$. Then,[7]

$$P_{RF} = \frac{r}{4}(i_1^2 - i_2^2) + \left(\frac{1}{\gamma_b} - \frac{1}{\gamma_a}\right)(r_{b2} - r_{b1}). \quad (4)$$

Comparison of (4) with (1) indicates that if the more complete theory is not taken into account, the measurement will be in error by the amount

$$\left(\frac{1}{\gamma_b} - \frac{1}{\gamma_a}\right)(r_{b2} - r_{b1}).$$

It will be noted that this term vanishes when $\gamma_b=\gamma_a$ or when $r_{b2}=r_{b1}$. The latter condition corresponds to the case where the resistance division between the two elements remains fixed.

Eq. (4) gives the error implicitly in terms of the change in resistance division $(r_{b2}-r_{b1})$ and the bolometer "constants" γ_a and γ_b. This is a useful expression in making an experimental determination of the magnitude of this error.

The shift in resistance division, however, is determined by the parameters of the bolometer mount, bolometer elements, and the applied power (dc and RF). Thus if a sufficiently detailed description of the bolometer mount were available, it would be possible to express the RF power as a function of these parameters and the dc power.

In the paragraphs which follow, an approximate expression is derived on the basis of a rather simple model for the bolometer mount. This provides a considerable amount of additional insight into the problem.

In the derivation which follows, the following approximations are used:

1) The bolometer mount is evaluated in terms of voltage and current concepts, and the equivalent circuit of Fig. 1 is used.
2) The series reactance of the bolometer elements and bypass condensers is ignored.
3) A linear relationship is assumed to exist between the bolometer resistance and power, *i.e.*, γ is assumed to be constant.
4) The dc-RF substitution process is assumed to be valid for the elements on an individual basis.

[Items 3 and 4 have been already introduced in conjunction with (2).]

These approximations, and the boundary conditions implicit in the technique, lead to the equations:

$$r_{a2} = r_{a0} + \gamma_a\left[I_2^2 r_{a2} + \frac{e^2}{r_{a2}}\right]$$
$$r_{b2} = r_{b0} + \gamma_b\left[I_2^2 r_{b2} + \frac{e^2}{r_{b2}}\right] \quad (5)$$
$$r = r_{a2} + r_{b2}$$

where the following definitions apply:

r_{a2}, r_{b2} = resistance of element a, b with RF applied
r_{a0}, r_{b0} = resistance of the element a, b in the absence of any power
γ_a, γ_b = resistance per unit power (ohms/milliwatt) coefficient of element a, b
e = RF voltage across element a, b
I = dc current through element a, b. (Note I takes the values I_1, I_2 without and with RF applied.)
r = (constant) sum of r_a and r_b.

In addition, the following definitions will be employed.

$P_b = r(I_1^2 - I_2^2)$ = power measured by conventional bolometric technique

$P_1 = rI_1^2$ = dc power required by bolometers for bridge balance in the absence of RF power

$P_{RF} = \dfrac{e^2 r}{r_{a1}r_{b2}}$ = "true" value of RF power absorbed by bolometers

$r_{a1} = \dfrac{r_{a0}}{1-\gamma_a I_1^2}$, $r_{b1} = \dfrac{r_{b0}}{1-\gamma_b I_1^2}$ = respective values of r_a, r_b when bridge is balanced by dc only.

From the immediately preceding system of eight equations it is possible to eliminate the variables r_{a2}, r_{b2}, r_{a0}, r_{b0}, e, I_1, and I_2. This yields the relationship between P_{RF}, P_b, P_1, r_{a1}, r_{b1}, γ_a, and γ_b:

$$P_{RF}^2 - P_b^2 + kP_{RF} - (k+\epsilon)P_b = 0 \quad (6)$$

where:

$$k = \frac{r_{a1}}{\gamma_a} + \frac{r_{b1}}{\gamma_b} - P_1 \quad (7)$$

$$\epsilon = \left(\frac{1}{\gamma_b} - \frac{1}{\gamma_a}\right)(r_{a1} - r_{b1}). \quad (8)$$

[7] Eq. (4) may also be written in the equivalent form:

$$P_{RF} = \frac{r}{4}(i_1^2 - i_2^2) + \left(\frac{1}{\gamma_b} - \frac{1}{\gamma_a}\right)(r_{a2} - r_{a1}),$$

since, $r_{a1}+r_{b1}=r_{a2}+r_{b2}=r$.

The locus of (6) is a rectangular hyperbola which passes through the origin. It can be shown that the hyperbola center lies in either the first or third quadrants.

For positive temperature coefficient elements (barretters) the center lies in the third quadrant as shown in Fig. 3. Although only that portion of the curve which lies in the first quadrant has any physical significance, the remainder is included as an aid in interpreting the result. It will be noted that the constant ϵ plays an important role in the geometry. In particular the *absolute* error is given by the difference between the hyperbola and the quadrant bisector. For large values of power this approaches the value $\epsilon/2$. The *fractional* error is determined by the slope of the straight line connecting the origin and operating point and is equal to the difference between this slope and unity. For small values of power this error is given by ϵ/k and decreases with increasing power.

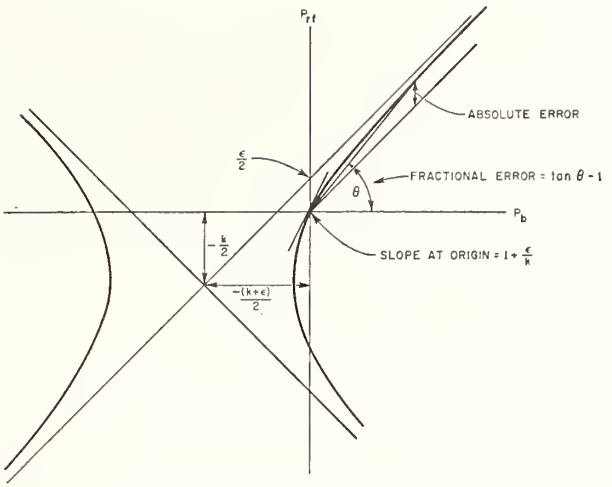


Fig. 3—Illustrating the behavior of positive temperature coefficient (barretter) elements.

For negative temperature coefficient elements (thermistors) the hyperbola center lies in the first quadrant as shown in Fig. 4. Here the fractional error is smallest for small values of power, and both the fractional and absolute errors increase with increasing power levels.

Within a limited neighborhood of the origin, this curve may be expected to give a reasonable approximation of the system behavior. For larger values of power a marked departure from the indicated operation may be expected because the $|\gamma_a|$ and $|\gamma_b|$, instead of being constant, decrease with increasing power. In order to predict the system behavior for large values of P_{RF}, a more complete description of the thermistor resistance law is required. This topic is beyond the scope of this paper. It is well known, however, that there is a limit to the RF voltage which may be applied if stable operating conditions are to be maintained.

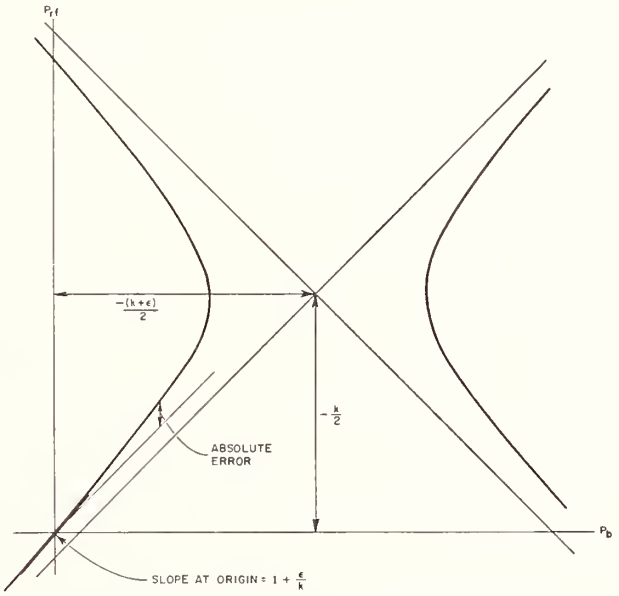


Fig. 4—Behavior of negative temperature coefficient (thermistor) elements.

Two points of interest are evident from this discussion. First, the error will vanish if either $\gamma_a=\gamma_b$ or $r_{a1}=r_{b1}$ [compare with (8)]. This first condition is also evident from (4), and may be explained by noting that if the sensitivities of the two elements are equal, the resistance division is unimportant. The second condition $r_{a1}=r_{b1}$ is explained by observing that when the initial resistances are equal, the power division ratios will also be equal for both sources of power and thus no shift in resistance division will occur. This result, if generalized to the case where circuit concepts are inapplicable and the series reactances non-negligible, leads to a symmetry requirement for the current paths associated with the two bolometer elements.

It is thus possible to minimize the error by matching the elements for both an equal (initial) division of resistance and equal sensitivities. The former tends to minimize the shift in resistance division, the latter its consequences. As a practical measure, the former condition is more easily recognized than the latter.

On the other hand, the $r_{a1}=r_{b1}$ condition is based on the circuit concept and negligible reactance approximations [1) and 2)]. This is in addition to approximations 3) and 4) which lead to the $\gamma_a=\gamma_b$ condition. In practice the largest deviations from the predicted behavior[8] are probably due to the failure to satisfy the negligible reactance approximation. Thus the initial equality of r_{a1} and r_{b1} does not always ensure little or no resistance shift. From an academic point of view, it is therefore more desirable to match the γ's. Unfortunately, as will be brought out later, this is much more difficult.

The second point of interest is that the relationship between P_{RF} and P_b is basically a nonlinear one. It follows that the fractional or percentage error will de-

[8] Excluding thermistor operation at large values of P_{RF}.

pend upon the power level. In this respect, this error differs from the substitution and mount efficiency errors for single element mounts, which, according to the available information, are either constant or negligible.[9] To be sure, the derivation of this result is based upon a number of simplifying approximations. Thus the predicted degree of nonlinearity is probably too large in some situations, too small in others. It cannot be anticipated, however, that a more exact description would significantly alter this basic conclusion.

Experimental Evaluation

If a sufficiently detailed description of the bolometer mount (including bolometers) were available, it would be theoretically possible to predict the magnitude of this error analytically. In practice an experimental evaluation based on (4) is more convenient.

Referring to (1) and (4), the fractional error E is given by:

$$E = \frac{P_{RF}' - P_{RF}}{P_{RF}} \approx \frac{\left(\frac{1}{\gamma_b} - \frac{1}{\gamma_a}\right)(r_{b1} - r_{b2})}{\frac{r}{4}(i_1^2 - i_2^2)} . \qquad (9)$$

The error thus depends upon γ_a, γ_b and the shift in resistance division $(r_{b2} - r_{b1})$ which occurs when the RF power is applied. In particular it should be noted that an observation of this resistance shift makes it unnecessary to account for the mechanism by which it occurs.

In order to measure the variations in resistance division, it is necessary to provide a dc path to the midpoint between the two bolometer elements since this is not normally found in commercially available dual-element mounts. As will be evident from Fig. 1, the coaxial center conductor may be used for this purpose provided that the blocking capacitor is eliminated. Because the ratio of RF current division between the two elements is, in theory, independent of this common series reactance, this modification should result in little or no change in the phenomenon of interest. This procedure was followed in the evaluation of three commercially available mounts. The series blocking condenser was eliminated and observations of the dc potential on the center conductor were then made possible by an appropriate bypass and blocking arrangement external to the mount.

The basic circuit for observing the resistance division is illustrated in Fig. 5. In operation the bridge current is maintained at the value required for bridge balance (G_1 indicates zero). The lower bridge arm includes a variable tap (potentiometer) as shown. If this tap is adjusted such that the galvanometer G_2 also indicates a null, this bridge arm will be divided in the ratio r_b/r_a. This circuit permits a determination of the resistance division with different combinations of dc and RF bias powers.

Although the γ's are given by the slope of the resistance vs power curve, an experimental evaluation of the term $(1/\gamma_b - 1/\gamma_a)$ proves to be more difficult. There are several practical problems associated with the direct approach. First, it is the *difference* between γ_b and γ_a which is important in the expression of interest. If these values are nominally equal, a high order of accuracy is required in the individual values if a reasonable degree of accuracy is to be maintained in the difference.

Second, the measurement of the resistance vs power curve, while straightforward in principle, is actually quite difficult. The resistance of a bolometer element is a function both of the power absorbed and the ambient temperature of its surroundings. Even if a constant temperature external heat sink is postulated, the application of a constant power will be accompanied by a drift in the resistance value until certain temperature gradients within the bolometer mount are established. When the power level is changed to a different value, this process repeats itself. This represents a serious problem because the available bolometer elements are not completely stable in their characteristics, that is, due to certain aging effects, the power required to maintain a given resistance value is subject to drift even with the assumed constant temperature heat sink. It is difficult to separate these phenomena in the experimental determination. Finally, element "b" is part of the environment of element "a" and the temperature of element "b" may well be reflected in the measured characteristics of element "a."

These problems are taken care of by an indirect measurement technique which utilizes the circuit illustrated in Fig. 6. Intuitively, it will be recognized that a simultaneous null in galvanometers G_1 and G_2 indicates a resistance division between the bolometers in the ratio required by the lower bridge arm. The two current sources permit the currents in r_a and r_b to take on different values as required. In practice these currents are provided by automatic or self-balancing bridges[10] (SBB) as shown in Fig. 7. Not indicated in the figure, but essential to the operation, are the feedback connections to the self-balancing bridges which replace the galvanometers shown in Fig. 6.

By use of this circuit it is possible to arbitrarily vary the resistance division between the two elements while keeping the sum of their resistances constant. The dc

[9] G. F. Engen, "Recent developments in the field of microwave power measurements at the National Bureau of Standards," IRE Trans. on Instrumentation, vol. I-7, pp. 304–306; December, 1958.

[10] G. F. Engen, "A self-balancing direct-current bridge for accurate bolometric power measurements," *J. Res. NBS*, vol. 59, pp. 101–105; August, 1957.

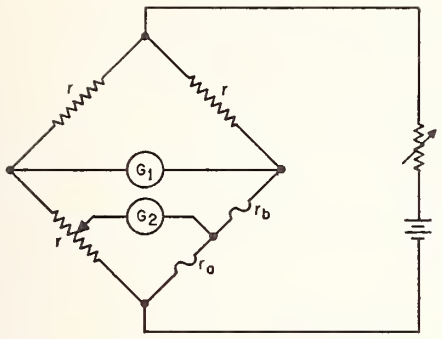


Fig. 5—Circuit for measuring shift in resistance division.

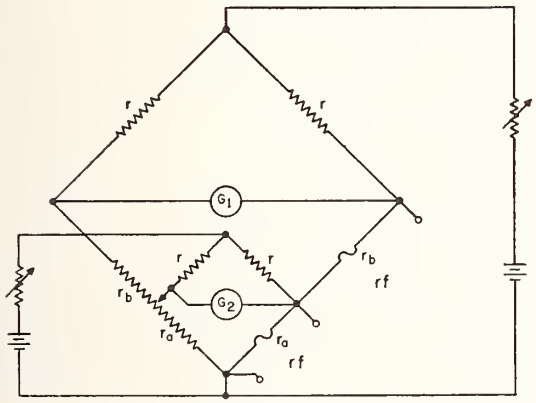
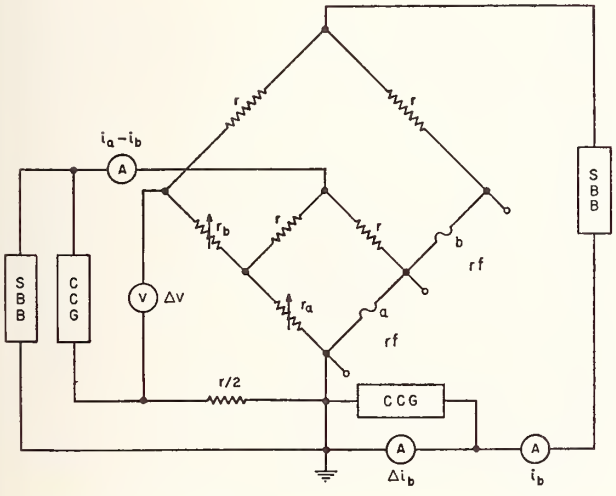


Fig. 6—Basic circuit for measuring error.



SSB-SELF BALANCING BRIDGE
CCG-CONSTANT CURRENT GENERATOR

Fig. 7—Detailed circuit for measuring error.

power dissipated in the bolometers may be determined from the appropriate ammeter readings.

The constant current generators[11] (CCG) provide a differential voltmeter and ammeter which are useful in measuring small changes in this power. These features permit a determination of the total dc bias power as a function of the resistance division ratio.

Eq. (4) may be written in the form:

$$P_{RF} = P_{1i} - P_{2i} - \left(\frac{1}{\gamma_b} - \frac{1}{\gamma_a}\right)(r_{b1} - r_{b2}) \qquad (10)$$

where P_{1i} and P_{2i} are the total dc powers required by the bolometers for bridge balance before and after the RF is applied. This equation, as written, is based on the implicit assumption that the powers P_{1i} and P_{2i} are measured subject to the condition that the dc currents in the two bolometer elements are equal. The circuit of Fig. 7 makes it possible to measure the RF power either in the "conventional" manner or with the resistance ratio fixed at an arbitrary value. In the latter case the power may be written:

$$P_{RF} = P_{1r} - P_{2r} \qquad (11)$$

where the subscript r indicates the values which obtain with the resistance ratio held constant. The term involving the resistance shift does not appear since these values are held constant.

Comparison of (10) and (11) indicates that the term

$$\left(\frac{1}{\gamma_b} - \frac{1}{\gamma_a}\right)(r_{b2} - r_{b1})$$

may be determined by measuring a given RF power by both methods and then taking the difference.

In practice this is done in the following way. First, by use of the circuit of Fig. 7, the bolometers are energized with dc only and the resistance ratio adjusted such that $i_{a1} = i_{b1}$. The power dissipated in the bolometers is now equal to P_{1i}. Next, an appropriate value of RF power is applied and the resistance ratio readjusted as required to yield the condition $i_{a2} = i_{b2}$. The dc power is now equal to P_{2i}. Since the resistance ratio in (11) was unspecified, it may now be set equal to that value which obtains with the RF present. In other words, let $P_{2r} = P_{2i}$. Finally the RF power is removed and the resistance values held constant. The dc power is now equal to P_{1r}. From (10) and (11) it is evident that

$$-\left(\frac{1}{\gamma_b} - \frac{1}{\gamma_a}\right)(r_{b1} - r_{b2}) = P_{1r} - P_{1i}. \qquad (12)$$

If the initial resistance division has been chosen such that $i_a = i_b$ and the constant current generators have been adjusted such that the associated differential current and voltage readings are initially zero, this change in power is given by:

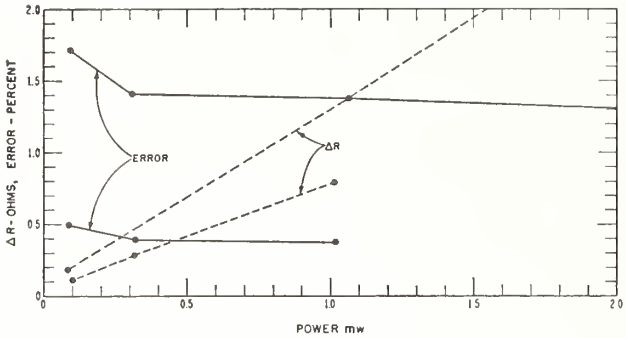


Fig. 8—Results of error measurements for barretter mounts.

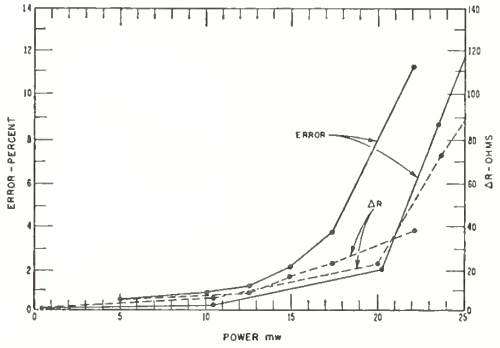


Fig. 9—Results of error measurement for thermistor mounts.

$$P_{1r} - P_{1i} = i_{b1}\Delta V + \tfrac{1}{4}\left[r_{a2}(\Delta i_a)^2 + r_{b2}(\Delta i_b)^2\right]. \quad (13)$$

The magnitude of this error has been evaluated for a small number of commercially available bolometers.

The results for barretter mounts are shown in Fig. 8, for thermistor mounts in Fig. 9. In each case, the worst behavior and typical behavior are indicated. The nonlinear nature is clearly evident in both illustrations.

For thermistor mounts it will be noted that the error is small over a considerable portion of the operating range but increases rapidly at the higher power levels. For the barretter mounts, the nonlinear behavior is less pronounced while the error, in the operating range of interest, is somewhat larger.

It is important to note that these results are based on only a small number of pairs of bolometer elements. In all, three different mounts were employed and approximately ten pairs of elements were measured. For this reason, the given results may not be necessarily representative of the existing art. The results do indicate, however, that the phenomena is large enough to warrant its evaluation in careful work.

Summary

A substitution error has been described which is peculiar to dual element bolometer mounts. This error occurs when the sum of the two bolometer resistances, but not their individual values, are constant.

The error is determined primarily by differences in the characteristics of the bolometer elements and, to some extent, by asymmetry in the bolometer mount. Thus, in most cases, the user is dependent upon the degree of quality control exercised by the manufacturer for any reduction in this difficult-to-detect error.

The rapid deterioration of thermistor mount performance, at the higher power levels, warrants more attention than it has received. In particuiar, it appears desirable to limit the RF dissipation in such mounts to a nominal 25 to 30 per cent of the total power. In many cases, this requirement is now being satisfied by the 10 mw upper power limit of the associated bridge circuitry.

Perhaps the most disturbing feature of this phenomenon is the nonlinear relationship which exists between the RF power as indicated by the bolometric method and its true value. This means that any calibration of the mount will be power level dependent. Although the problems in calibrating bolometer mounts are such that the technique is limited to a nominal one per cent in absolute accuracy, single element mounts have found use and evidently provide a much higher degree of accuracy in certain power *ratio*[11] applications.[7,12] The nonlinear characteristics of the dual element mounts generally render them unsuitable for this purpose.

Acknowledgment

The author acknowledges the assistance of F. R. Clague who performed the experimental evaluation and who checked most of the theoretical results. Thanks are extended to N. T. Larsen and R. W. Beatty for their helpful suggestions in the preparation of the manuscript.

[11] The type of "ratio" measurement alluded to here should not be confused with the type of measurement where the bolometer is provided with a constant dc bias current or voltage and the RF modulation envelope indicated by variations in the voltage or current. In the former case, the bolometer is operated in conjunction with a bridge and the RF power is frequently a substantial portion of the total bias power. As a rule this technique is employed with an unmodulated signal.

In the latter case, the RF power is limited to a small fraction of the total power and modulation of the signal is required. The implications of the described phenomena in an application of this type have not been investigated.

[12] G. F. Engen and R. W. Beatty, "Microwave attenuation measurements with accuracies from 0.0001 to 0.06 decibel over a range of 0.01 to 50 decibels," *J. Res. NBS*, vol. 64C, pp. 139–145; April–June, 1960.

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 70C, No. 2, April–June 1966

Coaxial Power Meter Calibration Using a Waveguide Standard*

Glenn F. Engen**

Institute for Basic Standards, National Bureau of Standards, Boulder, Colo.

(December 8, 1965)

The techniques associated with the calibration of one terminating type power meter in terms of a second terminating meter are useful both in calibration measurements and in the practical application of such devices. These techniques assume a variety of forms and represent an important segment of the microwave art. However their application to the calibration transfer problem between power meters with different input waveguides has long been inhibited by the requirement for an adaptor and the uncertainty which its losses can introduce into the procedure.

This paper describes a method of extending these existing techniques to this more general problem, in which the adaptor losses are only a second order effect. In addition, it provides limits for the error which is thus introduced.

Key Words: Power measurement, power calibration, adaptor efficiency.

1. Introduction

A large percentage of the power meters used at ultra-high and microwave frequencies are of the terminating type. This means that they (ideally) terminate the waveguide by its characteristic impedance and indicate the power which they absorb.

Given the problem of determining how much power is being delivered by a signal source to a particular load, it is common practice to substitute the terminating power meter for this load and thus measure the power it absorbs. Under ideal conditions this is also the power delivered to the load.

In practice, however, the impedance of the load and power meter will not be equal, and the ratio of the power delivered to the load and to the power meter will differ from unity. The determination of this ratio is of obvious importance in the practical application of terminating power meters. If the object of the measurement is that of calibrating one power meter in terms of another, this ratio determination is often called a "power calibration transfer." Today a variety of techniques are available for dealing with this problem.[1]

In practice, however, most if not all of these methods are limited to the case where the input waveguides to the two terminations (meter and load, etc.,) are of the same type or cross section. The more general problem of transferring power calibrations between power meters of rectangular waveguide and coaxial line inputs, for example, has received little attention. The reason for this lies in the implicit requirement for an adaptor and a detailed knowledge of its losses or other characteristics.

This paper will describe a method of effecting such a comparison in which the adaptor losses are only a second order effect and for which limits of error are given. Aside from the adaptors, the method requires little or no instrumentation beyond what is required to compare power meters having the same type of input. Its complexity may be judged by noting that it requires only the application of existing calibration transfer techniques. It does, however, call for two separate measurements which are then averaged to yield the final result. The procedure should prove a useful addition to existing measurement techniques.

2. General Description

As a specific example, the calibration of a coaxial bolometer mount in terms of a waveguide "standard" will be considered. The procedure for applying the method to similar problems will then be obvious.

The components to be considered explicitly include the "standard" or calibrated waveguide bolometer mount, the coaxial bolometer mount, and a waveguide to coaxial line adaptor of arbitrary characteristics.

A terminating type power meter may be calibrated either in terms of the "incident" power (power associated with the incident wave in a lossless line) or in terms of the *net* power absorbed (difference between "incident" and "reflected" powers). Although certain practical arguments can be given in favor of the "incident" power, the net power definitions are based

*A preliminary account of this work was given at the 1964 Conference on Precision Electromagnetic Measurements (Boulder, Colo.), at the IEEE International Convention, March 1965 (New York), and appears in Part 11 of the 1965 IEEE Convention Record (pp. 99–101). Because a different set of boundary conditions were employed, the error limits quoted in the Convention Record differ somewhat from those to be given here.

**Radio Standards Engineering Division, National Bureau of Standards, Boulder, Colo.

[1] For a brief survey of the existing techniques see the author's paper, A variable impedance power meter and adjustable reflection coefficient standard, J. Res. NBS **68C** (Engr. and Instr.), No. 1, 7–24 (Jan.–Mar 1964).

385-582 O - 70 - 4

on a more fundamental concept. In the case of a bolometer mount the parameter of interest is the *effective efficiency*, which by definition is the ratio of the bolometrically indicated value to the net power absorbed by the mount. The problem is thus one of measuring the *efficiency*[2] of the coaxial mount given the efficiency of the waveguide standard bolometer mount.

As already noted (see footnote 1), techniques exist for effecting such comparisons when the input terminals are alike. More specifically, the comparison procedure yields the ratio of powers actually delivered to the two mounts. This *ratio* is then multiplied by the power *ratio* which is observed by the bolometric technique and the *ratio* of mount efficiencies obtained. Finally if one of these efficiencies is known, the other may be determined.

The calibration procedure, which is the subject of this paper, requires two measurements m_1, m_2, of the efficiency ratios of the adaptor-bolometer mount combinations shown in figures 1 and 2. The actual measurement procedure is not specified but may consist of any technique, including those listed in reference 1, which provides these ratios. In figure 1 the adaptor is connected to the coaxial mount such that two similar waveguide terminals are available for the comparison procedure (m_1). In figure 2 the adaptor is connected to the waveguide standard mount, and comparison (m_2) effected at the coaxial terminals. In order to permit these connections both the waveguide and coaxial connectors must be of the "sexless" variety.

If the adaptor were lossless, the efficiency of the adaptor-bolometer mount combinations would equal that of the mounts alone, and either procedure would yield the desired efficiency ratio.

In practice, of course, the adaptor is not lossless. Thus the first measurement, m_1, yields

$$m_1 = \frac{\eta_1 \eta_c}{\eta_w}, \tag{2-1}$$

where η_c, η_w, and η_1 are respectively the efficiencies of the coaxial mount, standard (waveguide) mount, and adaptor. (Note that the adaptor efficiency is a function of the terminating load impedance and direction of power flow.) Equation (2–1) may be verified by noting that the product found in the numerator is the efficiency of the adaptor-coaxial mount combination. (This is one advantage of basing the efficiency definitions on net power.)

In a similar manner the second measurement, m_2, yields

$$m_2 = \frac{\eta_c}{\eta_2 \eta_w}, \tag{2-2}$$

where η_2 is the adaptor efficiency which obtains during the second measurement. In general $\eta_1 \neq \eta_2$.

It is convenient to regard the determination of the ratio η_c/η_w as the object of the measurement procedure. Since, by hypothesis, the efficiency of the standard, η_w, is known, the measurement of this ratio yields the coaxial mount efficiency, η_c.

Inspection of eqs (2–1) and (2–2) shows that the measurement results include the factors η_1 and η^{-1}, respectively. Then, because the efficiency [illegible] exceed unity, m_1 and m_2 yield a lower and upper bound to the desired ratio η_c/η_w.

The geometric mean of m_1 and m_2 yields

$$\sqrt{m_1 m_2} = \frac{\eta_c}{\eta_w} \sqrt{\frac{\eta_1}{\eta_2}}, \tag{2-3}$$

while the quotient

$$\frac{m_1}{m_2} = \eta_1 \eta_2 = \frac{1}{1+\epsilon}. \tag{2-4}$$

Equation (2–4) serves to define the parameter ϵ implicitly. This parameter is a measure of the adaptor losses and tends to zero as these losses are reduced.

[2] The term "efficiency" as employed in this paper is to be interpreted in a broad sense and may represent either "effective efficiency" of a bolometer mount, or adaptor "efficiency" (ratio of net power output to net power input).

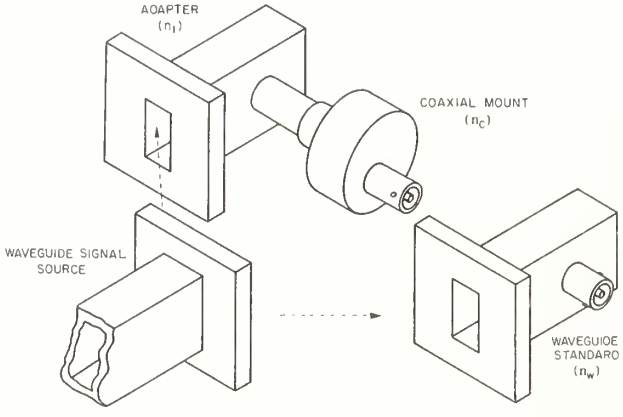


FIGURE 1. *Illustrating first step of measurement procedure.*

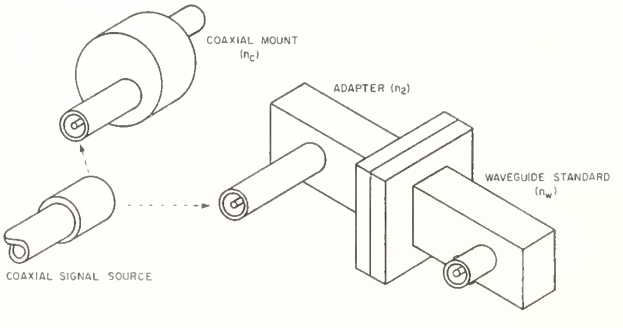


FIGURE 2. *Illustrating second step of measurement procedure.*

In summary, the measurement technique consists of making the measurements m_1, m_2, and taking the geometric mean as the desired ratio η_c/η_w. This includes the approximation $\sqrt{\eta_1/\eta_2}=1$. A knowledge of either η_c or η_w thus permits the determination of the other. As already noted, it is possible to establish limits for the error introduced by this approximation from the fact that the efficiencies cannot exceed unity. Much tighter limits of error, however, can be derived by utilizing the fact that the same adaptor is used in both measurements. More specifically, if reciprocity is satisfied, it is possible to obtain upper and lower limits to this approximation in terms of ϵ and the impedance conditions which prevail.

3. Limits of Error

As noted in the preceding section, the method is based on the approximation $\sqrt{\eta_1/\eta_2}=1$. The error, thus introduced, depends upon impedance conditions and adaptor losses. Three different modes of operation have been considered and are referred to as Cases I, II, III. For purposes of illustration it is convenient to visualize the procedure in terms of figure 3, where the adaptor and two bolometer mounts are connected together as shown.

The first measurement, m_1, consists of separating the assembly at terminal surface 1 and measuring the ratio of efficiencies of the waveguide mount to the adaptor-coaxial mount assembly. The second measurement, m_2, is analogous where terminal surface 2 instead of 1 is employed. Let Γ_w and Γ_c represent the reflection coefficients of the waveguide and coaxial mounts, and let Γ_1 be the reflection coefficient at terminal surface 1 of the adaptor when terminated by a load Γ_c (the coaxial mount) as shown in figure 3. Conversely, let Γ_2 represent the reflection coefficient which obtains at adaptor surface 2 with surface 1 terminated by Γ_w. Finally, let $|\Gamma_a|$ represent the magnitude of the adaptor reflection coefficient.[3]

[3] The adaptor reflection coefficient magnitude, $|\Gamma_a|$, is that value which obtains at one side or port of the adaptor with the other end connected to a matched (reflectionless) load. It thus corresponds to the "adaptor VSWR." In general the value of $|\Gamma_a|$ measured at terminal 1 differs from that at terminal 2. For high efficiency adaptors this difference is small and vanishes as the adaptor becomes lossless. Thus for most practical purposes the adaptor may be regarded as characterized by a single value of $|\Gamma_a|$.

The error expressions given for Case I are such as to give the correct limits if the value substituted for $|\Gamma_a|$ is the smaller of the two. Thus. if the larger value is used instead, somewhat wider limits will be obtained.

The failure to identify $|\Gamma_a|$ with either terminal is intentional in that this represents the most general case of practical significance. As will be shown later, tighter limits of error result if $|\Gamma_a|$ is identified with one terminal or the other.

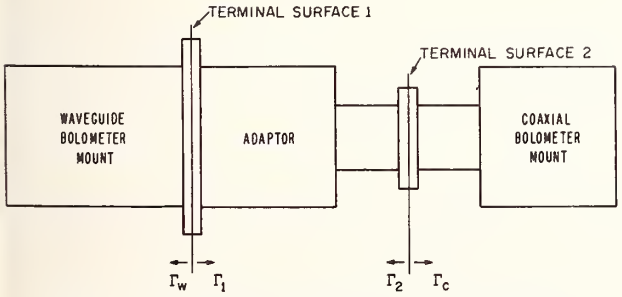


FIGURE 3. *Block diagram showing impedance relationships.*

The error, E, in the different modes of operation is based on the following definition:

$$E=\sqrt{\frac{\eta_1}{\eta_2}}-1. \tag{3-1}$$

Approximate (correct to the order given) limits for E are as follows:

Case I. The impedance conditions are assumed to be completely arbitrary. It will be shown that E lies within the following limits:

$$E_{\max}=\frac{\epsilon}{2}(|\Gamma_w|+|\Gamma_c|+|\Gamma_a|)+\frac{\epsilon^2}{8} \tag{3-2}$$

$$E_{\min}=-\frac{\epsilon}{2}(|\Gamma_w|+|\Gamma_c|+|\Gamma_a|)-\frac{\epsilon^2}{8}. \tag{3-3}$$

Case II. It is assumed that Γ_2 and Γ_c are equal (in both magnitude and phase) but unknown. This presupposes the incorporation and use of a tuning transformer in one of the components (usually the adaptor) to achieve this condition. It is then possible to express the limits for E in terms of $|\Gamma_1|$ and $|\Gamma_w|$ as follows:

$$E_{\max}=\frac{1}{2}(|\Gamma_1|+|\Gamma_w|)^2 \qquad \text{if } |\Gamma_1|+|\Gamma_w|\leq\frac{\epsilon}{2} \tag{3-4}$$

$$E_{\max}=\frac{\epsilon}{2}(|\Gamma_1|+|\Gamma_w|)-\frac{\epsilon^2}{8} \qquad \text{if } |\Gamma_1|+|\Gamma_w|\geq\frac{\epsilon}{2} \tag{3-5}$$

$$E_{\min}=-\frac{\epsilon}{2}(|\Gamma_1|+|\Gamma_w|)-\frac{\epsilon^2}{8}. \tag{3-6}$$

Case III. The reflection coefficients Γ_1 and Γ_w are assumed to be equal and of known magnitude. The limits for E now become

$$E_{\max}=2|\Gamma_w|^2 \qquad \text{if } |\Gamma_w|\leq\frac{\epsilon}{4} \tag{3-7}$$

$$E_{\max}=\epsilon|\Gamma_w|-\frac{\epsilon^2}{8} \qquad \text{if } |\Gamma_w|\geq\frac{\epsilon}{4} \tag{3-8}$$

$$E_{\min}=-\epsilon|\Gamma_w|-\frac{\epsilon^2}{8}. \tag{3-9}$$

The error limits for Case III may thus be obtained from Case II by letting $|\Gamma_1|=|\Gamma_w|$.

Although the errors associated with Cases II and III are somewhat smaller than in Case I, a more important argument in their favor is the simplification which they permit in the intercomparison measurements (m_1, m_2). Generally speaking, one of these measurements will be simplified if $\Gamma_2=\Gamma_c$ or if $\Gamma_1=\Gamma_w$.

In Case II, for example, the first step is to adjust the adaptor (transformer) such that $\Gamma_2 = \Gamma_c$. This permits a simplification in the measurement m_2. Although measurement m_1 must in general account for $\Gamma_1 \neq \Gamma_w$, this problem (along with the measurement of $|\Gamma_1|$ and $|\Gamma_w|$) is more easily handled in waveguide than in coaxial line (at least at higher frequencies!).

In Case III only one reflection coefficient magnitude $|\Gamma_w|$ $(=|\Gamma_1|)$ enters the error expressions and, all else being equal, gives the smallest error. The problem of making the calibration transfer between unequal impedances (m_2), however, has been shifted to the coaxial side where it is usually less convenient. An important application of Case III will be discussed in the following section.

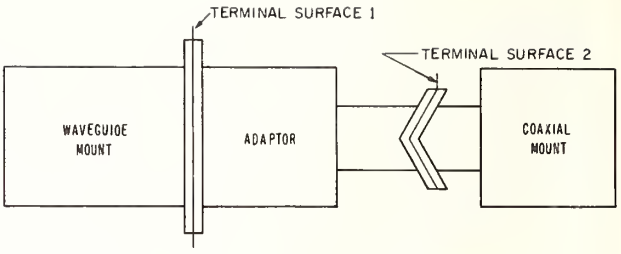


FIGURE 4. *Block diagram illustrating problem implicit in type* N *connectors.*

4. Extension to Type N Connectors

The foregoing techniques are based on the requirement that the connectors satisfy the "sexless" condition. Although a number of precision coaxial connectors are now available which meet this criterion, the extensively employed Type N does not. The technique may be extended to cover this case as follows.

It is now convenient to visualize the problem in terms of the assembly shown in figure 4. The first step of the measurement procedure is identical to that already described. However when the structure is separated at terminal surface 2, a problem is encountered in that it is not possible to mate both of these surfaces with a third one.

The solution requires the introduction of additional transitions and calibration transfer measurements as shown in figures 5a and 5b. The measurement "m_2" is thus replaced by a pair of measurements (m_{2a}, m_{2b}) which, by inspection, yield the ratios: $\eta_f\eta_c/\eta_2\eta_w$ and $\eta_c/\eta_m\eta_2\eta_w$, where η_f and η_m are the efficiencies of the additional transitions (adaptors) as shown. Since these efficiencies cannot exceed unity, it is possible to write

$$\frac{\eta_f\eta_c}{\eta_2\eta_w} = m_{2a} < m_2 < m_{2b} = \frac{\eta_c}{\eta_m\eta_2\eta_w}. \qquad (4\text{–}1)$$

The pair of measurements (m_{2a}, m_{2b}) thus yields upper and lower limits to m_2. If their arithmetic average is used in place of m_2, equation (2–3) becomes

$$\sqrt{m_1 \frac{(m_{2a}+m_{2b})}{2}} = \frac{\eta_c}{\eta_w}\sqrt{\frac{\eta_1}{\eta_2}}\sqrt{\frac{1}{2}\left(\eta_f + \frac{1}{\eta_m}\right)}. \qquad (4\text{–}2)$$

The error introduced by assuming $\sqrt{\eta_1/\eta_2} = 1$ has already been described. It is easily shown that the additional error in assuming $\sqrt{\frac{1}{2}(\eta_f + 1/\eta_m)} = 1$ is within the approximate limits $\pm\frac{1}{2}(m_{2b}-m_{2a})/(m_{2b}+m_{2a})$. As an illustration, it will be assumed that m_{2a} and m_{2b} differ by 1 percent. The average will then differ from m_2 by no more than ± 0.5 percent. Finally, this value is averaged with m_1. Since m_1 and m_2 are nominally equal, a ± 0.5 percent uncertainty in m_2 will give a ± 0.25 percent error in the final result.

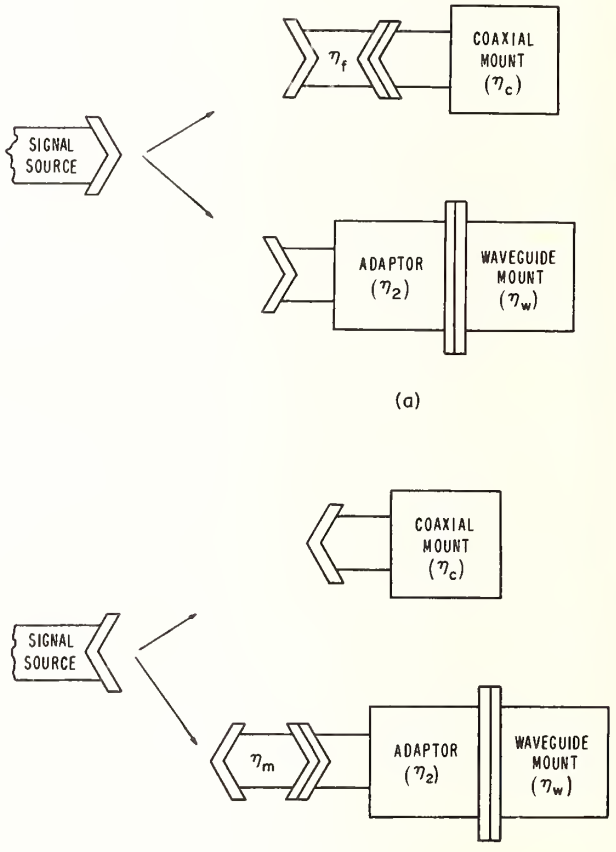


FIGURE 5. *Adaptor configuration for measurements* m_{2a} *and* m_{2b}.

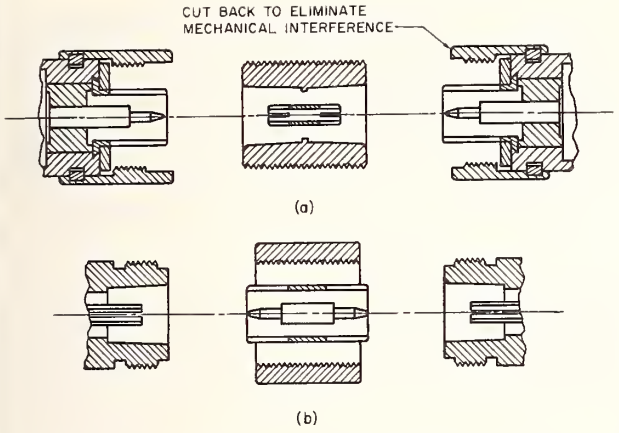


FIGURE 6. *Transition sections required in extending technique to type* N *connectors.*

Although these additional transitions can be made by commercially available components, it is desirable in practice to keep the associated losses as small as possible. Figures 6a and 6b show adaptors which have been built to satisfy this requirement. In particular it should be noted that the center conductor is supported only at its ends by the mating connectors.

The Type N connector suffers from a number of limitations including impedance discontinuities, which become increasingly important at high frequencies, and the lack of a well defined terminal plane or surface. As a consequence it is difficult to give a meaningful definition to impedance at the connector interface.

These considerations strongly suggest the use of a power calibration transfer procedure (measurement m_2) which is independent of the connector discontinuity. Such a technique has been described in a previous paper.[4] The complete procedure thus comes under Case III, where the impedances are matched and measured at the waveguide side (terminal 1), leaving the power calibration transfer between unequal impedances to be effected at terminal 2. (Note that according to the point of view adopted in the preceding paragraph, Case II cannot be applied because the impedance discontinuity makes it impossible to recognize when $\Gamma_c = \Gamma_2$.)

An alternative approach to using the Type N connector is based on the choice of reference plane indicated in figure 7. (This convention is called out in MIL Spec. C39012/1–5.)

If this point of view is adopted, it is desirable (in theory) to eliminate the shoulder in the outer conductor transition piece of figure 6a such that the two outside conductors are in physical contact.

The adaptor is thus, by definition, absorbed by the two Type N male connectors such that η_f may be given

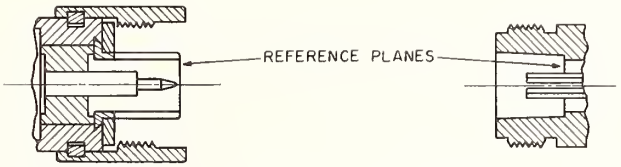


FIGURE 7. *Possible reference plane for type* N *connector.*

the value unity.[5] The measurement m_{2a} thus becomes m_2, and measurement m_{2b} is no longer required.

It is important to note that, according to this convention, the Type N male connector may be mated with either male or female, but this is not true of the Type N female. Thus this technique is limited to the case where the coaxial meter is fitted with the Type N male connector.

This convention also makes it possible to make the comparison measurements under Case II since the two assemblies shown in figure 5a may be adjusted (assuming tuning is available) for equal input impedances.

This alternative procedure is somewhat easier to implement, but more restricted in its application and interpretation.

5. Derivation of Error Expressions

A microwave measurements problem of long standing finds its solution in the foregoing techniques. However if these procedures are to be accepted by the scientific community, it appears desirable to record the derivation in sufficient detail to demonstrate its validity. Moreover, while the arguments are rather long, they have a potential application to related problems and are thus of some interest in their own right. The object of this section is the derivation of equations (3–2)–(3–9). (The reader who is willing to accept this "on faith" may, without loss of continuity, proceed to the next section.)

A complete description of the adaptor (at one frequency) requires six parameters—the real and imaginary components of its impedance matrix, for example. In addition, the adaptor efficiencies also depend upon the complex impedance of the terminating loads (bolometer mounts). Thus a total of 10 parameters is involved.

Obviously, if these 10 parameters were known, they would permit an exact determination of E. In practice, however, many of these parameters, especially those pertaining to the adaptor, do not lend themselves to ready measurement. Indeed, the value of this technique rests in a large measure upon its ability to provide limits for E with a minimum of supplementary information.

[4] G. F. Engen, A transfer instrument for the intercomparison of microwave power meters, IRE Trans. Instr. **I-9**, No. 2, 202–208 (Sept. 1960).

[5] This ignores any losses introduced by the failure to perfectly mate the outer conductors.

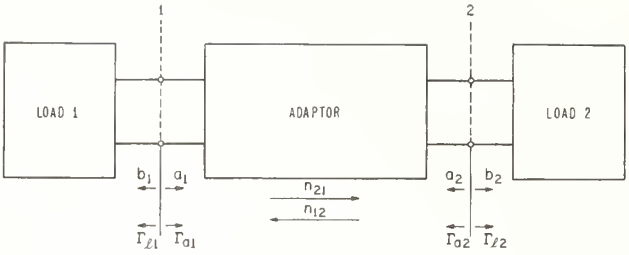


FIGURE 8. *Block diagram for error evaluation.*

Although the derivations are rather involved, it is easy to show that the error, E, is of second order. As already noted, the error vanishes in the ideal case of a lossless adaptor. The parameter, ϵ, is one measure or indication of the extent by which the adaptor fails to satisfy this condition. Thus the error tends to zero as ϵ goes to zero.

On the other hand, if an impedance match ($\Gamma=0$) is assumed for the different components, $\eta_1=\eta_2$ and again the error vanishes. This is also an idealization, but one which is approximately satisfied in practice. Since the error vanishes under either of these conditions, the error expressions may be expected to involve the product of ϵ and the Γ's.[6]

The problem may be formulated in a variety of ways; a convenient one is in terms of the normalized complex incident and emergent "voltage" wave amplitudes, a_1, a_2, b_1, b_2, as shown in figure 8 and the constants, α, β, γ, of the linear fractional transformation which relate the ratio b_1/a_1 to a_2/b_2. That is,

$$\frac{b_1}{a_1}=\frac{\alpha\dfrac{a_2}{b_2}+\beta}{\gamma\dfrac{a_2}{b_2}+1}. \tag{5-1}$$

In terms of the more familiar scattering matrix notation,

$$\begin{aligned}\alpha&=S_{12}^2-S_{11}S_{22}\\ \beta&=S_{11}\\ \gamma&=-S_{22}\\ \alpha-\beta\gamma&=S_{12}^2,\end{aligned} \tag{5-2}$$

and

$$\begin{aligned}b_1&=S_{11}a_1+S_{12}a_2\\ b_2&=S_{12}a_1+S_{22}a_2.\end{aligned} \tag{5-3}$$

It should be noted that reciprocity is assumed and impedance normalization made such that $S_{12}=S_{21}$. The complex constants α, β, γ thus provide a complete description of the adaptor.

The condition that the adaptor be source-free or passive imposes certain conditions on the parameters α, β, γ. Under the assumed normalization the scattering matrix S satisfies the matrix equation:[7]

$$\text{Det}\,(1-S^*S)\geq 0. \tag{5-4}$$

[In this equation (*) represents the Hermitian conjugate.]

For a two-arm junction this reduces to

$$(1-|S_{11}|^2-|S_{12}|^2)(1-|S_{12}|^2-|S_{22}|^2)$$
$$-|S_{11}S_{12}^*+S_{12}S_{22}^*|^2\geq 0, \tag{5-5}$$

and in terms of α, β, γ becomes

$$1-|\gamma|^2-|\beta|^2+|\alpha|^2-2|\alpha-\beta\gamma|\geq 0. \tag{5-6}$$

If eq (5–6) is multiplied by $1-|\gamma|^2$, the resulting relation can be expressed in the form

$$(1-|\gamma|^2-|\alpha-\beta\gamma|)^2-|\beta-\alpha\gamma^*|^2\geq 0. \tag{5-7}$$

It will also be shown in the next section [see eq (5–15)] that $1-|\gamma|^2-|\alpha-\beta\gamma|\geq 0$. Thus eq (5–7) may be written

$$1-\frac{|\alpha-\beta\gamma|}{1-|\gamma|^2}\geq\frac{|\beta-\alpha\gamma^*|}{1-|\gamma|^2}\geq 0. \tag{5-8}$$

Finally, it will prove convenient to make the following definitions:

$$1-\frac{|\alpha-\beta\gamma|}{1-|\gamma|^2}=x, \tag{5-9}$$

$$\frac{\beta-\alpha\gamma^*}{1-|\gamma|^2}=y. \tag{5-10}$$

Note that x is real and positive while y is complex. Equation (5–8) thus becomes

$$x\geq|y|\geq 0. \tag{5-11}$$

5.1. Analysis of a Special Case

Let η_{21} and η_{12} represent the adaptor efficiencies under the two conditions shown in figure 8, and let Γ_{l1}, Γ_{a1}, Γ_{a2}, Γ_{l2} represent the reflection coefficients which obtain at the different terminals as indicated. For example, $\Gamma_{l1}=a_1/b_1=1/\Gamma_{a1}$, etc. (This notation is somewhat more general than that employed in the earlier section and is introduced to simplify the application to other problems.)

[6] It will be noted that the error expressions also involve terms in ϵ^2. In the meaning of this paragraph a "matched adaptor" is matched at both ports or terminals while the error expressions assume the impedance is known at only one port.

[7] C. G. Montgomery, R. H. Dicke, and E. M. Purcell, Principles of Microwave Circuits (McGraw-Hill Book Co., Inc., New York, N.Y., 1948).

By definition,

$$\eta_{21}=\frac{\text{Net power delivered to load 2}}{\text{Net power input at terminal 1}}=\frac{|b_2|^2-|a_2|^2}{|a_1|^2-|b_1|^2}, \tag{5-12}$$

and similarly,

$$\eta_{12}=\frac{|b_1|^2-|a_1|^2}{|a_2|^2-|b_2|^2}. \tag{5-13}$$

The analysis of Case I is facilitated by considering first the special case where $\Gamma_{l1}=\Gamma_{l2}=0$. Applying these conditions (in turn) the adaptor efficiencies may be written in terms of α, β, γ as follows:

$$\eta_{21}=\frac{|\alpha-\beta\gamma|}{1-|\beta|^2}, \tag{5-14}$$

$$\eta_{12}=\frac{|\alpha-\beta\gamma|}{1-|\gamma|^2}. \tag{5-15}$$

The assertion made in connection with eq (5-8) is now evident, since $\eta_{12}\leq 1$.

It will prove instructive to consider the maximum and minimum values of η_{21} assuming η_{12} and $|\gamma|$ are given. From eqs (5-14) and (5-15),

$$\eta_{21}=\eta_{21}\frac{1-|\gamma|^2}{1-|\beta|^2}. \tag{5-16}$$

By inspection η_{21} increases in value when $|\beta|^2$ increases and conversely. Although the value of $|\beta|^2$ is not given, it is possible to determine limits for its value in terms of η_{12} and $|\gamma|$:

$$|\beta|^2\equiv\frac{|(\beta-\alpha\gamma^*)+(\alpha-\beta\gamma)\gamma^*|^2}{(1-|\gamma|^2)^2}$$

$$\leq\left[\frac{|\beta-\alpha\gamma^*|}{1-|\gamma|^2}+\frac{|\alpha-\beta\gamma|\cdot|\gamma|}{1-|\gamma|^2}\right]^2. \tag{5-17}$$

thus,

$$|\beta|^2\leq(|\gamma|+\eta_{12}|\gamma|)^2. \tag{5-18}$$

The maximum value of $|\beta|^2$ will occur for the maximum value of γ, that is when $|\gamma|=x=1-\eta_{12}$. Therefore,

$$|\beta|^2_{\max}=(1-\eta_{12}+\eta_{12}|\gamma|)^2. \tag{5-19}$$

Substitution in eq (5-16) yields

$$\eta_{21(\max)}=\frac{1+|\gamma|}{2-\eta_{12}(1-|\gamma|)}. \tag{5-20}$$

In order to determine the minimum value of $|\beta|^2$, it is convenient to write eq (5-17) in the form

$$|\beta|^2=\left|\gamma+\frac{(\alpha-\beta\gamma)\gamma^*}{1-|\gamma|^2}\right|^2. \tag{5-21}$$

Since the only restriction on γ is a limit to its magnitude, $|\beta|^2$ can be made to vanish by a suitable choice of γ for small values of $|\gamma|$. Comparison with eqs (5-8) and (5-15) shows that this is possible provided

$$|\gamma|\leq\frac{1-\eta_{12}}{\eta_{12}}. \tag{5-22}$$

For larger values of $|\gamma|$, the minimum value of $|\beta|^2$ is given by

$$|\beta|^2_{\min}=[\eta_{12}|\gamma|-(1-\eta_{12})]^2. \tag{5-23}$$

Substitution of these results into eq (5-16) yields

$$\eta_{21(\min)}=\eta_{12}(1-|\gamma|^2) \qquad \text{if } |\gamma|\leq\frac{1-\eta_{12}}{\eta_{12}}, \tag{5-24}$$

$$\eta_{21(\min)}=\frac{(1-|\gamma|)}{2-\eta_{12}(1+|\gamma|)} \qquad \text{if } |\gamma|\geq\frac{1-\eta_{12}}{\eta_{12}}. \tag{5-25}$$

Figure 9 shows a plot of eqs (5-20), (5-24), (5-25) with $|\gamma|$ equal to 0.2.

The area bounded by the $\eta_{21(\max)}$, $\eta_{21(\min)}$ curves, and the η_{21} axis represents possible combinations of η_{21} and η_{12} for the given choice of $|\gamma|$. In practice neither η_{21} nor η_{12} is known, but they are connected by the relation

$$\eta_{21}\eta_{12}=\frac{1}{1+\epsilon}. \tag{5-26}$$

This relationship is also plotted in figure 9 where $\epsilon=0.2$. (This value of ϵ is not representative, but chosen to better exhibit the characteristics of the problem.)

The final object of this section is to determine maximum and minimum values for the ratio η_{21}/η_{12} in terms of ϵ and $|\gamma|$. By inspection it is evident that this occurs at the intersection of eq (5-26) with the $\eta_{21(\max)}$ and $\eta_{21(\min)}$ loci.

The maximum value of η_{21}/η_{12} may be calculated from eqs (5-20) and (5-26) and is given by

$$\left.\frac{\eta_{21}}{\eta_{12}}\right|_{\max}=1+\frac{\epsilon|\gamma|+\frac{\epsilon^2}{4}(1+|\gamma|)^2}{1+\epsilon}. \tag{5-27}$$

In a similar way,

$$\left.\frac{\eta_{21}}{\eta_{12}}\right|_{\min}=1-|\gamma|^2 \qquad \text{if } |\gamma|\leq\frac{\epsilon}{2+\epsilon}, \tag{5-28}$$

$$\left.\frac{\eta_{21}}{\eta_{12}}\right|_{\min}=1-\frac{\epsilon|\gamma|-\frac{\epsilon^2}{4}(1-|\gamma|)^2}{1+\epsilon} \qquad \text{if } |\gamma|\geq\frac{\epsilon}{2+\epsilon}. \tag{5-29}$$

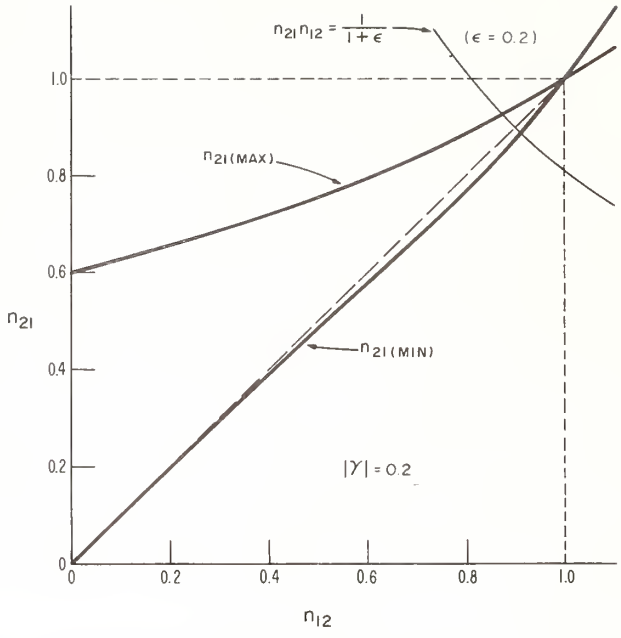


FIGURE 9. *Plot showing limits for* η_{21} *with* $|\gamma| = 0.2$.

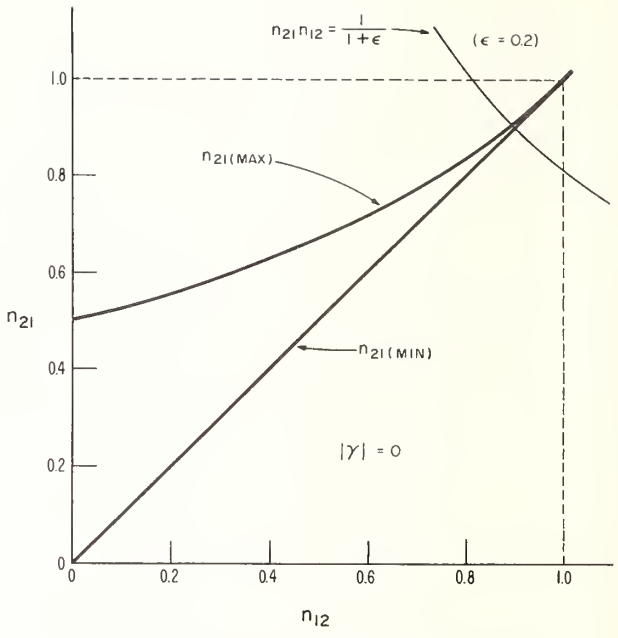


FIGURE 10. *Plot showing limits for* η_{21} *with* $\gamma = 0$.

These last three equations give the limits [8] for η_{21}/η_{12} in terms of ϵ and $|\gamma|$ subject to the condition that both loads (power meters) are matched ($\Gamma_{l1} = \Gamma_{l2} = 0$).

By definition, [eq (5–2)], $|\gamma|$ represents the reflection coefficient magnitude "looking into" adaptor terminal 2 with terminal 1 connected to a matched load. Figure 10 shows the reduction in the spread of possible values for η_{21}/η_{12} when $\gamma = 0$.

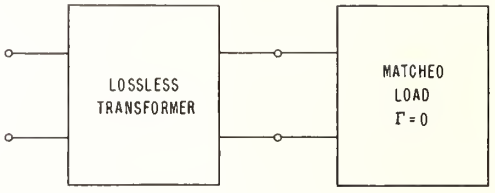


FIGURE 11. *"Equivalent circuit" of unmatched loads (power meters).*

5.2. Case I

In Case 1 each of the two loads (power meters) is assumed to have an arbitrary reflection of known magnitude.

The generalization of the foregoing results to Case 1 is simplified by use of the following artifice:

Returning to figure 8, it is possible to construct an equivalent circuit of loads 1 and 2 as shown in figure 11. The characteristics of the "lossless transformers" are adjusted to duplicate the impedances of the respective loads. Substitution into figure 8 results in the configuration shown in figure 12. It is now convenient to shift the terminal surfaces from the unprimed to the primed positions such that the lossless transformers become part of the adaptor. Because the efficiencies η_{21}, η_{12} are based on net power flow, their values are invariant to this shift in terminal surfaces. The results of the preceding section may thus be applied provided an appropriate change is made in $|\gamma|$ to account for the addition of the lossless transformers to the adaptor.

Let γ' represent the value γ takes when the terminal surfaces are shifted to the primed positions, and let Γ_{l1}, Γ_{l2} represent the reflection coefficients of the two loads: By definition γ' is equal to the ratio b_2'/a_2' which obtains at terminal 2′ assuming the matched load is removed and the assembly is excited at this port. The transformer bounded by terminals (2′ − 2) is thus terminated by a load of reflection coefficient $\Gamma_{a2} = b_2/a_2$, while terminal 1 of the adaptor is terminated by a load $\Gamma_{l1} = a_1/b_1$.

[8] It is of interest to compare this result with that which would be obtained if different adaptors were involved in the two measurements (as was done in extending the technique to Type N connectors). In the latter case the only condition which could be used is that the efficiencies do not exceed unity which leads to

$$\left.\frac{\eta_{21}}{\eta_{12}}\right|_{max} = 1+\epsilon, \quad \left.\frac{\eta_{21}}{\eta_{12}}\right|_{min} = \frac{1}{1+\epsilon}.$$

Although these expressions result in first order errors, they are of some interest in that they are completely independent of the impedance conditions.

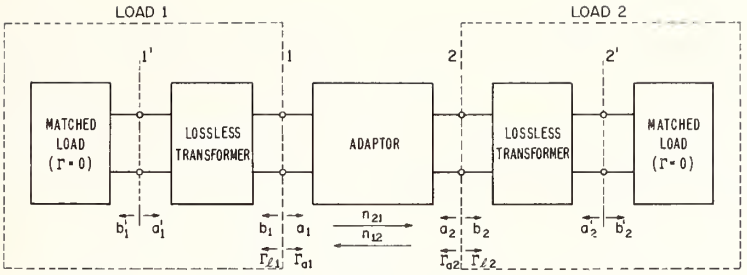


FIGURE 12. *Equivalent circuit used in extending analysis to unmatched loads.*

From eq (5–1),

$$\Gamma_{a2}=\frac{\alpha\Gamma_{l1}-\gamma}{-\beta\Gamma_{l1}+1}. \tag{5–30}$$

By hypothesis, the argument of Γ_{l1} is unknown (as well as the arguments of α, β, γ). The problem is now one of determining the maximum value of $|\Gamma_{a2}|$ as the argument of Γ_{l1} varies. For convenience the adaptor will be assumed lossless. The parameters α, β, γ then satisfy the conditions

$$\begin{aligned}\beta-\alpha\gamma^*&=0\\ |\alpha|&=1\end{aligned} \tag{5–31}$$

which follow from eq (5–8) when equality is assumed. These conditions may now be substituted back into eq (5–30) and the magnitude, $|\Gamma_{a2}|$, differentiated with respect to the argument of Γ_{l1}. This leads to [9]

$$|\Gamma_{a2}|_{\max}=\frac{|\gamma|+|\Gamma_{l1}|}{1+|\gamma\Gamma_{l1}|}. \tag{5–32}$$

By a straightforward extension of these arguments $|\gamma'|_{\max}$ is given by

$$\begin{aligned}|\gamma'|_{\max}&=\frac{|\Gamma_{a2}|_{\max}+|\Gamma_{l2}|}{1+|\Gamma_{l2}|\ |\Gamma_{a2}|_{\max}}\\ &=\frac{|\Gamma_{l1}|+|\Gamma_{l2}|+|\gamma|+|\Gamma_{l1}\Gamma_{l2}\gamma|}{1+|\Gamma_{l1}\Gamma_{l2}|+|\gamma\Gamma_{l1}|+|\gamma\Gamma_{l2}|}.\end{aligned} \tag{5–33}$$

As previously noted, the adaptor has been assumed lossless for this calculation. Because of losses, the actual value $|\gamma'|_{\max}$ will be somewhat smaller. In most practical applications the difference will be small, and in any case the use of eq (5–33) leads to error expressions which err on the side of being too large.

[9] Alternatively, eq (5–32) may be obtained by noting that $|\Gamma_{l1}|=$ constant represents a circle in the Γ_{l1} plane. The corresponding locus of Γ_{a2} is also a circle whose radius and displacement from the origin may be obtained in terms of α, β, γ, and $|\Gamma_{l1}|$. The maximum value of $|\Gamma_{a2}|$ is given by the sum of this radius and displacement.

Equation (5–33) is now substituted into eq (5–27), (5–28), (5–29), the square root taken, and only the lowest order terms retained. This results in the approximate expressions: [10]

$$\left(\frac{\eta_{21}}{\eta_{12}}\right)^{1/2}_{\max}=1+\frac{\epsilon}{2}(|\Gamma_{l1}|+|\Gamma_{l2}|+|\gamma|)+\frac{\epsilon^2}{8}, \tag{5–34}$$

$$\left(\frac{\eta_{21}}{\eta_{12}}\right)^{1/2}_{\min}=1-\frac{1}{2}(|\Gamma_{l1}|+|\Gamma_{l2}|+|\gamma|)^2$$

$$\text{if } |\Gamma_{l1}|+|\Gamma_{l2}|+|\gamma|\leqslant\frac{\epsilon}{2}, \tag{5–35}$$

$$\left(\frac{\eta_{21}}{\eta_{12}}\right)^{1/2}_{\min}=1-\frac{\epsilon}{2}(|\Gamma_{l1}|+|\Gamma_{l2}|+|\gamma|)+\frac{\epsilon^2}{8}$$

$$\text{if } |\Gamma_{l1}|+|\Gamma_{l2}|+|\gamma|\geqslant\frac{\epsilon}{2}. \tag{5–36}$$

5.3. Case II

In Case II the reflection coefficient magnitudes $|\Gamma_{l1}|$ and $|\Gamma_{a1}|$ are given. In addition, it is assumed that $\Gamma_{a2}=\Gamma_{l2}$. As in the previous problem, the analysis is facilitated by considering a special case: $\Gamma_{l1}=0$.

Subject to these conditions [eq (5–15)],

$$\eta_{12}=\frac{|\alpha-\beta\gamma|}{1-|\gamma|^2}$$

while

$$\eta_{21}=\frac{|\alpha-\beta\gamma|(1-|\Gamma_{l2}|^2)}{|1+\gamma\Gamma_{l2}|^2-|\beta+\alpha\Gamma_{l2}|^2}. \tag{5–37}$$

By hypothesis and by definition,

$$\Gamma_{l2}=\Gamma_{a2}=-\gamma. \tag{5–38}$$

[10] Aside from the change in notation, the generalization of eqs (5–34)–(5–36) to (3–2) and (3–3) involves the recognition that $|\gamma|$ is identified with terminal surface 2 while the counterpart $|\Gamma_a|$, as previously noted (footnote 3), is unspecified as to reference terminal. As a consequence, it is not possible to uniquely identify η_1 with either η_{21} or η_{12}. Both alternatives must be considered in order to determine the largest possible error.

The "deterioration" in error limit in going from eqs (5–34)–(5–36) to (3–2) and (3–3) is thus the "price" one pays for failure to identify the terminal surface associated with $|\Gamma_a|$.

Moreover,

$$\Gamma_{a1}=\frac{\alpha\Gamma_{l2}+\beta}{\gamma\Gamma_{l2}+1}=\frac{\beta-\alpha\gamma}{1-\gamma^2}. \tag{5-39}$$

By means of eqs (5-38) and (5-39) it is possible to eliminate β and Γ_{l2} from eqs (5-15) and (5-37), resulting in

$$\eta_{21}=\frac{|\alpha-\gamma\Gamma_{a1}|(1-|\gamma|^2)}{|1-\gamma^2|(1-|\Gamma_{a1}|^2)}, \tag{5-40}$$

$$\eta_{12}=\frac{|\alpha-\gamma\Gamma_{a1}|\ |1-\gamma^2|}{1-|\gamma|^2}. \tag{5-41}$$

In a similar way eq (5-8) becomes

$$1-\frac{|\alpha-\gamma\Gamma_{a1}|\ |1-\gamma^2|}{1-|\gamma|^2}$$

$$\geqslant\frac{|\alpha(\gamma-\gamma^*)+\Gamma_{a1}(1-\gamma^2)|}{1-|\gamma|^2}\geqslant 0. \tag{5-42}$$

Let

$$\frac{\alpha(\gamma-\gamma^*)+\Gamma_{a1}(1-\gamma^2)}{1-|\gamma|^2}=\delta. \tag{5-43}$$

Equation (5-42) now becomes

$$1-\eta_{12}\geqslant|\delta|\geqslant 0 \tag{5-44}$$

and eq (5-40) can be written:

$$\eta_{21}=\frac{\eta_{12}^2-|\delta-\Gamma_{a1}|^2}{\eta_{21}(1-|\Gamma_{a1}|^2)}. \tag{5-45}$$

In general the value of δ will be unknown. However, it is evident that η_{21} will have its maximum value if $\delta=\Gamma_{a1}$. This in turn is possible provided $|\Gamma_{a1}|\leqslant 1-\eta_{12}$, etc.

Therefore,

$$\eta_{21(\max)}=\frac{\eta_{12}}{1-|\Gamma_{a1}|^2}\qquad\text{if } |\Gamma_{a1}|\leqslant 1-\eta_{12}, \tag{5-46}$$

$$\eta_{21(\max)}=\frac{2\eta_{12}-(1-|\Gamma_{a1}|)}{\eta_{12}(1+|\Gamma_{a1}|)}\qquad\text{if } |\Gamma_{a1}|\geqslant 1-\eta_{12}, \tag{5-47}$$

$$\eta_{21(\min)}=\frac{2\eta_{12}-(1+|\Gamma_{a1}|)}{\eta_{12}(1-|\Gamma_{a1}|)}. \tag{5-48}$$

As in the previous problem, it is now possible to plot $\eta_{21(\max)}$ and $\eta_{21(\min)}$ in terms of η_{12}. It will then become apparent [as may also be shown by solving eqs (5-46), (5-47), and (5-48) for η_{12}] that this problem has been reduced to the previous one where the roles of η_{12} and η_{21} have been exchanged and $|\gamma|$ is replaced by $|\Gamma_{a1}|$. The same arguments for extending the result to the case where $\Gamma_{l1}\neq 0$ may also be used.

The expressions of interest may thus be obtained directly from eqs (5-34) and (5-35),

$$\left.\frac{\eta_{21}}{\eta_{12}}\right|_{\max}=\left.\frac{\eta_{12}}{\eta_{21}}\right|_{\min}\text{ etc.}$$

Therefore, the approximate limits for Case II are

$$\left(\frac{\eta_{21}}{\eta_{12}}\right)_{\max}^{1/2}=1+\frac{1}{2}(|\Gamma_{a1}|+|\Gamma_{l1}|)^2$$

$$\text{if } |\Gamma_{a1}|+|\Gamma_{l1}|\leqslant\frac{\epsilon}{2}, \tag{5-49}$$

$$\left(\frac{\eta_{21}}{\eta_{12}}\right)_{\max}^{1/2}=1+\frac{\epsilon}{2}(|\Gamma_{a1}|+|\Gamma_{l1}|)-\frac{\epsilon^2}{8}$$

$$\text{if } |\Gamma_{a1}|+|\Gamma_{l1}|\geqslant\frac{\epsilon}{2}, \tag{5-50}$$

$$\left(\frac{\eta_{21}}{\eta_{12}}\right)_{\min}^{1/2}=1-\frac{\epsilon}{2}(|\Gamma_{a1}|+|\Gamma_{l1}|)-\frac{\epsilon^2}{8}. \tag{5-51}$$

5.4. Case III

The boundary conditions on Case III are $\Gamma_{l1}=\Gamma_{a1}$ ($=\Gamma$ for brevity) and $|\Gamma|$ is given. The treatment of the previous problems has been simplified by first assuming one or both loads to be matched, but this approach does not lend itself to the present problem. The analysis of Case III proceeds in a different manner.

The efficiencies may be written:

$$\eta_{21}=\frac{|\alpha-\gamma\Gamma|^2-|\beta-\Gamma|^2}{|\alpha-\beta\gamma|(1-|\Gamma|^2)}, \tag{5-52}$$

$$\eta_{12}=\frac{|\alpha-\beta\gamma|(1-|\Gamma|^2)}{|1-\beta\Gamma|^2-|\gamma-\alpha\Gamma|^2}. \tag{5-53}$$

In terms of the previously defined parameters x, y [eqs (5-9), (5-10)] these become

$$\eta_{21}=\frac{(1-x)^2-g^2}{(1-x)(1-|\Gamma|^2)}, \tag{5-54}$$

$$\eta_{12}=\frac{(1-x)(1-|\Gamma|^2)}{h^2-(1-x)^2|\Gamma|^2}, \tag{5-55}$$

where

$$g=|y-\Gamma|, \tag{5-56}$$

$$h=|1-y\Gamma|. \tag{5-57}$$

Moreover

$$\eta_{21}\eta_{12}=\frac{(1-x)^2-g^2}{h^2-(1-x)^2|\Gamma|^2}=\frac{1}{1+\epsilon}. \tag{5-58}$$

These equations may be combined to yield the ratio η_{21}/η_{12} as a function of ϵ, Γ, and y.[11]

$$\frac{\eta_{21}}{\eta_{12}}=\frac{1+\epsilon}{(1+\epsilon+|\Gamma|^2)(1-|\Gamma|^2)^2}\cdot\frac{h^2\left(1-\frac{g^2}{h^2}|\Gamma|^2\right)^2}{1+\frac{g^2}{h^2}(1+\epsilon)} \tag{5-59}$$

The terms ϵ and $|\Gamma|$ are parameters of the problem while y is unknown and subject only to the restriction [eq (5-11)]

$$|y|\leq x.$$

The problem is thus one of determining y and arg Γ such that η_{21}/η_{12} has its maximum and minimum values.

Inspection of eq (5-59) shows that η_{21}/η_{12} decreases as h decreases and as g increases. Thus, the minimum value of η_{21}/η_{12} occurs when $g=x+|\Gamma|$ and $h=1-x|\Gamma|$.

If these relations are now substituted into eq (5-58), the resulting equation may be solved for x in terms of ϵ and Γ. This in turn determines g and h as functions of ϵ and Γ. Finally, substitution into eq (5-59) yields

$$\left.\frac{\eta_{21}}{\eta_{12}}\right|_{\min}=1-\frac{2\epsilon|\Gamma|}{\left(1+\frac{\epsilon}{2}\right)(1+|\Gamma|^2)+\epsilon|\Gamma|}$$

$$-\frac{\frac{\epsilon^2}{4}(1-|\Gamma|^2)^2}{\left[\left(1+\frac{\epsilon}{2}\right)(1+|\Gamma|^2)+\epsilon|\Gamma|\right]^2}, \tag{5-60}$$

or if only the lowest order terms are retained,

$$\left(\frac{\eta_{21}}{\eta_{12}}\right)^{1/2}_{\min}=1-\epsilon|\Gamma|-\frac{\epsilon^2}{8}. \tag{5-61}$$

Conversely, it is also evident from inspection that η_{21}/η_{12} increases as h increases and as g decreases. Thus, the arguments of y and Γ should be chosen such that

$$g=\Big|\,|y|-|\Gamma|\,\Big|, \tag{5-62}$$

and

$$h=1+|y\Gamma|. \tag{5-63}$$

In this case, however, $|y|$ is not determined since $g_{\min}$ occurs for $|y|=|\Gamma|$ while $h_{\max}$ obtains for $|y|=x$. It thus becomes necessary to substitute the above expressions for g and h into eq (5-59) and differentiate with respect to $|y|$. The derivative vanishes and a maximum occurs for

$$|y|=\frac{|\Gamma|(2+\epsilon)}{1+\epsilon-|\Gamma|^2}. \tag{5-64}$$

Substitution back into eq (5-59) leads to

$$\frac{\eta_{21}}{\eta_{12\,\max}}=\frac{(1+|\Gamma|^2)^2}{(1-|\Gamma|^2)^2}. \tag{5-65}$$

This however is subject to the condition that

$$|y|=\frac{|\Gamma|(2+\epsilon)}{1+\epsilon-|\Gamma|^2}\leq x. \tag{5-66}$$

The associated value of x may now be computed by substituting eqs (5-62), (5-63), and (5-64) back into eq (5-58). This leads to the condition that the solution given by equation (5-65) is valid in the range

$$\frac{|\Gamma|(2+\epsilon)}{1+\epsilon-|\Gamma|^2}\leq 1-\frac{(1+|\Gamma|^2)(1+\epsilon)^{1/2}}{(1+\epsilon-|\Gamma|^2)}, \tag{5-67}$$

which may be solved to yield

$$|\Gamma|\leq\frac{1}{\epsilon}[2+\epsilon-2(1+\epsilon)^{1/2}]=\frac{\epsilon}{4}-\frac{\epsilon^2}{8}\cdots \tag{5-68}$$

For larger values of $|\Gamma|$, $|y|$ is set equal to x and the problem handled as was done for $\eta_{21}/\eta_{12}|_{\min}$. Again, if only the lowest order terms are retained, $(\eta_{21}/\eta_{12})^{1/2}_{\max}$ is given by

$$\left(\frac{\eta_{21}}{\eta_{12}}\right)^{1/2}_{\max}=1+2|\Gamma|^2 \qquad \text{if } |\Gamma|\leq\frac{\epsilon}{4}, \tag{5-69}$$

$$\left(\frac{\eta_{21}}{\eta_{12}}\right)^{1/2}_{\max}=1+\epsilon|\Gamma|-\frac{\epsilon^2}{8} \qquad \text{if } |\Gamma|\geq\frac{\epsilon}{4}. \tag{5-70}$$

With an appropriate change in notation, the results of this section leads to eqs (3-2) –(3-9).

6. Experimental Results

An application of the foregoing techniques which is of immediate interest is the extension of the existing

[11] Note that g and h are merely abbreviations for certain functions of y and Γ.

NBS calibration capability in X-band waveguide to coaxial thermistor and barretter mounts. A variety of these items, with an advertised upper frequency limit in the 10–12 GHz range, is commercially available. A calibration near this upper limit is of particular interest because it appears reasonable to anticipate decreasing efficiency with increasing frequency. To the extent that this is true, the X-band calibration provides a lower limit to the efficiency over the entire operating range.

Measurements on a group of four mounts from different manufacturers (with Type N connectors) yielded efficiency values in the range 86–97 percent at 9 GHz. Another series of measurements on four different mounts but of the same make and model yielded values in the range 94–96.5 percent. These results indicate that high efficiency values are possible at X-band frequencies in coaxial mounts but also suggest there may be a much wider variation in different makes than is found in waveguide mounts.

The "waveguide-coax adaptor" used in the measurement was a eommercial adaptor connected to a five-stub waveguide tuning transformer. The combination provided an average efficiency of approximately 98 percent.

Because of the Type N connectors, it was necessary to employ the procedures described in section 4. The two measurements m_{2a}, m_{2b} typically differed by 0.4 percent, thus the error limit from this source was ± 0.1 percent. The error introduced by assuming $\sqrt{\eta_{21}/\eta_{12}}=1$ was computed from eqs (3–7)–(3–9) and did not exceed ± 0.1 percent. (The waveguide standard was matched with $|\Gamma_w| < 0.01$.) These errors are in addition to those introduced by the calibration transfer procedure itself. Although this latter error can be held to a few tenths of a percent or so when the transfer is between waveguide mounts, the performance of certain of the coaxial components (sliding short, connector repeatability, etc.) is not up to that of the waveguide counterparts, and this calls for a wider estimate of the error limits. A tabulation of the errors in a typical calibration is thus as follows:

1. Uncertainty in efficiency of waveguide standard	0.2%
2. Calibration transfer procedure	0.6–1.0%
3. $\sqrt{\eta_{21}/\eta_{12}}=1$ approximation	0.1%
4. Difference in m_{2a} and m_{2b}	0.1%
Total	1.0–1.4%

7. Other Measurements

Although attention has been focused primarily upon a specific application, the developed techniques have potential use in many other measurement problems where a change in waveguide is involved.

For example, the measurement of the adaptor efficiency may be the prime objective. Equation (2–4) may be combined with (3–1) to yield

$$\eta_1=\sqrt{\eta_1\eta_2}\,(1+E)=\sqrt{\frac{m_1}{m_2}}\,(1+E). \qquad (7\text{–}1)$$

The square root of the quotient of the two measurements thus yields the efficiency η_1 within the limits given for E. A similar expression may be obtained for η_2.

The author expresses his appreciation to several of his colleagues who provided experimental demonstrations of the technique, checked the mathematics, and provided constructive suggestions in the preparation of the manuscript. These include Mrs. Anne Rumfelt, Fred R. Clague, John W. Adams, and Robert W. Beatty.

(Paper 70C2–223)

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 71C, No. 2, April–June 1967

A Dual-Load Flow Calorimeter for RF Power Measurement to 4 GHz

M. L. Crawford* and P. A. Hudson*

Institute for Basic Standards, National Bureau of Standards, Boulder, Colo. 80301

(December 6, 1966)

A new dual-load flow coaxial calorimeter power meter has been constructed at the National Bureau of Standards, Boulder Laboratories. Designed for use as a reference standard, the frequency range of the calorimeter extends up to 4 GHz and beyond. The power range extends from 2 W to 100 W with an error limit of 0.38 percent.

Design details, error analysis, and results of intercomparison with other standards are given.

Key Words: Coaxial, flow calorimeter, radio frequency power.

1. Introduction

The increasing complexity and higher performance characteristics of radio frequency and other electronic equipment in recent years has resulted in the need for more accurate measurement of the rf quantities. In rf power measurements, for example, 1 percent uncertainty for measurements in industrial standards laboratories is often required. Formerly, uncertainties of 5 percent were tolerable. Because uncertainties are accumulated in a chain of calibrations, the uncertainties in reference standards maintained by the National Bureau of Standards must be less than 1 percent.

The dual-load flow calorimeter described here is essentially a refinement of earlier calorimeters of this type [1][1] and was developed to meet the need for greater accuracy, and extend the range of NBS reference power standards up to 100 W. This development made possible the intercomparison between this standard and the NBS dry-load calorimeter [2] (50 mW to 5 W), thereby increasing the confidence in each.

2. Theory of Operation

The calorimetric principle has been considered the most accurate method for the measurement of rf power. This principle is based upon the first Law of Thermodynamics, or the conservation of energy. The measurement of electrical power using this principle depends upon the complete conversion of the electrical energy as delivered by a generator into thermal energy in a resistive load. The heat generated in the load results in a temperature rise in the load and its surroundings. This temperature rise is a monotonic function of the input power level and may be detected, for example, with a thermopile located between the load and a reference body whose temperature is stable with time.

Two types of calorimeters could be constructed to measure power at the levels of interest. They are the absolute flow calorimeter and the substitution flow calorimeter. (The use of dry-load calorimeters at these power levels, 5–100 W, is not practical because of the large physical size of loads required and the long measurement time constant.) An example of a "true" or "absolute" calorimetric system is shown in figure 1. In this system, power is measured in terms of mass, time, and temperature by the equation $P=Fc\Delta T$. In this equation F is the mass flow rate of the calorimeter fluid, c is its specific heat, and ΔT is the equilibrium temperature rise of the fluid. A conservative measure of the uncertainty with which such a measurement can be made is the sum of the uncertainties with which F, c, and ΔT can be determined. The value of c is known very accurately for the common fluids. The value of ΔT, however, is difficult to determine accurately for low input power levels, and F can be measured with an uncertainty no better than 0.5 percent at usual values of flow rate. In addition great care must be taken to prevent heat exchange with surroundings, and a correction must be made for heating due to friction flow of the fluid. Therefore, the uncertainty in power measurement using an absolute flow calorimeter is usually 1 percent or more.

Many of the above difficulties can be overcome by use of the dual-load calorimeter employing d-c substi-

*High Frequency Electrical Standards Laboratory, National Bureau of Standards Laboratories, Boulder, Colo. 80301.

[1] Figures in brackets indicate the literature references on page 116.

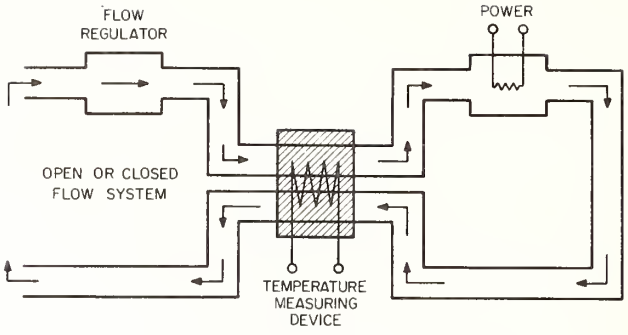


FIGURE 1. *Flow type "absolute" calorimeter.*

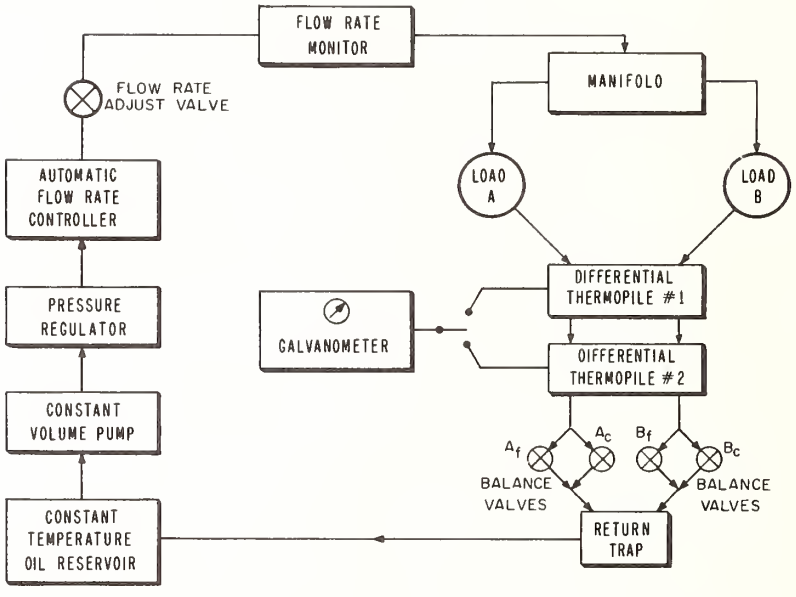


FIGURE 2. *Block diagram of dual-load flow calorimeter power meter 2–100 W, dc to 4 GHz.*

tution. A block diagram showing the system is given in figure 2. As the name implies, the dual-load calorimeter consists of two nearly identical loads connected in series or parallel to the fluid flow supply. Radio frequency power is applied to one load while d-c power is applied to the other. A differential thermopile, or other temperature sensing device, detects the temperature difference between the two streams on the downstream side of the loads when the system has reached thermal equilibrium. The d-c power level is adjusted to make the temperature difference zero, and the rf power is equated to the d-c power. Prior to the above measurement, it is necessary that the system be balanced by applying equal d-c power to each load at or near the rf power level to be measured. If the loads are in parallel to the fluid flow, then the flow rate through either one or the other may be adjusted for a null at the differential thermopile output.

Thus, in the dual-load substitution calorimetric technique, accurate knowledge of flow rate and temperature is not required for accurate measurement results. In addition heat exchange with surroundings due to external sources is not a problem as long as it remains constant during the measurement period. The substitution principle is based upon the assumption that heat generated by d-c or low frequency power absorbed by a load will have the same effect on the device used to sense the temperature rise of the load or a fluid surrounding it as heat generated by an equal amount of rf power. However, in general, equal quantities of rf and d-c power will not produce exactly the same calorimeter sensor response. Thus, an error, commonly known as the rf-dc substitution error, may exist. This error arises primarily because the rf current distribution in the load resistor is not identical to the d-c current distribution. This results in a difference

in the distribution of heat sources in the load in the two cases. In flow calorimeters, this error is minimized since the load resistor and mount structure are in intimate contact with a moving fluid, and the heat generated by the load tends to be transferred to the liquid regardless of the distribution of heat sources.

The main problems associated with measuring power by the substitution calorimetric method are the relatively long measurement time constants and the difficulty in maintaining thermal environmental and flow stability over this time period. The dual-load flow calorimeter was designed with emphasis on reducing some of these problems. For example, by reducing the time constant of the system to a minimum, the need for extremely stable rf sources and temperature control devices was reduced.

3. Description of Calorimeter and Operating Procedure

3.1. Loads

The loads (see fig. 3) were designed for rapid transfer of heat and low VSWR (≤ 1.03 referenced to 50 Ω at frequencies up to 4 GHz). The load resistor is a thin metal film deposited by vacuum evaporation onto a truncated conical substrate. The load is mounted inside a cylindrical outer conductor. This type of design was proposed by D. Woods [3]. The sheet resistivity, ρ, of the film is uniform over the area of the cone and the resistance is given by

$$R=\frac{\rho}{2\pi \sin\theta}\ln b/a \qquad (1)$$

where θ is the cone semiangle, and b/a is the ratio of outer to inner diameter of the coaxial line. The temperature coefficient of the resistive film is less than 10 ppm/°C thus insuring a nearly constant value of R at all power levels. The characteristic impedance of the section of coaxial line containing the conical resistor is

$$Z_0=\frac{Z_m}{2\pi}\ln b/a \qquad (2)$$

where $Z_m=\sqrt{\mu/\epsilon}$ is the wave impedance of the medium. By making $\rho/\sin\theta$ equal to Z_m, $Z_0=R$ at any point along the length of the load. This condition tends to assure the same current distribution along the resistor for both rf and dc. Measurements using a time domain reflectometer indicated that small discontinuities of the order of 0.1 Ω did exist along the length of the load. Thus, it appeared probable that a small substitution error could exist. The upper limit of the error was evaluated and is discussed in the following section.

The resistors are capable of absorbing up to 100 W of power with no significant change in their impedance characteristics. The physical dimensions of the loads were made small to reduce the measurement time. In order to absorb 100 W without damage to the film, the oil had to be circulated rapidly around and through the resistor body. At low levels of input power (2–5 W) the minimum oil flow rate was such that the temperature rise of the oil was approximately 5 deg. This flow rate per watt was about $\frac{15\ \text{cm}^3/\text{min}}{\text{watt}}$. At higher levels of input power, the flow rate per watt was reduced to about $\frac{7.5\ \text{cm}^3/\text{min}}{\text{watt}}$ so that the temperature rise was 10 deg centigrade.

3.2. Temperature Detection System

This system consists of two differential thermopiles (T.P. #1 and T.P. #2), of 10 junctions each, and a sensitive galvanometer. A selector switch allows either thermopile to be connected separately or both in series, to the galvanometer. The junctions of the thermopiles are located in the path of the moving oil as shown in figure 2. The oil is mixed thoroughly in a chamber downstream from the loads and then

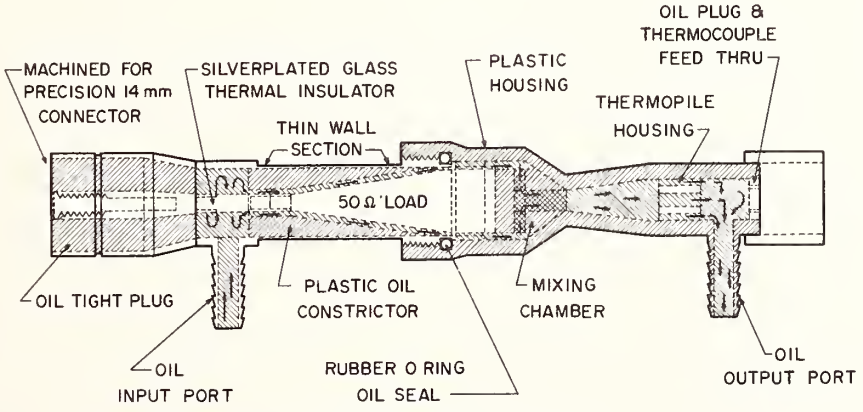


FIGURE 3. *Oil flow calorimeter load.*

passes around and through a plastic block in which the junctions of T.P. #1 are located. The close proximity of the thermojunctions of T.P. #1 to the load resistors results in rapid temperature detection and reduces the possibility of heat leakage before detection. Thermopile #2 was installed as a cross check against thermopile #1. Its junctions are mounted symmetrically inside plastic holders located farther downstream from T.P. #1. Any differences in the temperature of the oil flowing in channel A as compared to channel B will result in an emf generated by both T.P. #1 and T.P. #2. Either one or the sum of these emf's is detected by the galvanometer and within the limits of ± 0.02 percent, it is possible to detect any balance in the calorimeter system.

3.3. Flow System and Reservoir Temperature Control

The flow system of the dual-load flow calorimeter employs precision needle valves, a flow controller and meter, pressure regulator, constant volume pump, and reservoir, all incorporated into a closed circulation system (see fig. 2). Oil pumped from the reservoir passes through the pressure regulator which is set at 15 psi. It is then filtered and flows on to the flow rate meter and controller which is adjusted for the desired flow rate. Following the flow meter, the oil flow is divided into two channels, A and B, by means of the manifold. The flow path through each load is indicated by the arrows in figure 3. Oil leaving the loads passes through the mixing chamber, across thermopile #1, and across thermopile #2, on through the balance valves, to the return trap, and thence to the reservoir. The flow division between the two channels is regulated by means of the balance valves $A_F - A_C$ and $B_F - B_C$. These valves were placed downstream from the loads because this arrangement causes a back-pressure which resulted in better stability and control and insured that the load bodies were completely filled with oil at all times. Each valve has an adjustment range of 350:1 and by properly setting the ratio of flow between the fine and course valves, very small adjustments can be made in the flow division between channels.

The temperature of the oil in the reservoir is controlled at 28 ± 0.02 °C by a conventional automatic control circuit-cooling coil combination. This provides a nearly infinite heat sink for the system and helps to reduce the measurement time constant.

4. Estimation of Uncertainties

Uncertainties in measurement of power with the dual-load flow calorimeter were minimized by careful design and precision flow control and adjustment. The uncertainty of measurement is the difference between the true value and the measured value. This difference is usually expressed as a percentage of the true value. Sources of significant error and a brief explanation of each are given below.

4.1. Flow Division Instability and Thermal Effects

The instability in the flow division through each channel and the thermal drift are both reflected as a null shift or unbalance in the outputs of the thermopiles between loads A and B. As mentioned earlier, the effect of heat exchange with surroundings is much reduced in the dual-load configuration as compared to the single-load "absolute" calorimeter. Efforts were, nevertheless, made to minimize and equalize heat exchange between the two loads and their surroundings. Unequal heat exchange is not critical because the effect can be cancelled by proper adjustment of the flow rate through the individual loads during the d-c balancing operation. It is required, however, that the heat exchange be constant during the time a measurement is being made. After initial null (zero temperature difference between the oil leaving loads A and B) was achieved, the emf output of the thermopiles was recorded over a time period much greater than the measurement time constant which was approximately 5 min. Tests at several power levels indicated that for periods of up to 30 min the maximum drift was no greater than 6 μV. Since an oil temperature rise of 10 deg centrigrade produces a net thermopile output of 4000 μV, a maximum emf unbalance of ± 6 μV, corresponds to 0.15 percent shift in the thermal balance between the loads.

4.2. Detection System

The degree to which the two channels are balanced is a function of the overall resolution of the system. The sensitivity of the complete system is limited by the temperature fluctuations in the oil and not by the detection system which consists of the galvanometer and differential thermopile. These fluctuations limited the resolution to 0.02 percent.

4.3. RF–D-C Substitution Error

As mentioned earlier, the calorimeter loads and temperature detecting system were designed to minimize the rf–d-c substitution error. The loads were designed to provide matched terminations and thus insure, as nearly as possible, identical rf and d-c current distributions. In addition, the oil flowing over the resistor surfaces tends to absorb all the heat generated in the load regardless of the distribution of heat sources in the load. A small portion of heat may be conducted away at the points where electrical connection is made between the load resistor and the mount. Negligible heat conduction to the mount occurs at the input end where the center conductor is immersed in the moving oil stream for a distance of 1½ cm. At the grounded end of the resistor, where it contacts the outer conductor, some heat conduction is possible. In order to check for a temperature difference in this area a 10-junction thermopile was cemented to the outer conductor of one of the loads and referenced to an ice bath. "Equal" levels of d-c or rf power were then applied alternately to the load.

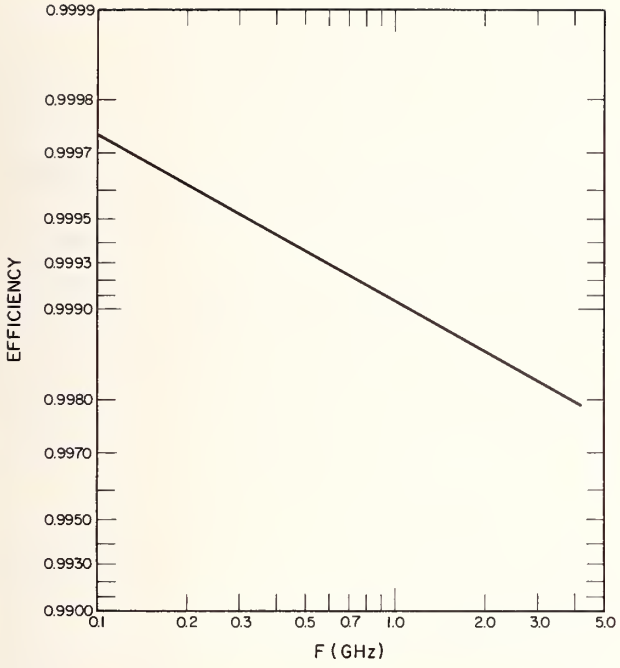


FIGURE 4. *DLFC input line efficiency versus frequency.*

At a frequency of 4 GHz, a difference in temperature of 0.035 °C was observed while at 2 GHz no measurable difference was discernible. Calculations were made (appendix) which show that at 4.0 GHz, with a temperature difference of 0.035 °C, an error less than 0.1 percent exists.

On the basis of the above tests the rf–d-c substitution uncertainty was estimated to be 0.1 percent at frequencies of 1 GHz and above and 0.05 percent at frequencies below 1 GHz. The resolution of the measurement, due to noise and ambient temperature variations, was approximately 0.05 percent.

4.4. Uncertainty in Input Lines and Mismatch Losses

In the load bodies as described in the preceding section, the only portion of input transmission line not immersed in the oil stream is a length of approximately 3.7 cm. The portion immersed in oil can be considered as a part of the load since the I^2R loss will be absorbed by the oil. The loss in the remaining 3.7 cm of input line was calculated using the following equation: [4]

$$\alpha_T = 8.686l \left[\alpha_0 \sqrt{f} \frac{1/a + 1/b}{\ln b/a} + \frac{\omega}{2} \sqrt{\mu\epsilon} \tan \delta \right] dB \quad (3)$$

where $\alpha_0 = \frac{1}{2} \sqrt{\frac{\pi\epsilon}{\sigma}}$ and $\omega = 2\pi f$,

ϵ = dielectric constant,
μ = permeability,
σ = conductivity of metal used for inner and outer conductors,
a = inner conductor radius,
b = outer conductor inside radius,
f = frequency,
l = input line length, and
$\tan \delta$ = loss tangent of dielectric.

The results of this calculation as a function of frequency are plotted in figure 4. The uncertainty in the calculations of the loss factor is due to the combined uncertainty in measuring a, b, l (length = 3.7 cm), and in the assumed values of ϵ, μ, σ, and $\tan \delta$. The combined uncertainty of these factors could be as great as ±35 percent at 4 GHz. This was primarily due to the uncertainty in the thickness of the silver plating on the inner and outer conductors of the line, hence an uncertainty in the value of σ. For example, at 4 GHz the following data were obtained from calculations for the loss of the input line:

Loss in input connectors	0.0026 ± 0.0005 dB
Loss in straight region	.0030 ± .0009 dB
Loss in taper region	.0028 ± .0019 dB
Total loss	.0084 ± .0033 dB

The loss is then 0.2 ± 0.07 percent or an efficiency of 99.8 ± 0.07 percent. Thus the power at the input connector was higher than that measured by the calorimeter by the factor 1.002 ± 0.0007. The rf efficiency factor versus frequency is shown in figure 4. At lower frequencies the loss in the input line decreases and below 10 MHz the efficiency is assumed equal to unity.

Due to the fact that the calorimeter loads are not perfectly matched to Z_0, part of the incident power will be reflected. Because a measurement of the incident power is usually desired, reflections cause the calorimeter to read low with respect to the incident power. The impedance of both loads was matched to Z_0 (50 Ω) using a time domain reflectometer so that VSWR < 1.03 ±0.01 at frequencies from 1.0 to 4.0 GHz and < 1.015 ±0.005 at frequencies below 1.0 GHz. A plot of VSWR versus frequency for each load is shown in figure 5.

At the maximum VSWR of 1.03, 0.02 percent of the incident power will be reflected resulting in a 0.02 percent error in the power measurement if the reflection loss is neglected. For a VSWR of 1.015, the reflected power is only 0.01 percent of the incident source. Combining the maximum expected uncertainty in the correction factor for rf efficiency with the maximum mismatch error gives 0.09 percent above 1 GHz and 0.04 percent below 1 GHz.

4.5. D-C Power Measurement Error

The d-c power substituted in the loads is calculated from measured values of current and voltage. The voltage drop across a 1 Ω standard resistor in series with the termination is measured with a potentiometer

385-582 O - 70 - 5

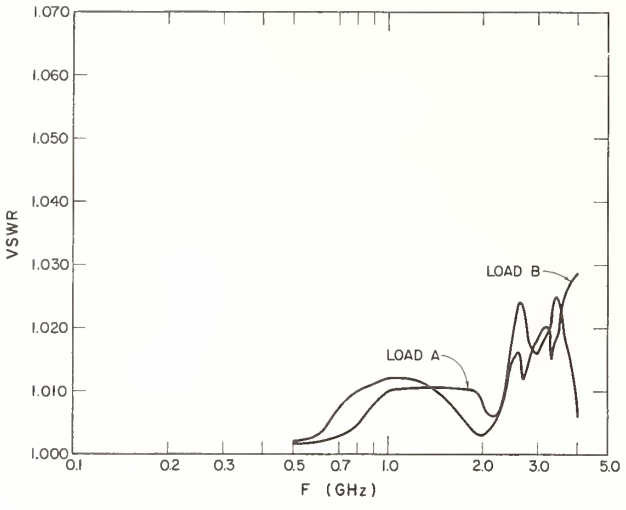


FIGURE 5. *VSWR versus frequency of DLFC loads.*

to obtain the current flowing into the loads, and the voltage drop across the loads is measured directly with a digital voltmeter. The power is then found by the simple equation $P=VI$. Both V and I can be determined accurately to 0.01 percent giving a maximum d-c measurement error of 0.02 percent.

The errors discussed are believed to be the only significant ones in the calorimeter and measuring system described herein. The overall limit of error can be found by adding the maximum values:

	0–1 GHz	1–4 GHz
1. Flow division stability	0.15	0.15
2. Detection system	.02	.02
3. RF–D-C substitution	.05	.10
4. Line loss and rf reflection (mismatch errors)	.04	.09
5. D-C measurement error	.02	.02
Limit of error	0.28%	0.38%

5. Conclusions

The dual-load flow calorimeter was designed and constructed to provide a reference standard for cw power measurements in the range 2 to 100 W at frequencies up to 4.0 GHz. A maximum uncertainty or error limit of 0.38 percent was achieved. This is a significant improvement over prior capabilities at these power levels. Also the frequency range of the NBS reference standards was extended from 1.0 GHz up to 4.0 GHz in the power range of 5 to 100 W.

Intercomparisons at 1, 3, and 4 GHz have been made between this calorimeter and the reference standard dry-load calorimeter which has a total estimated uncertainty of 0.35 percent. The intercomparisons included measurements at power levels of 2 W and 4 W and the disagreement was no greater than 0.2 percent.

6. References

[1] J. P. Vinding, An accurate calorimeter for high microwave power, The Microwave Journal **4**, No. 1, 41–46 (Jan. 1961).
[2] P. A. Hudson, A calorimetric reference standard power meter for the frequency range 0–4000 MHz. Paper in process.
[3] D. Woods, Improvements in precision coaxial resistor design, IRE Trans. on Instr. **I–II**, Nos. 3–4 305–309 (Dec. 1962).
[4] S. Ramo and J. R. Whinney, Fields and waves in modern radio, (John Wiley & Sons, Inc., New York, 2d edition, fifth printing, Nov. 1960). See Table 9.01 between pages 364–365. Equation (3) was derived using equations in this table.

7. Appendix

7.1. Analysis of RF–D-C Substitution Error

In the following analysis of the rf–d-c substitution error, an upper bound is arrived at which, though derived from approximate considerations, is felt to have significance.

The total input power to the calorimeter, P_t, is related to the power absorbed by the oil, P_a, and the power P_c absorbed elsewhere, by

$$P_t=P_a+P_c. \tag{1A}$$

Using additional subscripts, rf and d-c, to indicate when the input power is respectively rf and d-c (1A) becomes

$$P_{t\mathrm{dc}}=P_{a\mathrm{dc}}+P_{c\mathrm{dc}} \tag{2A}$$

and

$$P_{t\mathrm{rf}}=P_{a\mathrm{rf}}+P_{c\mathrm{rf}}. \tag{3A}$$

The measuring technique used with the calorimeter results in

$$P_{a\mathrm{dc}}=P_{a\mathrm{rf}}. \tag{4A}$$

Using (2A), (3A), and (4A), one can obtain the relation

$$\frac{P_{t\mathrm{rf}}}{P_{t\mathrm{dc}}}=1+\frac{P_{c\mathrm{dc}}}{P_{t\mathrm{dc}}}\left[\frac{P_{c\mathrm{rf}}}{P_{c\mathrm{dc}}}-1\right]. \tag{5A}$$

Thus, if $P_{c\mathrm{rf}}=P_{c\mathrm{dc}}$, there would be no substitution error. However, this condition may not hold for all frequencies and a measure of the ratio of $P_{c\mathrm{rf}}$ to $P_{c\mathrm{dc}}$ is obtained from the following considerations.

As noted in the text, there is little chance for heat losses to the external environment except possibly at the junction of the load resistor with the outer conductor. A thermopile connected externally between this area and an ice bath indicated that there was no

difference in the temperature for frequencies up to 2 GHz when equal rf and d-c power were alternately applied to the calorimeter. However, at 4.0 GHz a temperature difference of 0.035 °C was measured. Because of the construction and nature of operation of the calorimeter, the principle mode of heat transfer to the external environment is by conduction. Since the rate at which heat flows by conduction is proportional to the temperature difference between the source and sink,

$$\frac{P_{c\,\mathrm{rf}}}{P_{c\,\mathrm{dc}}}=\frac{T_{\mathrm{rf}}-T_{\mathrm{amb}}}{T_{\mathrm{dc}}-T_{\mathrm{amb}}}, \tag{6A}$$

where T_{rf} and T_{dc} are the respective temperatures at the junction of the load resistor and outer conductor when rf and d-c power are alternately applied to the calorimeter. The ambient temperature is noted by T_{amb}. From experimental data at 4 GHz,

$$\frac{P_{c\,\mathrm{rf}}}{P_{c\,\mathrm{dc}}}=1.023. \tag{7A}$$

An upper bound for the ratio of $P_{c\,\mathrm{dc}}/P_{t\,\mathrm{dc}}$, obtained from using the dual-load calorimeter as an absolute flow calorimeter, is

$$\frac{P_{c\,\mathrm{dc}}}{P_{t\,\mathrm{dc}}}=0.03. \tag{8A}$$

Substituting (7A) and (8A) into (5A), the upper limit for the substitution error is 0.07 percent which was increased to 0.1 percent because of the approximations used.

(Paper 71C2–250)

1.a. **Mismatch errors in microwave power measurements,** R. W. Beatty and A. C. MacPherson, Proc. IRE 41, No. 9, 1112-1119 (September 1953).

Expressions are derived for error due to mismatch when a UHF or microwave power meter is calibrated by comparison with a standard power meter. Three different methods are considered: (a) alternate connection to a stable power source, (b) the use of a microwave junction which simultaneously supplies power to the uncalibrated power meter and the standard power meter in a known ratio, (c) alternate connection to a microwave junction. The relative merits of the methods are discussed.

Expressions are derived for error due to mismatch when using a calibrated power meter in the following situations: (a) direct connection of power meter to power source, (b) reduction of power into the power meter by means of an attenuator, (c) reduction of power into the power meter by means of a directional coupler.

1.b. **An improved method of measuring efficiencies of ultra-high-frequentcy and microwave bolometer mounts,** R. W. Beatty and Frank Reggia, J. Res. NBS 54, No. 6, 321-327 (June 1955).

A method is presented for measuring efficiencies of bolometer mounts used for ultra-high-frequency and microwave power measurement. It is based upon the impedance method of Kerns, but avoids the direct measurement of impedance. Pertinent theory is developed, and the errors in measuring efficiency by this method is analyzed and discussed. Experimental results are given.

1.c. **A transfer instrument for the intercomparison of microwave power meters,** G. F. Engen, IRE Trans. Instr. I-9, No. 2, 202-208 (September 1960).

1.d. **A survey of microwave power-measurement techniques employed at the National Bureau of Standards,** Glenn F. Engen, Proc. IEE, 109, Part B, Suppl. 23, 734-739 (May 1962).

The bolometric technique of power measurement is an important part of the microwave art. The paper describes certain refinements and extensions of this basic method which have been developed at the Boulder Laboratories of the National Bureau of Standards and which provide the basis for a microwave power-calibration service. The attendant problems may be divided into three categories: (i) measurement of the substituted or bolometric bias power, (ii) evaluation of the d.c. r.f. substitution error, and (iii) determination of the bolometer mount efficiency.

1.e. **A variable impedance power meter, and adjustable reflection coefficient standard,** Glenn F. Engen, J. Res. NBS 68C, (Eng. & Instr.), No. 1, 7-24 (Jan.-Mar. 1964).

Most microwave power meters, such as those of the bolometric and calorimetric types, completely absorb the power which they indicate. The use of these meters is thus usually accompanied by the requirement to either (1) determine the power at another place in the microwave circuit from the meter reading, or (2) determine the power which will be delivered to a different load which is to be substituted for the meter.

The evaluation of these relationships plays an important role in the power measurements art.

These problems are alleviated by the power measuring device to be described. This instrument, which is based on reflectometer techniques, may be used either as a feed-through power meter which indicates the power delivered to loads of arbitrary impedance, or as a terminating power meter whose input impedance can be adjusted to arbitrary values without requiring recalibration.

In a slightly modified form the instrument also gives an indication of the reflection coefficient magnitude to which it has been adjusted. A comprehensive error analysis permits the accuracy to be determined under general operating conditions.

1.f. **A precision RF power transfer standard,** P. A. Hudson, IRE Trans. Instr. I-9, No. 2, 280-283 (September 1960).

1.g. **A microwave microcalorimeter,** A. C. MacPherson and D. M. Kerns, Rev. Sci. Instr. **26,** No. **1,** 27-33 (January 1955).

This paper gives a brief account of a calorimeter developed at the National Bureau of Standards for the measurement of microwave power with relatively high accuracy (better than 1 percent) and at low (milliwatt) power levels. The calorimeter is identifiable as an electrically (dc) calibrated, aneroid microcalorimeter of the Joule twin type. A bolometer mount assembly serves as a wave-guide termination for the absorption of microwave power and constitutes the calorimetric body whose temperature rise is observed. This arrangement enables the use of a novel indirect measurement technique and tends to insure calorimetric equivalence of dc and rf heating. The equivalence error is investigated with the aid of heat flow analysis and auxiliary experiments, and an upper bound for the error is assigned.

1.h. **Microwave power measurements employing electron beam techniques,** Harold A. Thomas, Proc. IRE 45, No. 2, 205-211 (February 1957).

A new electron beam technique for measuring microwave power flow, either cw or pulse, in waveguides is described. This technique consists of accelerating an electron beam transversely through an evacuated section of waveguide carrying power in the TE_{10} mode. The transit time of the electrons is adjusted to a value which gives maximum interaction of the field in the guide with the electrons, i.e., electrons gain maximum energy. The energy gained by the electrons is measured by means of a dc stopping potential which can be related to the field. Power is then calculated from the Poynting vector.

2. Sinusoidal Voltage and Current

Papers

Abstract

Voltage Measurement at High and Microwave Frequencies in Coaxial Systems

M. C. SELBY, SENIOR MEMBER, IEEE

Abstract—**The progress and up-to-date state-of-the-art in measuring voltages at frequencies to 10 GHz and higher is briefly described. The trend towards higher frequencies in voltmeter design is indicated. Advantages of voltage measurements and standards over computation of voltages from power and impedance measurements are briefly discussed. Some pending development problems are indicated and major steps are proposed to improve the application of voltmeters at frequencies above 30 MHz.**

Introduction

IN SECURING a national and international coherent system of echelons to measure a quantity or measurand[1] (e.g., voltage, within a given range of frequency, magnitude, and other parameters), three major steps are generally involved: 1) development of basic reference standards of magnitude (standard sources), or standard measuring instrumentation, 2) development of interlaboratory reference standards, and 3) development of "field" reference sources and measuring instruments. Satisfactory progress seems to have been made in all three steps during the last two decades to meet needs on measurement of sinusoidal voltages with one possible exception: field applications of available instrumentation does not appear to be in step with potential capabilities and benefits; this lag will be discussed to some extent in a later section.

There are two schools of thought as to which of the above three development steps should come first: 1) the development of field instruments presumably justified by immediate or impending real needs, or 2) the development of basic standards without the pressure of needs in order to avoid the serious time lag in meeting future needs. It will be very briefly shown below that, in case of voltage, fortuitously, development of both standards and field instruments has been

Manuscript received March 7, 1967.

The author is with the National Bureau of Standards, Boulder, Colo.

[1] The ASA in "Definitions of electrical terms," Group 30, Instruments, Meters and Meter Testing, C42.30, p. 5, 1957, defines "measurand" as a "physical" quantity, property, or condition which is to be measured. This term is used here in a more restricted sense, namely, as a quantity in a given limited range or domain of its parameters including frequency, magnitude, waveform, circuit configuration, etc.

progressing almost simultaneously during the last couple of decades. Thus, this particular ambivalent development offers no resolution of the difference of opinion; perhaps additional experience over the next decade or so and a close exploration of the field of application of voltmeters will help to provide an answer.

Up-to-Date Status

A voltage between points a and b is defined as the work required to carry a unit change from a to b, i.e.,

$$V_{ab} = -\int_a^b E \cdot dl \qquad (1)$$

where

$E \equiv$ field strength along the path (a vector)
$l \equiv$ length and direction of path.

This definition applies when the value of the integral is not a function of the path chosen between a and b, i.e., when the system is conservative [1]. This condition is met in all electrostatic and magnetostatic systems, and in all dynamic systems operating in the TEM mode provided the integration is performed in the transverse plane. Voltage measurements are practicable, for example, even in a 7-mm 50-Ω coaxial system (one of the most widely used) at frequencies to 18 GHz and in a 3.5-mm coaxial system to 35 GHz. Such a wide frequency range, starting with dc, of necessity, covers a number of specific measurands of voltage as a quantity. The development of measurement techniques and of standards reached its first plateau at "broadcast" frequencies, say, at several MHz, another plateau at about 30 MHz, and a third at 1 GHz. Indications are that the next plateau will be at 10 or 20 GHz depending on how profitably the accomplishments to date are digested in practical applications.

The first standard source of RF voltage (which may be termed "basic" for reasons indicated below) was developed at the National Bureau of Standards (NBS) to cover the approximate frequency range of 500 kHz to 1 GHz; it proved successful for levels of about 0.02 to 1 volt and was reproduced in a number of laboratories in the U.S.A. and abroad [2]. The basic principle of this standard is as follows. A nonreactive resistance element R_m is placed in a TEM-mode field in such a way that the spatial form of the field is negligibly altered. Irrespective of any other reactive or resistive components or networks which may be shunting R_m, the resultant conduction current through R_m renders the true sinusoidal voltage across it. Consequently, R_m integrates the transverse electric field components. This voltage can be detected and measured by employing the well-known bolometric, thermoelectric, or photoelectric effects. Figure 1 shows a schematic diagram illustrating this principle. Radio frequency energy is fed into a coaxial line generating a TEM mode along the line, including the connector to the voltmeter being standardized. The input plane of the connector is placed at the plane of R_m; thus the voltage across R_m is applied to the voltmeter. The figure shows the plane

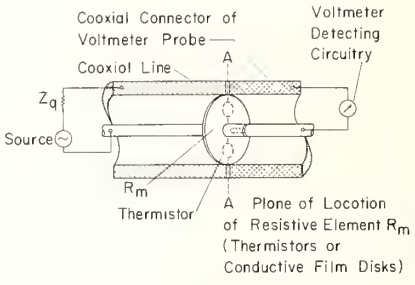


Fig. 1. Schematic diagram illustrating principle of standardizing RF voltmeter in terms of known voltage generated in R_m located in a transverse plane A of a TEM mode.

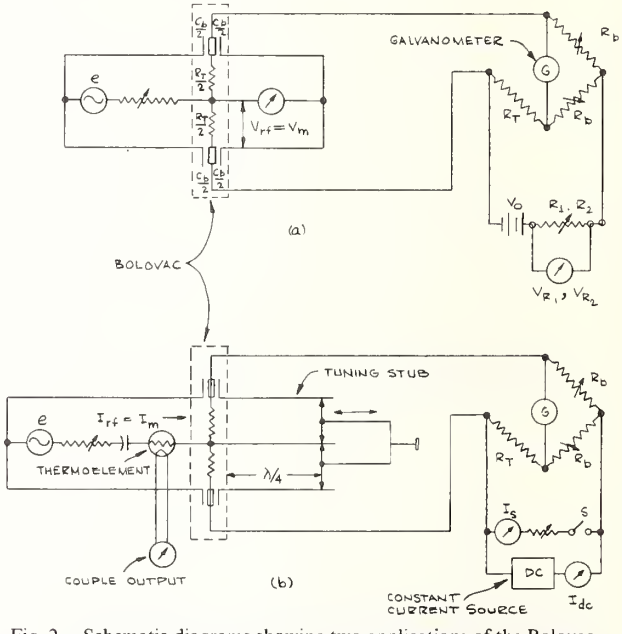


Fig. 2. Schematic diagrams showing two applications of the Bolovac. (a) For voltage standardization. (b) For current standardization.

where R_m is located and where the probe of the voltmeter is connected. In the first NBS standard, thermistors were used to 1 GHz and the RF-for-dc power substitution technique was most convenient. Thus, though power substitution is used, the manner of its application is fundamentally different from the way power dissipation is employed in a power calorimeter or in a power-bolometer head. In the latter two cases, one must determine the power absorbed by a system of components. For example, when RF power is substituted for dc, other losses in the system (outside the bolometers or other sensing elements) must be considered. Another source of uncertainty here is the relatively low sensitivity of calorimetric and bolometric conventional measurements of power (absorbed by a load) to reflected power in case of low voltage standing wave ratios (VSWR's). For example, a VSWR of 1.05 will reduce the absorbed power by about 0.05 percent; this reduction cannot be detected by present-day calorimetry. Therefore, better power measurement accuracy is inherently possible from voltage measurements than with the above power measurement techniques.

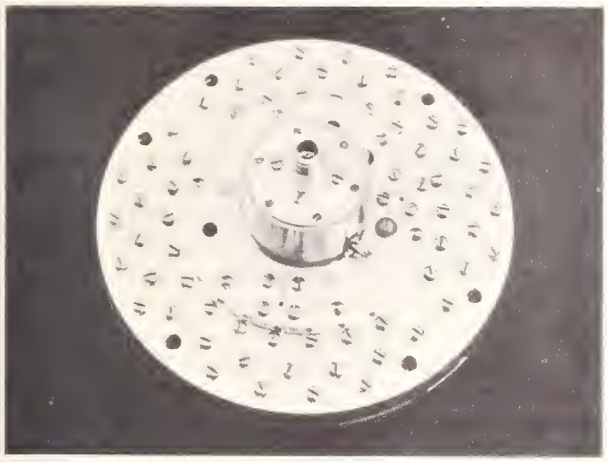


Fig. 3. Bolovac assembled for RF current standardization.

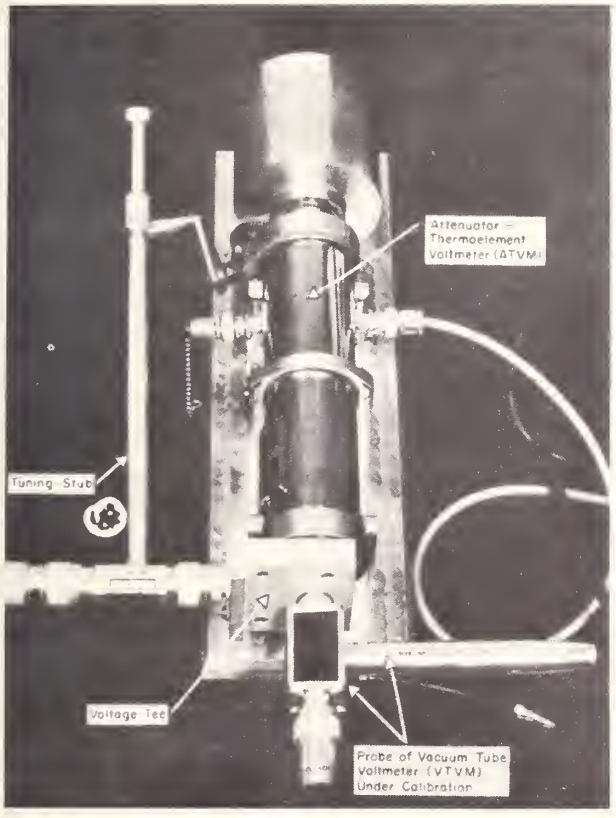


Fig. 4. Head of "unknown" VTVM under calibration connected with a Tee to an ATVM interlaboratory standard.

The NBS thermistor voltage standard was limited to frequencies below 1 GHz with uncertainties to 1 percent because of the residual inductances of one-mil leads and the questionable equivalence of the dc-to-RF resistances of the thermistor beads. These and other limitations have been greatly reduced in a new NBS bolometric voltage and current standard (the Bolovac) employing a split-disk film-type bolometer [3]. This standard has, so far, been successfully used to 8 GHz for voltage and to about 2 GHz for current; it has a potential range up to 18 GHz and higher. Its present level range is 0.1 to 5 volts and the potential range is much wider, subject to perfecting fabrication techniques of the bolometer disks. Currents of 5 to 100 mA were standardized with the Bolovac. The estimated uncertainties were ± 4 percent at 8 GHz and ± 1 percent at 2 GHz for voltage, and ± 1 percent for current at 1 GHz. These uncertainties are based on agreement with conventional bolometer-bridge power measurements. Figure 2 shows the principle of application of the Bolovac, and Fig. 3 shows the Bolovac assembled for current standardization.

Another basic NBS standard of voltage to 50 MHz, still in use, is the old "slide-back" technique employing an electron beam deflection for levels of 5 to 100 volts [4]. This standard years ago earned its first-echelon classification in a number of laboratories on the basis of analytical reasoning and agreement with other independent techniques.

Three widely used interlaboratory reference standards were developed at the NBS, namely the Thermal Converter [5], the RF Micropotentiometer [6], and the ATVM (attenuator-thermoelement type of voltmeter) [7]. They are noted for superior calibration stability (a prime requisite for interlaboratory functions) and cover a range of frequencies from about 30 kHz up to 1 or 2 GHz; their combined voltage range is approximately 1 μV to 20 V at 1 GHz and higher at lower frequencies, e.g., 0.2 μV to 200 V at 30 MHz.

Another tool singularly useful in calibration work was developed by the NBS and is available on the market; this is a coupling voltage Tee [8] making possible the intercomparison of two voltage-indicating instruments at frequencies to 1 or 2 GHz irrespective of the input impedances of these instruments. This Tee essentially eliminates the detrimental effects of the standing waves and higher modes and reduces the RF power requirements to a level negligible compared with that necessary when matched reference standards are used. In application, the voltages on both sides of this Tee are equal to within an uncertainty which does not exceed 1.5 percent at 1 GHz, even with an indicator input VSWR of 200. Figure 4 shows a VTVM being calibrated in terms of an ATVM employing one of these voltage Tees.

Interlaboratory standards for frequencies above 1 or 2 GHz are still to be developed. There need be no strict demarcation line between basic, interlaboratory, and field instruments. Either a voltage bolometric head or a field voltmeter could be used to establish the chain of calibration. Nevertheless, past experience urges independent interlaboratory standards in the interest of better accuracy, economy, and expediency. Prospects seem fair of employing the same principles for higher-frequency interlaboratory standards as at the lower frequencies. Mechanical dimensions of sensors will have to be scaled down and perhaps lower accuracies tolerated until radically new methods are devised.

An extensive review of the up-to-date state of field voltmeters and of voltage ("signal") generators at the fre-

quencies up to, say, 10 GHz is beyond the scope of this paper. Facts indicate a trend over the last two or three decades towards marketing sinusoidal voltage measuring devices for frequencies well into the GHz range. For example, millivoltmeters to 2.5 GHz were described in the British literature in 1962; the U. S. market has been offering meters to 2.5 GHz at least since 1954; a German source marketed a meter to 5 GHz in 1957; a U. S. supplier is presently offering a meter for both amplitude and phase to 1 GHz and an oscilloscope to 12 GHz, both employing the sampling technique. Many signal generators on the market specify their output to 10 GHz in terms of volts rather than watts; however, this may be a result of habit rather than of need. It is difficult to ascertain whether this trend is based on existing and potential future needs, or on mere speculation. Repeated efforts in the past to determine "needs" seem to lead to the conclusion that the areas of R&D and application welcome the best measurement accuracies available; these accuracies are incorporated in projected activities and the "trade-off" principle (i.e., improving performance of one component to compensate for a deficiency in that of another) is applied to predict overall system performance. This is contrary to the hoped-for clear statement of "required" or "needed" accuracies of measurands. Nevertheless, "needs" are still continuously sought to justify priorities and expenditures, apparently as a matter of established practice or necessity. It may, therefore, be worth while *first* to explore briefly the justifications for advancing voltage measurement capabilities, say, to frequencies practicable in coaxial lines, and *second*, to critically examine direct voltage measurements as opposed to the alternate approach of computing voltages from measured power and impedance values at various points in a system. It should perhaps be pointed out here that observations and statements (outside of factual information) given below shall be construed as personal deductions, opinions, and recommendations of the author.

The apparent reasons why voltages should be measured directly are as follows.

1) Voltage is to be treated as an independent quantity wherever possible, just as power, current, impedance, etc. are treated.

2) Under certain conditions, it seems more profitable to measure voltage and VSWR (instead of obtaining voltage from power and impedance measurements) because this procedure yields incident and reflected powers, impedance, voltage and current distributions, and power absorbed by a load with higher accuracy.

3) The measurement of voltage and VSWR is a more sensitive way of detecting troubles in some components of a system. It may show up drifts of load components sooner than power measurements.

4) Voltage and VSWR can be measured with the system in operation, whereas operation is generally interfered with if power and impedance are to be measured. (A directional coupler can be used as a voltmeter in order to obtain magnitude and phase information.)

5) Voltage distribution measurement yields phase information; power does not.

6) Voltage measurements are applicable in coaxial and strip-line systems, both of which are becoming more prominent due to miniaturization of microwave systems and to the development of baseband systems extending to the order of 20 GHz.

7) In modern communication systems there is a need to measure pulsed-voltage parameters. One very useful tool in assessing pulsed-system performance is the frequency-domain transfer function whose determination requires voltage amplitude and phase measurements as functions of frequency.

8) Voltage measurements may be the basis of field strength and field polarization measurements, voltage spectrum analysis, distortion measurements, dispersion measurements, *E*-field sensing device studies, determination of voltage characteristics of materials and components, voltage divider characteristics, and breakdown limits of circuitry (e.g., commercial miniature coaxial cables are designed for operation to 1 or 2 kV peak at frequencies to 10 GHz).

To illustrate the practical opportunities rendered by voltage standardization, one may consider the power propagated along a lossless coaxial transmission line given by

$$P = \frac{V_1 V_3}{R_0}\left[1 - \frac{1}{4}\left(\frac{2V_2^2 - V_1^2 - V_3^2}{V_1 V_3}\right)^2\right]^{1/2} = \frac{V_1^2}{R_0}\frac{V_3}{V_1}\left[1 - \frac{1}{4}\left(\frac{2V_2}{V_1}\frac{V_2}{V_3} - \frac{V_1}{V_3} - \frac{V_3}{V_1}\right)^2\right]^{1/2} \tag{2}$$

where V_1, V_2, and V_3 are voltages measured in three respective planes spaced $\lambda/8$ apart, and $R_0 \equiv$ characteristic impedance of the transmission line (assumed to have negligible losses). Presently, commercial precision 50-Ω coaxial line sections have an uncertainty of ± 0.05 percent in R_0.

The uncertainty in the absolute value of either V_1 or V_3 is the source of the largest error, because the uncertainty contributed by the voltage ratio terms may be made negligibly small. This approach effectively accounts for the reflected power.

The above voltages may be measured by means of a calibrated voltmeter connected to probes located in the desired planes of a transmission line. A combination of any probe and voltmeter may be readily calibrated (see Fig. 5). A standardizing voltage is applied to the open end of the line section containing three probes, the other end of the section is short-circuited, and the ratio of the voltage at the probes to the standardizing voltage is

$$\frac{V_x}{V_m} = \frac{\sinh \alpha x \cos \beta x + j \cosh \alpha x \sin \beta x}{\sinh \alpha l \cos \beta l + j \cosh \alpha l \sin \beta l} \tag{3}$$

where

V_m and V_x are the standardizing and unknown probe voltages, respectively,

l and x are the respective distances from the short-

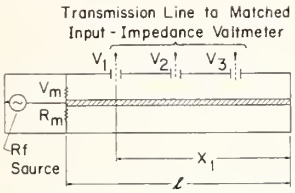


Fig. 5. Schematic diagram illustrating setup to calibrate an RF voltmeter in combination with transmission-line probes.

circuited end of the line section to V_m and V_x,
α is the attenuation constant of the line, and
β is the phase constant.

For negligible α,

$$V_x = V_m \frac{\sin \beta x}{\sin \beta l}. \tag{4}$$

However, the most impelling arguments in favor of direct voltage measurements at the frequencies in question are the same as those for dc and low frequencies. One could not generally justify the use of power and impedance measurements as contrasted to voltage measurements for frequencies, say, to 10 MHz. The former must of necessity be used as long as voltages cannot otherwise be measured, but techniques for direct voltage measurements are desirable, if not indispensable, in all TEM-mode operating systems.

Assuming it is agreed that voltages should be measured, in preference to power and impedance, one may still justly question the necessity of having voltage reference standards because one may use existing power and impedance standards for this purpose. The objections to the latter approach are largely the same as indicated in the above reasons in favor of measuring voltage. The major advantages in having independent voltage standards may be briefly summed up again.

1) Voltage standards eliminate the need of impedance and power measurements for standardizing voltages.
2) Considerably less power is needed to develop voltages with high impedance meters than with "matched" meters.
3) When voltage measurements are applicable, it seems much easier and cheaper to fabricate film disks of arbitrary resistance for voltage standardization than low-VSWR tractorial or other loads for power standards.
4) There are no efficiency measurement problems present with voltage measurements; once the VSWR and α of the coaxial line is known, the power can be accurately determined.
5) The accuracy of the standard voltages is inherently better when measured directly.

It should be emphasized that the author does not advocate abandonment of calorimetric, bolometric, or other direct measurements of power. The attempt above is to point out the fallacy of abandoning direct measurement of voltage.

Apparent Lag in Application of Voltmeters in the Field

It was pointed out above that the art of developing voltage standards and field measuring instruments has been progressing during approximately the last two decades. However, the voltmeters available for field applications do not seem to be utilized to the expected extent and efficiency possible. Repeated interrogations of users resulted in the same revelation that high-frequency voltmeters are usually used merely as indicators or detectors (and not for measuring actual voltage values), which was the practice 20 or more years ago. One gets the impression that voltmeters are not widely used to accurately measure voltages above about 30 or 100 MHz. Significant advantages of voltage measurements are thus lost.

Three steps seem necessary to increase the use of voltmeters at the higher frequencies. First, the system components should be designed with properly placed voltage probes, e.g., capacitive probes of permissible insertion depths, so as not to unduly perturb system operation. A similar approach is employed, for example, by incorporating Pitot tubes to measure fluid velocity or airplane speeds as a matter of general practice. Second, voltmeter heads should be provided with properly designed connectors to mate with the above voltage probes; it may be possible to employ, for example, line sections with probes, spaced $\lambda/8$ apart, for insertion into transmission lines of a system for purposes of voltage, power, and impedance checks. Third, competence of personnel should be developed in the proper use of field voltmeters at the higher frequencies so that maximum accurate information may be derived under various conventional and special applications.

In conclusion, the reader's attention is called to the "IRE Technical Committee Report on the State-of-the-Art of Measuring Sine-Wave Unbalanced RF Voltage" published in 1963 [9]. This report reflects the state up to 1962 and will hopefully be brought up to date in the near future. The trend towards voltage measurements into the GHz range is indicated and a bibliography of the subject matter is included. This committee report plus the bibliography of an NBS (1949) circular [4] and the bibliography at the end of this paper should serve as a fairly comprehensive guide of references back to about 1930.

References

[1] R. I. Sarbacher and W. E. Edson, *Hyper and Ultrahigh Frequency Engineering*. New York: Wiley, 1943, p. 8.
[2] M. C. Selby and L. F. Behrent, "A bolometer bridge for standardizing radio-frequency voltmeters," *J. Res. NBS*, vol. 44, pp. 15–30, January 1950.
[3] M. C. Selby, "Bolometric voltage and current (Bolovac) standard for high and microwave frequencies," to be published.
[4] ——, "High-frequency voltage measurement," Nat'l Bur. Standards, NBS Circular 481, September 1949.
[5] F. L. Hermach and E. S. Williams, "Thermal voltage converters for accurate voltage measurements to 30 megacycles per second," *AIEE Trans.* (*Communication and Electronics*), vol. 79, pt. I, pp. 200–205, July 1960.
[6] M. C. Selby, "Accurate radio-frequency microvoltages," *AIEE Trans.* (*Communication and Electronics*), vol. 72, pt. I, pp. 158–163, May 1953.

[7] ——, "Stable radio-frequency voltmeters," *NBS Tech. News Bull.*, vol. 40, pp. 29–30, February 1956.
[8] ——, "Coaxial *T* for radio-frequency voltmeter calibrations," Nat'l Bur. Standards, NBS Tech. Note 263, June 1965; patent pending.
[9] "IRE technical committee report on the state-of-the-art of measuring sine-wave unbalanced RF voltage," *Proc. IEEE*, vol. 51, pp. 575–580, April 1963.

Bibliography

[10] M. C. Selby, "The system of electromagnetic quantities at 30 kHz to 1 GHz," *Metrologia*, vol. 2, January 1966.
[11] F. K. Weinert, "The RF vector voltmeter for amplitude and phase measurements from 1 MHz to 1000 MHz," *Hewlett-Packard J.*, vol. 18, pp. 2–12, May 1966.
[12] "Voltmeter spans 1 Kc to 1 Gc using random sampling," *Electronic Design*, p. 58, December 1965.
[13] M. C. Selby, W. J. Blank, and R. P. Chariton, "Voltmeter calibration to 1 GHz," presented at the WESCON Electronic Show and Conv., August 1965.
[14] T. D. Towers, "Electronic laboratory instrument practice," *Wireless World*, pp. 198–203, April 1965.
[15] E. Dilatush, "Low-level RF voltage measurements," *Electronic Equipment Engrg.*, vol. 13, pp. 88–92, January 1965.
[16] K. H. Klarer, "Measuring high frequency voltage with photometric volometers," *ISA J.*, vol. 11, pp. 69–72, July 1964.
[17] M. C. Selby, "Single-bolometer series-parallel circuit for accurate voltage and power measurement," ISA Preprint 21.6-1-64, October 1964.
[18] ——, "Pulsed and CW sinusoidal voltage and current measurements," *J. Res. NBS (Radio Science)/USNC-URSI*, vol. 68D, pp. 533–536, May 1964.
[19] F. X. Ries and G. Rebuldela, "High frequency microvolt measurements," ISA Preprint 37.2.63, September 1963.
[20] B. P. Hand, "Radio-frequency voltage measurement," *Electro-Tech.*, vol. 72, pp. 109–116, October 1963.
[21] R. E. Lafferty, "Improving the accuracy of RF voltage measurements," *Electronic Ind.*, July 1963.
[22] D. Woods, "A coaxial millivoltmeter/milliwattmeter for frequencies up to 1 Gc/s," *Proc. IEE (London)*, vol. 109, pt. B, Suppl. 23, pp. 750–756, May 1962.
[23] I. A. Harris, "Electron transit time and other effects in a valve voltmeter operating at extremely low current," *J. Brit. IRE*, vol. 24, November 1962.
[24] "Summary of vacuum tube voltmeters" (Spec. Rept.), *Electronic Ind.*, January 1962.
[25] D. Woods, "Precision instruments for coaxial line measurements up to 4 Gc/s," *Proc. IEE (London)*, vol. 108, pt. B, pp. 490–494, September 1961.
[26] T. C. Anderson, "A new RF millivoltmeter for convenient measurements to kmc," *Hewlett-Packard J.*, vol. 12, September 1960.
[27] A. M. Fedorov, "Technique and apparatus for determining the transit error in diode voltmeters," *Izmeritel'naya Tekhnika*, pp. 47–48, January 1961; for English transl. see *Meas. Techniques*, pp. 63–65, September 1961.
[28] E. Uiga and W. F. White, "Techniques and errors in high frequency voltage calibration," presented at the NBSBL Conf. on Standards and Electronic Measurements, June 1960.
[29] M. C. Selby, L. F. Behrent, and F. X. Ries, "RF voltmeter calibrating consoles," *IRE Nat'l Conv. Rec.*, pt. 5, pp. 251–257, 1958.
[30] M. C. Selby, "High-frequency standards of the Electronic Calibration Center, NBSBL," *IRE Trans. on Instrumentation*, vol. I-7, pp. 262–270, December 1958.
[31] J. Turban, "A sensitive crystal-diode voltmeter/powermeter for 30 c/s to 5000 Mc/s," *Siemens Z. (Germany)*, vol. 31, pp. 100–102, July 1957.
[32] W. K. Volkers, "Conclusive voltage calibration of high frequency signals," Millivac Instruments, Inc., Schenectady, N. Y., Tech. Rept. TR-43, vol. 9, February 1954.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967
pp. 877-882
The Institute of Electrical and Electronics Engineers, Inc.
PRINTED IN THE U.S.A.

The Measurement of Current at Radio Frequencies

WINSTON W. SCOTT, JR., MEMBER, IEEE, AND NOLAN V. FREDERICK, STUDENT MEMBER, IEEE

Abstract—**The state-of-the-art of radio-frequency current measurement is reviewed with emphasis on the most useful standards. In particular, thermocouple and electrodynamic ammeters are discussed in detail. Reference is made to photoammeters, air thermometer milliammeters, and other types of current-measuring apparatus, some of which deserve additional development. Extensive referencing is included for the convenience of investigators interested in an intensive review of radio-frequency current measurements.**

I. Introduction

IN THE EARLY DAYS of radio, considerable interest existed in the construction of radio-frequency ammeters. Literature in later years, during the 1930's and early 40's, seemed more concerned with improving existing methods of measuring current than with the development of new methods. However, relatively little interest was shown following World War II. The National Bureau of Standards (NBS) has now resumed work in radio-frequency current measurements [1]. Presently an attempt is being made to accurately interrelate quantities of voltage, current, power, and impedance at radio frequencies.

Thermocouple ammeters are most extensively used for radio-frequency current measurements, but they must be calibrated using absolute standards or by comparison to dc currents, using instruments with RF-dc corrections which can be calculated. Such instruments include the electrodynamic ammeter, photoammeter, and air thermometer milliammeter. Several other types of current meters, such as the loop and detector [2] or the hot-wire ammeter, are of limited interest.

II. General Comments on Calibration

When radio-frequency ammeters are used to measure current, it is usually desirable to calibrate them or otherwise check on their performance. This requires:

1) the choice of a method for comparing the current through the ammeter with that through the standard,
2) the choice of a standard appropriate for the frequency and uncertainty level desired.

When choosing a method of current measurement at radio frequencies, consideration must be given to the effects of standing waves and stray impedances. These effects may alter the magnitude of the current at the two locations where current is being compared. Stray impedance effects, for instance, becomc increasingly significant at higher frequencies. Before the choice of a standard is made, each type of proposed standard should be studied for advantages and limitations at the desired current level and operating frequency. To obtain the best calibrations, it is desirable to intercompare at least two standards which depend on different operating principles in order to verify the unit measurement. If agreement between these standards is obtained over a wide frequency range, then there is a good probability that the value is accurate.

Manuscript received April 3, 1967.
The authors are with the National Bureau of Standards, Boulder, Colo.

III. Thermocouple Ammeters

Thermocouple ammeters are used extensively because of their low cost, wide frequency range, and relative portability. Most thermocouple ammeters use a thin wire or tube (heater) in the radio-frequency circuit to heat a thermal junction. The output of the thermal junction (often about 10 mV with rated heater current) is connected to a dc millivoltmeter which usually has a scale marked in amperes (see Fig. 1).

Thermocouple ammeters are of simple construction but are very difficult to design properly. In the past, at frequencies above 10 MHz, skin effect within the heater was responsible for the existence of large corrections. For example, at 100 MHz an ammeter rated at 5 amperes which used a 0.28-mm diameter platinum-alloy heater indicated 62.4 percent high because of skin effect [3]–[5]. Skin-effect corrections can be reduced by using thin-walled tubes. One design uses a platinum-alloy tube of 0.71-mm diameter and 0.025-mm wall thickness which theoretically requires a 2-percent correction at 80 MHz [6]. However, to avoid unreasonably large skin-effect corrections at frequencies up to 1 GHz, even thinner walls are required. In fact, evaporated metal films deposited on dielectric substrates offer interesting possibilities for heater design at 1 GHz. The dielectric substrate increases the capacitive effect which opposes the inductive effect in the heater [7].

Thermocouple instruments for one ampere and above are generally constructed with the heater exposed to ambient air. For operation in the milliampere ranges, however, the heater is placed in an evacuated glass envelope, so as to increase heater temperature by reducing heat loss due to air convection. This construction permits the heater resistance and therefore the voltage drop across the heater to be reasonably small. The voltage drop should be small to reduce stray currents which tend to bypass the heater through unavoidable capacitances between leads or between leads and ground. Stray currents may cause an instrument to read either high or low compared with the actual current flowing to a load.

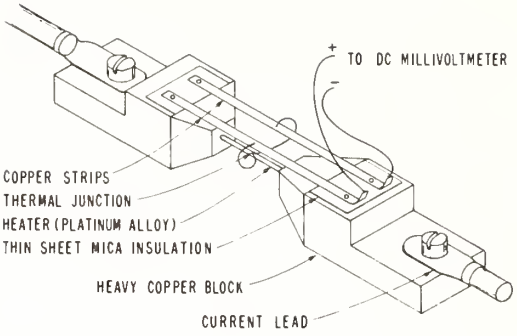


Fig. 1. Typical construction of a thermocouple ammeter.

As the frequency is increased, residual immittances can cause progressively worse problems. For example, dc resistance for a heater designed for operation at 5 amperes may be only 0.03 ohm. However, at 200 MHz the series impedance (mainly caused by series inductance) of the heater can be 60 or 70 ohms and the voltage drop, at rated current, may be 300 to 350 volts [8]. This voltage can force considerable heater bypass current through stray capacitances. This condition must be avoided in current measurements. In an effort to reduce series inductance effects in the heater, one design by McAninch [8] folds the heater (a flat strip) back on itself and separates the two halves by a thin sheet of mica. This arrangement allows use of a coaxial connection which also reduces the series impedance. With this construction, the series impedance is only about 6 ohms and no resonance effects occur below 200 MHz. The ammeter uncertainty at 50 MHz reportedly is 5 percent and at 200 MHz is 20 percent. Several circuits for intercomparing ammeters have been designed to reduce some of the stray impedance effects [9]. At particular frequencies, the distributed capacitances and inductances can interact and create resonances where localized currents become large. Both series and parallel resonances can occur [10], [11]. For example, lead inductance and capacitance between the heater circuit and the thermal junction can series-resonate and allow current to bypass the heater. Under this condition, the ammeter indication is low. Also, heater inductance and capacitance between the heater leads can parallel-resonate and allow current to circulate. In this case, the ammeter indication is high.

An additional uncertainty, which is a function of current amplitude and the magnetic properties of the heater, was reported by Gainsborough in England [12]. He states that under certain circumstances, as the temperature of the heater increases with increasing current, a change in magnetic properties of the heater occurs which substantially alters the skin effect.

An inconvenient problem with thermocouple instruments is that their overload capacity is usually only about 150 percent of full-scale current. Further, the dc output from the thermal junction tends to change from the time current is applied [13], [14]. Conduction of heat from the hot junction may warm the cold junction of the thermocouple, producing a change in the dc output. Another heat source, contact resistance at the ammeter terminals, may also cause difficulties under high current conditions. Goodwin [3] compensates for effects of unequal terminal temperatures and air-temperature variations by connecting the cold junction to thermal compensating conductors.

IV. Electrodynamic Ammeters

Thermocouple ammeters are small, portable, have a wide frequency range, and are convenient to use. However, they should be calibrated at several radio frequencies. One instrument excellently suited to serve as a standard in this calibration is the electrodynamic ammeter. It can be used to determine current over a broad current range (1 to 100 amperes or more) and a broad frequency range by reference to measurements of the basic quantities: mass, length, and time.

Electrodynamic ammeters, also known as electrodynamometers, that are used in the audio-frequency range normally use a movable coil located between two stationary coils. Each coil has many turns of wire and at higher frequencies the inductance and capacitance of these coils cause errors in the ammeter.

One method of reducing these errors at high frequencies is to make the coils of very few turns. This causes a decrease in the sensitivity of the instrument which can be counteracted, to some extent, by reducing the restoring force of the movable element springs. Recent radio-frequency electrodynamic ammeters have a single-turn movable ring supported by a quartz-fiber suspension in such a way that the magnetic field of the current causes a torque on it. Such an arrangement is practical for measuring currents in the frequency range from about 1 to 1000 MHz.

The torque phenomenon between the movable coil and the field created by the alternating current was first described by Thomson and Fleming in 1887 [15], [16]. Pierce used such a device to measure "feeble antenna currents" resulting from nearby radio-frequency wireless transmissions [17], [18]. Pierce's electrodynamometer consisted of a thin disk of silver backing on a plane-glass galvanometer mirror which was suspended by a fine quartz fiber in a tube. The "field" current was provided by a small coil consisting of about 30 turns of insulated wire. It provided rapid and consistent detection of current but could not be used as a detector for wireless signals coming from a great distance.

In the late 1930's an electrodynamic ammeter was described by Turner and Michel for measurement of currents from 1 to 5 amperes and for frequencies from 1 to 100 MHz [19], [20]. This ammeter can be thought of as an air-core transformer with the current to be measured passing through a fixed single-turn primary which induces a current in a short-circuited-ring secondary. The secondary was suspended by a quartz fiber so that it could swing about a vertical axis in response to the interaction of the magnetic fields. To avoid disturbances from air currents, the shorted ring and suspension were placed in a sealed air-filled glass tube. The single-turn primary was not placed in the tube because heat generated from current in the primary disturbed the suspension.

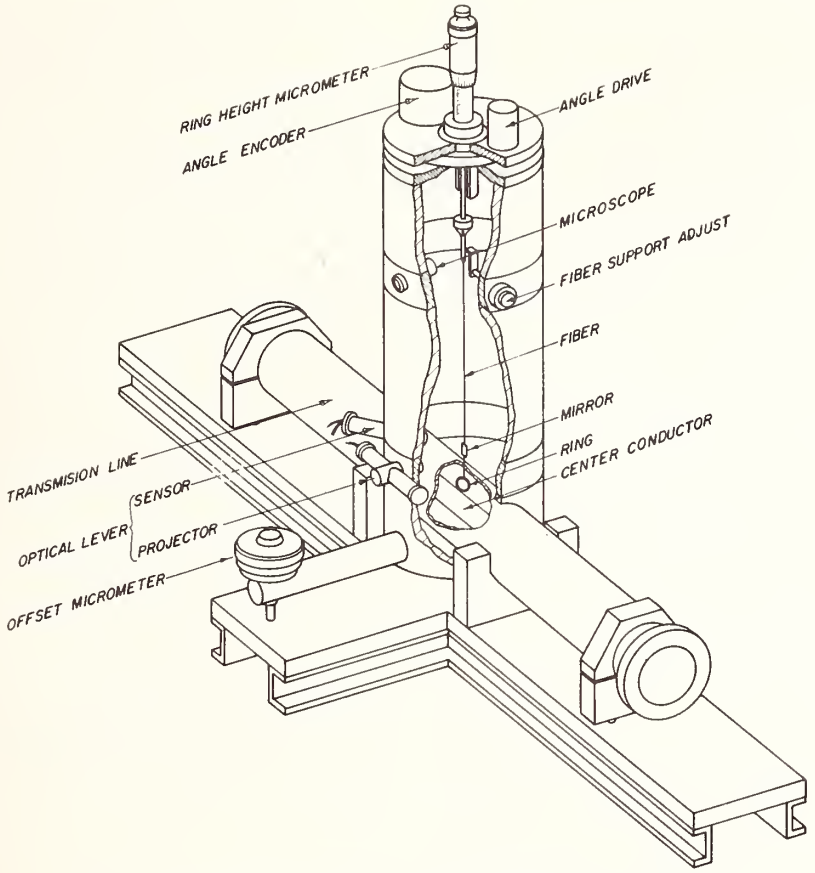


Fig. 2. Cut-away drawing of NBS short-circuited-ring electrodynamometer.

Results reported by Turner and Michel [19], [20] in comparing their electrodynamic ammeter and a thermocouple ammeter at about 5 amperes showed good agreement for frequencies up to 5 MHz. At 10 MHz the thermocouple ammeter read higher by 4 percent and at 80 MHz, by 80 percent. Most of this difference was attributed to skin effect in the thermocouple ammeter (50-percent error at 80 MHz). The remainder was attributed to the effect of parallel resonance.

Another electrodynamic ammeter was built by Meahl [21] which used jewel bearings to obtain a more sturdy instrument and to simplify the technique of its use.

In the 1950's electrodynamic ammeters were designed which allowed insertion of the shorted ring into an air coaxial transmission line. Solow [23] reported in 1950 on a theoretical study at the National Bureau of Standards of such an electrodynamic ammeter, but the construction of a successful model was not completed. Meahl and Allen [22] built their ammeter so that an instrument being calibrated terminated the coaxial transmission line. Their ammeter had provision for moving the ring along the transmission line so that measurements of current could be made at three or more positions. This provision was intended for use at frequencies above 300 MHz where the line length was an appreciable part of a wavelength. If the observed values of current distribution along the line showed good agreement with the calculated values, then a correction could be applied for the transmission-line position of the ammeter under calibration. Stated calibration uncertainty was within one percent for currents 3 to 10 amperes over the frequency range of 1 to 350 MHz.

In 1963, results were announced of the development in Japan of a short-circuited-ring type of electrodynamic ammeter [24], with a claimed uncertainty of about 0.7 percent under optimum conditions. The useful frequency range claimed for this model is 100 kHz to 1.5 GHz. The current range is from a few tens of milliamperes to a few hundred amperes.

In a new attempt to establish an independent radio-frequency standard ammeter, the Radio Standards Laboratory of NBS recently constructed a short-circuited-ring electrodynamometer [1]. Figure 2 is a cut-away drawing showing the principal elements of the ammeter. The vertical quartz-fiber suspension is about 20 cm long and supports both a plane-surface mirror and a copper ring.

In general, the torque acting on a thin, perfectly conducting, shorted ring near a current-carrying conductor can be relatively easily calculated if the electromagnetic fields are undisturbed by conductor or dielectric boundaries in the neighborhood of the ring [25]. This "free-space" or idealized condition has been the basis for calculations of the sensitivity of previously mentioned short-circuited-ring

electrodynamometers. Research is presently being carried out at the Radio Standards Laboratory in an effort to determine the effects of the boundary conditions imposed by the coaxial transmission-line conductors on the current-torque relation. An independent experimental check on this research is expected to be made soon, using a method originated by Cullen [26]. Cullen's method relates the torque on the short-circuited ring to the force on a piston at the end of the transmission line when the transmission-line ring assembly is connected as a cavity resonator. The force on the piston can be easily calculated, whereas precise calculation of the torque on the ring is far more difficult.

A disadvantage of the shorted-ring electrodynamic ammeter is its sensitivity to mechanical vibration. Mechanical vibration disturbs the torsion-balance system, but by good design in the mechanical support for the ammeter, these disturbances can be minimized sufficiently.

Other references of interest are given which involve different physical forms of an electrodynamometer or deal with topics involved in the operating principles [27]–[32], [41].

V. Photoammeters

The photoammeter usually consists of a straight-filament tungsten lamp and a photocell arranged to detect the emitted light when current is passed through the filament. Comparison can then be made between direct current and radio-frequency current when the current is of sufficient magnitude to cause incandescence. A tungsten filament is used because a very small-diameter filament is capable of carrying a large current. For example, at 100 MHz, a 5-ampere thermocouple ammeter may typically have an RF-to-dc resistance ratio of 2.57, while a tungsten filament lamp may have a similar ratio of only 1.065 [6]. This method needs little auxiliary apparatus near the heater and therefore does not tend to disturb the RF current distribution. The photoammeter has high precision because the output of the photocell is approximately proportional to the sixth power of the filament current. The high sensitivity does greatly limit the operating current range, however, because the useful range of a lamp is from its maximum current capacity to about 65 percent of that capacity. Smaller currents do not adequately operate the photocell. Usually two or more lamps are used to calibrate a single-range thermocouple instrument.

Currents from 50 milliamperes to 20 amperes may be measured by a photoammeter of the proper range. Wallace and Moore [7] used single-filament lamps below 3 amperes and multiple parallel filaments for higher ranges. They felt that the largest source of uncertainty was stray capacitance. Uncertainty in calibrations was about 5 percent at 15 MHz and within 12 percent at 100 MHz. Precision with this method was stated to be 3 percent at the highest test frequency and 1 percent at the lower frequencies.

A high-current photoammeter for use between 0.1 and 10 MHz was reported by Hoffmann and Weber [33]. They used a small platinum tube with very thin walls for the filament. The current range is 10 to 100 amperes. Each measurement takes 10 to 15 minutes.

In 1948, Miller [13] compared a number of 5-ampere thermocouple ammeters, using photoammeters as standards. Results showed the instruments read high and a detailed study showed that skin effect accounted for the total uncertainty and that there were no other factors producing errors of any appreciable magnitude.

Stanek (as reported by Jones [11]), reduced the impedance of the heater both by decreasing the length and increasing the cross section of the filament and by substituting tubes for solid conductors. He increased the current capability by operating the filament in air instead of vacuum and by using platinum instead of tungsten.

The use of a photoammeter to calibrate thermocouple instruments is controversial. In 1953 Meahl and Allen [22] argued that the close agreement obtained between photoammeters and thermocouple ammeters is due to the fact that both types operate from the same basic principle and are therefore subject to the same sources of error. They found errors as large as 60 percent at 5 amperes and 200 MHz, using (apparently) a shorted-ring electrodynamometer. The photoammeter has the additional disadvantage of using more power than other methods. This power is required to make the filament incandescent.

The photoammeter is particularly vulnerable at radio frequencies because of its small overload capacity. For example, if a resonance condition occurs, the sudden rise in current may cause the photoammeter to burn out.

VI. Air Thermometer Milliammeters

One satisfactory standard for calibrating low-current-range instruments is the air thermometer milliameter. Advantages of the design are high sensitivity and relatively good isolation between the radio frequency and sensing circuits.

An air thermometer milliammeter (see Fig. 3) consists of two air-filled bulbs connected by a capillary tube, which contains a small bubble of alcohol or ether. Each bulb is sealed to the atmosphere and contains a heater. The index is initially located by passing direct current through both heaters in series and observing the bubble location. When one heater is carrying RF current and the other heater direct current, the bubble indicates the balance condition of the RF-dc "air" bridge. A microscope is used to increase the precision in locating the index. Schmitz [34] used this construction to measure currents of 1 and 10 mA with an uncertainty of less than 5 percent. A similar differential air-milliammeter was used by Takaya [14] to calibrate hot-wire and thermocouple instruments from 10 to 1000 mA to a frequency of 10 MHz. Repeatability at dc was within 0.6 percent. Disadvantages of the air thermometer milliammeter arise because the heater has to be relatively long and at high frequencies, if standing waves exist, a significant amount of energy will be radiated, creating measurement uncertainty. Another model reported by Gainsborough [12] reproduced indications at 10 mA to within 0.1 percent. The sensitivity of the milliammeter was altered by using capillaries of different bore and heaters of different resistance.

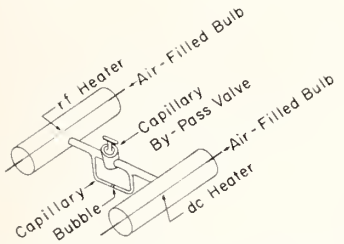


Fig. 3. Typical construction of air thermometer milliammeter.

Comparisons between the air thermometer and thermocouple milliammeters were reported at frequencies up to 700 MHz where they agreed with a low-frequency calibration within the experimental uncertainty, which was about one percent of the maximum current. The principal problem at 700 MHz was the difficulty, because of standing waves, in obtaining the same current through both ammeters. This requires knowledge of their precise location. Further, if standing waves do exist in the ammeters, it is questionable if meaning can be attached to the ammeter indications because the current is not uniformly distributed in the heaters.

A similar instrument described by Strutt and Knol [35] featured heaters constructed of constantan (20 microns in diameter) in order to reduce skin-effect errors to less than three percent up to 1.5 GHz. With the RF-dc arrangement previously described, they believed currents of a few milliamperes could be measured with an uncertainty of less than about one percent.

The development of the air thermometer milliammeter has mainly occurred in England and Germany. Because of its advantages it deserves more consideration.

VII. Circuits and Miscellaneous Devices for Measuring Current

Besides the devices described in the previous sections, there are several circuits which could be used to measure radio-frequency currents. One circuit suggested by Miller [13] uses a one-ampere thermocouple in conjunction with a loop inserted through the outer conductor of a coaxial line. The coupling of the loop to the magnetic field can be adjusted. A serious difficulty, however, is that the loop couples to both magnetic and electric fields and unless standing waves exist and the loop is located at a voltage null, or unless the loop is electrically shielded, an uncertainty due to the electric field will occur [36].

A calorimeter method, which is particularly suited for measurement of currents from 50 to 200 mA, is described by Meahl and Allen [22]. Their method describes essentially a "feed-through" wattmeter in which relatively little power is consumed as compared with an "absorption" wattmeter, where all the power is consumed. They use a coaxial system in which a section of the center conductor is removed and replaced with a low-valued, thin-film resistor. The resistor is low-valued (30 ohms) compared to the characteristic impedance of the coaxial line (149 ohms) and at maximum current dissipates only 120 mW. The resistor is enclosed in an insulated cylinder which also contains a thermal junction. As RF current is passed through the resistor, a temperature rise occurs within the insulation. The temperature stabilizes when the heat loss equals the heat generated. This temperature is indicated by the dc output of the thermal junction. If the same temperature is obtained with direct as with RF current, then the two currents are effectively equal within the correction and uncertainty for the method [4], [5]. A disadvantage of this method is that the leads are customarily brought through the center conductor of a one-quarter-wavelength short-circuited isolation stub to avoid error due to shunt impedance. This makes the calorimeter very frequency-sensitive. The calorimeter method is reported to be useful to about 1 GHz. If the resistor is made to equal the characteristic impedance (Z_0), and is terminated in a short circuit, then the calorimeter can be used as an absorption wattmeter. The calorimeter method requires a very stable generator which provides steady RF current over the time interval necessary to obtain temperature equilibrium.

Another method, which uses a dc bridge with thermistors heated by radio-frequency current, has application in the calibration of thermocouples from about 5 to 15 mA over the frequency range from about 100 MHz to 1 GHz [22]. Again this instrument can be used as an RF wattmeter where, if the impedance Z_0 is known, the radio-frequency current can be calculated.

Hot-wire expansion ammeters were very popular for a few decades around the turn of the century [37]. The hot-wire ammeter uses a mechanical linkage to amplify thermal linear expansion caused by radio-frequency current in a wire, but is difficult to compensate for errors due to ambient temperature variations. Also, the expansion element is easily damaged by a moderate overload. Ammeters using electrical methods of detecting temperature changes [38] are now preferred over instruments of this type.

Another device for determining current makes use of a thermionic vacuum tube. The input circuit uses a filament which is heated first by low-frequency currents and then by radio-frequency currents. Electrons emitted by the heated filament are drawn to an anode sealed into the tube. Comparison of known low-frequency current and unknown radio-frequency currents can be made by equating the space currents [39].

Some other devices use rectifiers, transformers, and shunts but have serious inherent disadvantages, especially at the higher radio frequencies, which preclude their general use [11], [40].

VIII. Conclusion

The principal standards which are used in the measurement of radio-frequency currents have been described. But what is the present state-of-the-art? Ignoring the dangers of generalization for the advantages of brevity, radio-frequency currents can probably be measured within one percent over a current range of 5 milliamperes to 10 amperes and over a frequency range of 100 kHz to 1 GHz.

Challenges do exist for significant improvements in the measurement of RF currents. In the Introduction, it was

385-582 O - 70 - 6

mentioned that relatively little interest was shown in RF current measurements in the years following World War II as compared with the years preceding. A search of the literature reveals this fact. The decline of interest in RF current measurements probably came about when the attention of metrologists in the radio-frequency field was drawn to the more immediately fruitful areas of power, attenuation, and impedance, resulting in few significant advances (with the exception of the electrodynamometer) in the development of improved current standards over the past 30 years. Recent developments of stabilized RF power generators, precision connectors and lines, and thin-film resistors (to mention but a few items) should make improvements in the measurement of current at radio frequencies possible with modern electronics in the hands of a knowledgeable, interested metrologist.

References

[1] N. V. Frederick, "An absolute electrodynamometer for high frequency currents," to be published.*

[2] H. N. Edwardes, "The measurement of current distribution in aerials at VHF," *Proc. IRE* (*Australia*), vol. 24, pp. 351–355, April 1963.

[3] W. N. Goodwin, Jr., "The compensated thermocouple ammeter," *Elec. Engrg.*, vol. 55, pp. 23–33, January 1936. (Discussion and closure p. 407.)

[4] "Radio instruments and measurements," NBS Circular 74, 1924.

[5] W. E. Forsythe, "Smithsonian physical tables," 9th rev. ed., Smithsonian Institution, Washington, D. C., 1954.

[6] J. H. Miller, "Thermocouple ammeters for ultra-high frequencies," *Proc. IRE*, vol. 24, pp. 1567–1572, December 1936.

[7] J. D. Wallace and A. H. Moore, "Frequency errors in radio-frequency ammeters," *Proc. IRE*, vol. 25, pp. 327–339, March 1937.

[8] O. G. McAninch, "Thermocouple-type ammeters for use at very high frequencies," *Elec. Engrg.*, vol. 73, pp. 431–435, May 1954.

[9] C. L. Fortescue and L. A. Moxon, "A method of comparing ammeters at very high frequencies," *J. Sci. Instr.*, vol. 8, pp. 94–97, 1931.

[10] "New type 493 vacuum thermocouples for use at high frequencies," *Gen. Radio Exper.*, vol. 8, March 1939.

[11] T. I. Jones, "Measurement of current at radio frequencies," *Reports on Progress in Physics*. London: Cambridge University Press, 1938.

[12] G. F. Gainsborough, "Experiments with thermocouple milliammeters at very high radio frequencies," *J. IEE* (*London*), vol. 91, pt. 3, pp. 156–161, 1944.

[13] J. H. Miller, "Thermocouple ammeters for very high frequency," *Weston Engrg. Notes*, vol. 1, pp. 7–8, April 1946.

[14] M. Takaya, "Characteristics of thermoammeters for use in radio frequencies," *Res. Electrotech. Lab.* (*Tokyo*), no. 329, March 1932.

[15] E. Thomson, "Novel phenomena of alternating current," *The Electrician*, vol. 19, pp. 546–547, April 29, 1887.

[16] J. A. Fleming, "A copper disc galvanometer for alternating current," *The Electrician*, vol. 18, p. 561, May 6, 1887.

[17] G. W. Pierce, "Experiments on resonance in wireless telegraph circuits—part 1," *Phys. Rev.*, vol. 19, pp. 196–217, September 1904.

[18] ——, "Experiments on resonance in wireless telegraph circuits—part 2," *Phys. Rev.*, vol. 20, pp. 220–251, April 1905.

[19] H. M. Turner and P. C. Michel, "An electrodynamic ammeter for use at frequencies from one to one hundred megacycles," *Proc. IRE*, vol. 25, pp. 1367–1374, November 1937.

[20] P. C. Michel, "A high frequency electrodynamic ammeter," Ph.D. thesis, Yale University, New Haven, Conn., 1935, unpublished.

[21] H. R. Meahl, "A bearing-type high-frequency electrodynamic ammeter," *Proc. IRE*, vol. 26, pp. 734–744, June 1938.

[22] H. R. Meahl and C. C. Allen, "Calibrating ammeters above 100 Mc," *Proc. IRE*, vol. 41, pp. 152–159, January 1953.

[23] M. Solow, "Theoretical study of an electrodynamic ammeter for very high frequencies," NBS Central Radio Propagation Lab. Preprint 50–15, January 1950.

[24] Y. Aida, "Research on the primary standard of current using short-circuited ring for high frequencies," *Res. Electrotech. Lab.* (*Tokyo*), no. 639, 1963.

[25] W. R. Smythe, *Static and Dynamic Electricity*. New York: McGraw-Hill, 1950, pp. 317–318.

[26] A. L. Cullen, "A general method for the absolute measurement of microwave power," *J. IEE* (*London*), vol. 99, pt. 4, no. 24, pp. 112–120, 1952.

[27] E. B. Moullin, "The development of a precision ammeter for very high frequencies," *J. IEE* (*London*), vol. 68, no. 396, pp. 544–555, 1930.

[28] "Electrodynamic ammeter for very high frequencies," *NBS Tech. News Bull.*, pp. 103–104, July 1950.

[29] E. B. Moullin, "An ampere meter for measuring alternating currents of very high frequency," *Proc. Roy. Soc.* (*London*), ser. A, vol. 121, pp. 41–71, December 1928.

[30] P. K. Taylor, "The action of a high-frequency alternating magnetic field on suspended metallic rings and disks," *Proc. IRE*, vol. 22, pp. 886–896, July 1934.

[31] H. R. Meahl, P. C. Michel, M. W. Scheldork, and T. M. Dickinson, "Measurements at radio frequencies," *Elec. Engrg.*, vol. 59, pp. 654–659, December 1940.

[32] P. G. Agnew, "A tubular electrodynamometer for heavy currents," *NBS Bull.*, vol. 8, pp. 651–658, 1912.

[33] E. Hoffmann and K. H. R. Weber, "On measuring high currents in the frequency range of 0.1 to 10 MHz," *Tech. Mitt. RFZ* (*Germany*), vol. 6, pp. 129–132, September 1962.

[34] W. Schmitz, "Differential air thermometer for ac current measurement," *Z. Hochfrequenztech.*, vol. 27, pp. 18–19, 1926.

[35] M. J. O. Strutt and K. S. Knol, "Measurements of currents and voltages down to a wavelength of 20 centimeters," *Proc. IRE*, vol. 27, pp. 783–789, December 1939.

[36] C. M. Allred and R. A. Lawton, "A precision current comparator," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-16, June 1967.

[37] F. K. Harris, *Electrical Measurements*. New York: Wiley, 1952, pp. 429–430.

[38] C. L. Fortescue, "The theory and design of hot-wire ammeters for frequencies of 25 to 100 megacycles," *J. IEE* (*London*), vol. 79, pp. 179–193, August 1936. (See also discussions p. 200.)

[39] H. C. Hazel, "A new method for the calibration of ammeters at radio frequencies," *Proc. IRE*, vol. 16, pp. 70–74, January 1928.

[40] A. Nyman, "Condenser shunt for measurement of high-frequency currents of large magnitude," *Proc. IRE*, vol. 16, pp. 208–217, February 1928.

[41] V. L. Lopan, "Electrodynamic ammeter for measuring high-frequency currents," *Meas. Techniques* (English transl. of *Izmeritel'naya Tekhnika*), no. 1, pp. 85–90, 1958.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967
pp. 886-891
The Institute of Electrical and Electronics Engineers, Inc.

---

* [1] N. V. Frederick, "A new high frequency current standard," IEEE Trans. Instr. & Meas. IM-17, December 1968.

Reprinted from IEEE TRANSACTIONS
ON *INSTRUMENTATION AND MEASUREMENT*
Volume IM-16, Number 2, June, 1967
pp. 142-145
THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

A Precision Current Comparator

C. McKAY ALLRED, MEMBER, IEEE, AND ROBERT A. LAWTON

Abstract—A technique capable of precise comparison of currents in different parts of a network or between currents in different networks is discussed. With proper conditions, this comparison can be done with negligible perturbation of the networks under measurement. Experiments on a coaxial system at 30 MHz are discussed.

INTRODUCTION

THE USUAL method of obtaining a measure of the current in a circuit is to measure the voltage across an impedance inserted in the circuit.[1] This includes, in general, coupled devices such as clamp-on probes.[2] Unless the impedance is an inherent part of the circuit, undue alteration of the circuit can occur. The technique described herein, to the authors' knowledge, is novel and permits the precision comparison of currents at some particular point in a network to some other current, either at the same point or at some other point at another time. Under proper conditions negligible disturbance of the network occurs.

THEORY

Consider the arbitrary network in Fig. 1, termed Network A. Here, the circuit with a particular current I has been separated from the rest of the network for discussion purposes. By Thévenin's theorem we may replace the above network by its equivalent as shown in Fig. 2. Let us now open the circuit and connect it to a somewhat arbitrary 3-terminal pair (3-port) linear network, as indicated in Fig. 3.

As indicated in the figure, the "opened" network is connected to terminal pair 3; a sensitive null detector is connected to terminal pair 2, while a source supplying a voltage E_1 is connected to terminal pair 1. This latter source must be coherent with the source giving rise to E_{eq}. The relationship, in matrix form, between the voltages and currents at the terminal-pairs of the 3-port is

$$\begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{21} & Z_{22} & Z_{23} \\ Z_{31} & Z_{32} & Z_{33} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix}. \tag{1}$$

With terminal pair 3 shorted, certain 3-port networks can be constructed so that a null exists at terminal pair 2. There are numerous networks with this characteristic. For example, any immittance bridge is such a network. If a null exists at terminal pair 2 with terminal pair 3 shorted, the Z matrix elements are related by

$$\frac{Z_{23}}{Z_{21}} = \frac{Z_{33}}{Z_{31}}. \tag{2}$$

This is the only restriction on the 3-port network other than it be linear and stable. Reciprocity need not hold. Hereafter, the 3-port is used without a short at terminal pair 3.

Manuscript received September 29, 1966; revised January 18, 1967.

The authors are with the National Bureau of Standards, Boulder, Colo.

[1] P. N. Miljanic, N. L. Kusters, and W. J. Moore, "The development of the current comparator, a high-accuracy ac ratio measuring device," *AIEE Trans.* (*Communications and Electronics*), pp. 359–368, November 1962.

[2] Charles O. Forge, "A new clip-on oscilloscope.voltmeter probe for 25 cps to 20 Mc current measurements," *Hewlett-Packard J.*, vol. 11, nos. 11 and 12, July–August 1960.

It will now be shown that for any value of E_{eq} and Z_{eq}, if E_1 is adjusted in phase and amplitude so as to produce a null at the detector, the equivalent generator of Network A looking into terminal-pair 3 will see a a short circuit; i.e., Network A has not been perturbed.

With Network A connected to the 3-port,

$$I = I_3 \tag{3}$$

and

$$E_3 = E_{eq} - Z_{eq}I_3. \tag{4}$$

Using (1), (4), and the null condition

$$E_2 = 0 \tag{5}$$

$$I_2 = 0, \tag{6}$$

one obtains

$$E_{eq} - Z_{eq}I_3 = \left(-\frac{Z_{23}Z_{31}}{Z_{21}} + Z_{33}\right)I_3. \tag{7}$$

But (2) makes the right-hand term zero, and hence

$$E_{eq} = Z_{eq}I. \tag{8}$$

This means the 3-port presents a short circuit at terminal-pair 3 and therefore does not perturb or affect Network A in which it has been inserted..

Let I_a and I_b represent the current I at two different times. The currents could be at the same point or at different points in a network.

If now E_1 is always adjusted in phase and amplitude such that a null exists no matter what circuit terminal-pair 3 is inserted in, the ratio of the currents at the point of insertion, I_a and I_b, is equal to the ratio of the corresponding voltages at terminal-pair 1, E_{1a} and E_{1b}. That is,

$$\frac{I_a}{I_b} = \frac{E_{1a}}{E_{1b}}. \tag{9}$$

This equation is rather obvious from (8) and the linear relationship between I and E_1 involving only the invariant elements of the 3-port. This is the desired relationship which expresses the current ratios in external circuits in terms of voltage ratios at terminal-pair 1. Because (8) is valid for null conditions, the impedance looking into terminal-pair 1 of the 3-port is constant when a null is achieved.

Thus, precision phase shifters and attenuators are natural devices to use for the measurement of voltage ratios at terminal-pair 1. A possible overall system is suggested in Fig. 4. If A and ϕ are the values indicated by the attenuator and phase shifter respectively,

$$\frac{E_{1a}}{E_{1b}} = \frac{A_a}{A_b}e^{j(\phi a - \phi b)} = \frac{I_a}{I_b}. \tag{10}$$

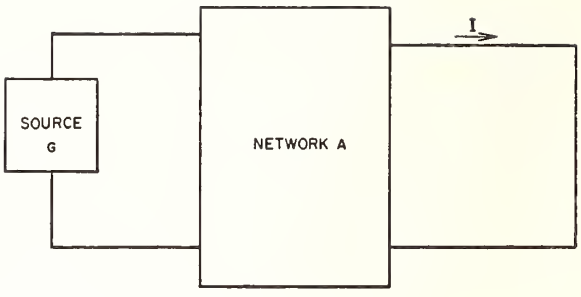


Fig. 1. Network A.

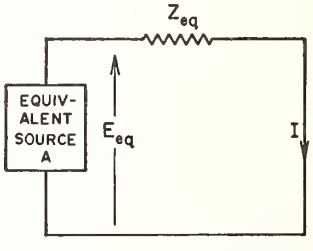


Fig. 2. The equivalent of network A.

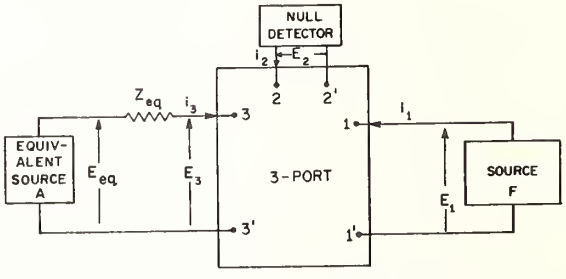


Fig. 3. Network A combined with a 3-port.

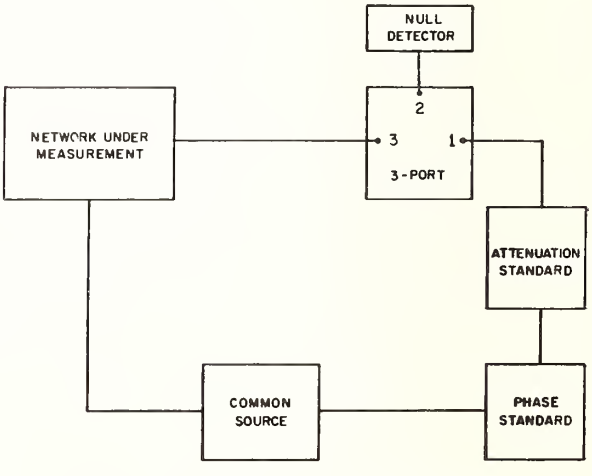


Fig. 4. Complete current comparison system.

An interesting extension to the above theory can be made by replacing the short circuit by an impedance Z, while making the original adjustments to the 3-port for the null condition. It may be shown that terminal pair 3 then presents an impedance of value $-Z$ at its terminals.

At sufficiently high frequencies, care must be exercised to see that the short used in adjusting the 3-port is at the precise location where the currents are to be compared. The dimensions of the short should be small compared to the wavelength. Also, coupling between other parts of the 3-port and the network under measurement must be avoided.

Errors

As already indicated, if the impedance ΔZ of the "short" is small but finite, then an impedance equal to $-\Delta Z$ is inserted into the network under measurement. (This is true in general and holds for any type of network used as the 3-port.) The effect of this small impedance depends on the network under consideration. A high-Q circuit at resonance could be appreciably affected. There are additional imperfections due to imperfect nulling since the detector always has finite sensitivity. Errors may occur in two cases: a) in adjusting the bridge to meet the condition expressed by (2), and b) in achieving a null when the 3-port is inserted in a network. The effect, in case a), is similar to the one caused by not using a perfect short, because there exists an impedance ΔZ which actually produces a null at the detector. The magnitude of the effect depends on the nature of the 3-port. Consider, for simplicity, the bridge shown in Fig. 5, and let ΔE_2 be the minimum detectable voltage resulting from insensitivity in the detector. Also, assume the input impedance to the detector is R. It can be shown that if $\Delta R \ll R$ (terminal-pair 3 is shorted)

$$\frac{\Delta R}{R} = 8\,\frac{\Delta E_2}{E_1}. \tag{11}$$

For example, if

$$\Delta E_2 = 10^{-6} \text{ volts}$$
$$E_1 = 1 \text{ volt}$$
$$R = 10 \text{ ohms},$$

then

$$\Delta R = 80 \times 10^{-6} \text{ ohms}.$$

Thus, because of imperfect sensitivity, the bridge may possibly be unbalanced by the amount ΔR. The same condition would have occurred had there been perfect sensitivity and an impedance $-\Delta R$ placed across terminal-pair 3 instead of the "short." Hence, if terminal-pair 3 is inserted into a network and a perfect null is now obtained, the network under measurement will see an impedance ΔR inserted in it rather than a "short."

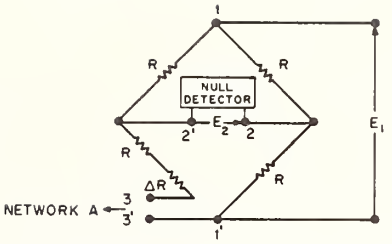


Fig. 5. Bridge unbalance due to detector insensitivity.

In case b), the problem of detector sensitivity increases as the current being compared becomes smaller and smaller, since E_1 must correspondingly decrease while ΔE_2 remains fixed. Again, assume the 3-port is that of Fig. 5, the short is perfect, and the bridge has been perfectly adjusted, i.e., ΔR of case a) is zero. Network A is now connected to terminal pair 3 and E_1 is adjusted until a voltage ΔE_2 appears at the detector. Analysis shows that

$$E_{eq} = I\left(Z_{eq} + \frac{4\Delta E_2}{I}\right). \tag{12}$$

Hence, an equivalent impedance equal to $4\Delta E_2/I$ appears inserted into the network under measurement. For $\Delta E_2 = 10^{-6}$ volts and $I = 10^{-3}$ amperes, this equivalent impedance is 4×10^{-3} ohms. However, since coherent signals are being considered, coherent detectors can be used with much greater sensitivities than indicated here. For example, the authors have constructed a detector at 30 MHz with a sensitivity of 40×10^{-12} volts when using an integration time of 30 seconds.[3]

Experimental Verification

An experiment was conducted at 30 MHz in order to verify some of the aspects of the theory. Since the system is a linear one, the current voltage relationship given by (9) is rather obvious. Hence, it was felt that the emphasis on the experimental verification should be placed on the impedance perturbation problem as this aspect of current measurements is often very critical. To this end a special 50-ohm coaxial termination was constructed, so that the impedance of a normal termination could be compared with one in which the 3-port network had been inserted. This special termination had a replaceable center conductor. One center conductor had effectively a normal solid configuration while two others contained gaps of 0.001 inch and 0.01 inch, respectively. The gaps were effectively the unshorted third terminal pair of the 3-port, while the normal center conductor, which maintained the internal configuration as shown in Fig. 6, served as a short for the third terminal-pair.

[3] C. M. Allred and R. A. Lawton, "Precision detector for complex insertion ratio measuring systems," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-13, pp. 76–81, June–September 1964.

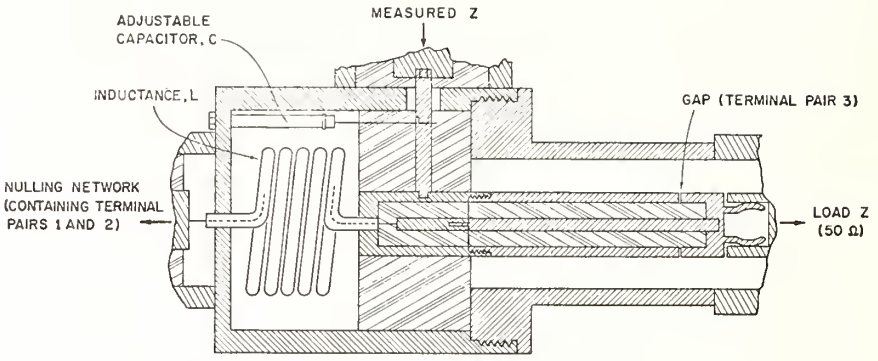


Fig. 6. Coaxial current comparison network.

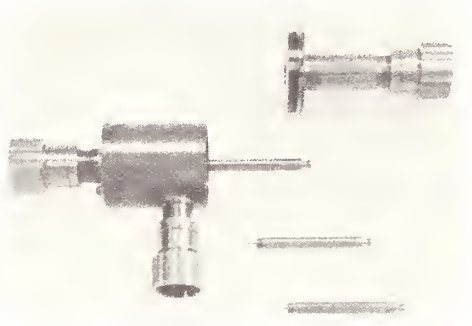


Fig. 7. Coaxial network with interchangeable center conductors.

The parallel-tuned circuit in Fig. 6 consists of an inductance L, which is the outer conductor of a small coaxial line, and a capacitance C. This circuit, together with the stray inductance and capacitance, is resonant at 30 MHz, which enables one to connect to the internal transmission circuit of the center conductor of the regular coaxial system without shorting it. The unit, with the interchangeable center conductors of different gap widths, is shown in Fig. 7.

With the normal center conductor (without a gap) in use, the impedance looking into the unit was measured. The nulling network (which was actually an immediately available impedance bridge) was then adjusted for the condition described by (2). The center conductor with a gap of 0.001 inch was then inserted, and the impedance looking into the unit was again measured. The impedance measurements were the same within the sensitivity of the measuring device which was somewhat better than 0.2 percent. The tests were repeated for the center conductor having a 0.01-inch gap with the same results. Instead of the "shorting" system used, it would probably be safer to use a center conductor with a gap and a tight-fitting ring to slip over the center conductor in order to "short" the gap. This would help to maintain the same 3-port network when nulling for condition (2) as when measuring a current through terminal pair 3.

Conclusions

A technique of comparing currents has been shown which can be done with precision and without undue reaction with the circuit under measurement.

Such a technique has a number of uses. For example, in a coaxial system, an arrangement such as indicated in Fig. 6 could be used to compare currents over wide amplitude levels to a known standard current at an optimum level. The smallness of the gap would allow the measurement to be placed extremely close to the desired point. Another example is the measurement of the relative amplitude and phase distribution of the current in a "thick" circular cylindrical antenna above a ground plane. (In fact, it was this problem that gave rise to the technique under discussion.) A section of the antenna could be made similar to the gapped center conductor of Fig. 6. This section could then be moved to different parts of the antenna to obtain the current distribution.

Abstract of Related Paper

2.a. URSI National Committee Report, XIV General Assembly, Tokyo, September 1963: Commission 1. Radio measurement methods and standards. 4. Pulsed and CW sinusoidal voltage and current measurements, M. C. Selby, Radio Sci. J. Res. 68D, No. 5, 533-536 (May 1964).

See also 10.d.

3. Electromagnetic Fields and Antennas

Papers

Abstracts

See also 7.e.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967
pp. 981-990
THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

Field Strength Above 1 GHz: Measurement Procedures for Standard Antennas

RONALD R. BOWMAN

Abstract—**To calibrate antennas for state-of-the-art field-strength measurements above 1 GHz, standard antennas are needed that have gain values known to within ±0.1 dB. Since this requirement exceeds the verified accuracy of calculated gain values, these standards must be established by making absolute gain measurements. The discussion primarily concerns absolute gain measurements for horn antennas by the two-antenna method. However, much of the discussion is pertinent to high-accuracy field-strength measurements in general. The two-antenna method is considered to be essentially an insertion-loss measurement (with many additional problems and sources of error), and this concept is used to derive a working formula that is suitable for high-accuracy gain measurements. The two most intractable problems—insufficient antenna separation and multipath interference—are discussed in detail. Some important experimental details are included that have previously been overlooked or inadequately discussed, and it is concluded that previous error estimates of less than ±0.1 dB for horn-gain measurements have been somewhat optimistic. To facilitate the design and evaluation of high-accuracy gain measurements, some simple terms, concepts, and formulas are provided that are useful in analyzing multipath interference.**

Manuscript received March 8, 1967.
The author is with the National Bureau of Standards, Boulder, Colo.

I. INTRODUCTION

A. Field-Strength Measurements Above 1 GHz: A Review of Some Basic Definitions and Formulas

WITH reference to field-strength measurements, the terms "above 1 GHz" and "below 1 GHz" usually denote a division between the use of antennas larger than a wavelength and antennas smaller than a wavelength.[1,2] Though this division is rather arbitray as regards frequency,[3] it is logical as regards antenna "size" since this factor is of primary importance in de-

[1] Field-strength measurements and standards below 1 GHz are discussed in a paper by F. M. Greene in this issue.

[2] For a general survey of radiation measurements and techniques, see Cumming [5].

[3] For absolute field-strength measurements, at least, this choice of frequency is not as arbitrary as it may seem. Even though the use of electrically small antennas at frequencies above 1 GHz is fairly common, this practice is mainly limited to measurements of relative field strength—as when investigating the structure of the field near a large antenna.

termining both the techniques and the terminology used for a field-strength measurement. With reference to electrically large antennas, the term "field strength" usually refers to the power density of the electromagnetic field.

The power density at a field point is given by the magnitude of the time-averaged Poynting vector S. For free space,

$$S = \frac{1}{2}\left(\frac{\mu_0}{\varepsilon_0}\right)^{-\frac{1}{2}} E^2$$

where E is the peak value of the electric component of the field and $(\mu_0/\varepsilon_0)^{\frac{1}{2}}$ is the characteristic impedance of free space. If "S" or "power density" are used without qualification, it is usually understood that they refer to a field that does not differ significantly, in the region of interest, from a plane-wave field with uniform power density. The receiving characteristics of an antenna are specified in terms of the power P_R delivered to a conjugately matched load when the antenna is placed in a uniform, plane-wave field. Under free-space conditions, and with the polarization parameters of the antenna and field conjugately matched ([44]; [57], p. 442), the effective receiving area (or absorption cross section) A is defined by

$$P_R = SA. \tag{1}$$

So defined, A is a function only of the direction angles between the axis of the antenna and the propagation direction of the field. If "A" or "receiving area" are used without qualification, it is usually understood that they refer to the maximum value of this function.

The transmitting characteristics of an antenna are usually specified by the power gain function G, which is essentially a performance rating of the antenna as compared to an isotropic, point-source, lossless radiator. If P_Ω is the power per unit solid angle radiated by an antenna as a function of direction when P_T is delivered to (accepted by) the antenna, then

$$G = \frac{P_\Omega}{P_T/4\pi}. \tag{2}$$

Except for the hypothetical case of a point-source antenna, G is also a function of the distance to the field point. If "G" or "gain" are used without qualification, it is usually understood that they refer to the maximum value of the power gain function—that is, in the direction of maximum radiation intensity, the asymptotic value that is approached with increasing distance. The region of space in which G does not differ significantly from its asymptotic values is called the far-field region. With reference to the near-field region, "near-field gain" and "G_N" are commonly used for the power gain function; and the functional dependence on distance is implied when it is not stated explicitly.

The power density in the field radiated into free space by a transmitting antenna with power gain function $G = G_T$ is given by

$$S = \frac{P_\Omega}{r^2} = \frac{G_T P_T}{4\pi r^2} \tag{3}$$

where r is the distance to the field point. If a receiver is located in this field at any point where the field parameters are essentially constant over the receiving aperture and where the separation distance is great enough so that multiple scattering between the transmitting and receiving antennas is negligible, then, from (1) and (3), the received power is given by

$$P_R = \frac{G_T P_T}{4\pi r^2} A_R \tag{4}$$

where $A = A_R$ is the effective receiving area function of the receiving antenna. It is worth noting that this formula is valid for near-field values of G_T if the receiving antenna is sufficiently small.

The effective receiving area function and the far-field power gain function of an antenna are related by

$$G = \frac{4\pi}{\lambda^2} A, \tag{5}$$

provided that nonreciprocal elements are excluded from the antenna.

The performance of an antenna is usually adequately specified if the gain value and polarization parameters are accurately known for the direction of maximum radiation and approximately known for the other directions.

B. Gain Measurements Above 1 GHz: State-of-the-Art Requirements for Standard Antennas

In practice, the gain of an antenna is almost always measured by comparing its performance with the performance of a standard antenna whose gain is accurately known ([18]; [55], p. 581; [57], p. 454). Such measurements are referred to as relative gain measurements. If the gain of an antenna is measured without using an antenna whose gain is already known, the measurement is referred to as an absolute gain measurement. For absolute gain measurements, the well-known two-antenna method is generally considered to provide maximum accuracy ([18]; [55], p. 583; [57], p. 454).

Pyramidal horn antennas are the most commonly used standard antennas above 1 GHz. In addition to the fact that their gain can be easily and fairly accurately calculated [13], they are widely used because they are rugged, easy to construct, and have wide bandwidth properties. It has been well established [4], [10], [11], [13] that the measured gain of these horns agrees with the calculated gain to within ± 0.3 dB. For most antenna calibrations, this degree of accuracy is sufficient; however, advancements in radio astronomy, communications, and radar systems have created important demands for improved accuracy. Excellent examples of requirements for gain standards with uncertainties less than ± 0.1 dB are provided by Jull and Deloli [11] and by Chu and Semplak [4]. These reports are also notable because

they provide, to the author's knowledge, the only careful efforts to measure horn gain to within ±0.1 dB. The error estimates in these reports are somewhat questionable, however, because of certain difficulties that were not adequately considered. The two-antenna method was used for both of these efforts, and it would appear that further discussion of some aspects of this method is essential to establish the necessary procedures for high-accuracy measurements of horn gain.

C. Objectives and Approach

The main objectives of this paper are to review the literature concerning accurate horn-gain measurements, to present further discussion of the problems that have limited the accuracy of previous measurements, and to provide an adequate basis for future measurements accurate to within ±0.1 dB. The two-antenna method will be briefly discussed with reference to high-accuracy measurements, and certain aspects of the overall measurement problem will then be chosen for detailed discussion.

D. Terms

Scattering: For convenience, this term is considered to include reflection as a special case.

Multipath propagation: Wave propagation from a radiating system to some reference point by multiple transmission paths (i.e., by way of scattering objects in addition to the direct path).

Multipath interference (MPI): The existence of fields or voltages at some reference point (usually the receiving load) due to multipath propagation.

Interhorn (or interantenna) scattering: A form of multipath propagation caused by multiple scattering directly from horn to horn (or antenna to antenna).

Primary field: The radiated field of a radiating system that would exist under free-space conditions (i.e., if multipath propagation did not exist).

Proximity correction: A factor (usually expressed in dB) applied to a measured gain value to correct for the error caused by insufficient antenna separation. (This correction is often referred to as the ratio of "far-field gain" to "near-field gain," but this terminology is very inaccurate. For a discussion of this point, refer to Soejima [41]).

II. High-Accuracy Horn-Gain Measurements by the Two-Antenna Method

A. General

The usual formula used for gain measurements is obtained directly from (4) and (5), but it is seriously inadequate for high-accuracy gain measurements. Matching losses must be carefully considered and a detailed formulation is highly desirable. It is usually possible to connect the generator and load directly by an essentially perfect connector or adaptor, and the gain measurement can then be considered as involving an insertion-loss measurement [17], [18], [49]. Using a basic insertion-loss formula, it is shown in the Appendix that

$$G_T G_R = \frac{|1 - \Gamma_R \Gamma_L|^2 |1 - \Gamma_G \Gamma_T|^2}{(1 - |\Gamma_R|^2)(1 - |\Gamma_T|^2)|1 - \Gamma_G \Gamma_L|^2} \left(\frac{4\pi r}{\lambda}\right)^2 \left(\frac{{}^f P_L}{{}^i P_L}\right), \tag{6}$$

where ${}^i P_L$ is the power delivered to the load when the generator and load are directly coupled, ${}^f P_L$ is the power delivered to the load when the transmission path is "inserted" between the generator and load, and the ratio ${}^i P_L / {}^f P_L$ is the insertion loss. The reflection coefficients Γ_G, Γ_L, Γ_T, and Γ_R refer, respectively, to the generator, load, transmitting antenna, and receiving antenna.[4]

For a measurement made at a single frequency, Γ_G and Γ_L will usually be made "zero" by using tuners. Also, it may be desirable to incorporate tuners with the antennas to make Γ_T and Γ_R "zero." If measurements are to be made at many frequencies, however, it will probably be easier to measure the various reflection coefficients at each frequency before or after the insertion-loss measurement and use (6) without simplification.

The essential conditions involved in (6) can be summarized by grouping them as follows.

1) Antenna range—free-space conditions, uniform, plane-wave field at the receiving antenna.[5]
2) Transmitting medium—linear, lossless, reciprocal, isotropic.
3) Antennas—reciprocal, polarizations conjugately matched ([44]; [57], p. 442).
4) Equipment operation—ideal initial coupling between generator and load, single sinusoidal frequency, single waveguide mode, stable generator and load.

Most of these conditions require careful attention for horn-gain measurements of high accuracy, but the limiting problems are associated with the conditions grouped under 1). Because the errors resulting from violations of conditions 1) are usually relatively large, and because the techniques and corrections used to eliminate these errors are not well known, the remaining discussion will concern only the difficulties involved with these conditions.

B. A Representative Example of Gain Errors Caused by Multipath Interference and Insufficient Antenna Separation

Suppose that two identical horns are located as shown in Fig. 5 and oriented for a gain measurement. Using material presented in Sections III and IV, it is easy to estimate the error that will result, if not corrected, from multipath interference and insufficient antenna separation. The estimated errors shown in Fig. 1 are for typical gain measurements on a high-quality range: horns with 20-dB nominal gain,

[4] The two-antenna method assumes that the two antennas are identical. If there is a possibility that the antennas are significantly different, then a third antenna should be used to measure any difference by the comparison technique ([5], p. 732; [12], p. 442). This consideration is emphasized in (6) by using $G_T G_R$ rather than G^2.

[5] In this condition is satisfied when using identical antennas, the receiving antenna will also be in the far field of the transmitting antenna.

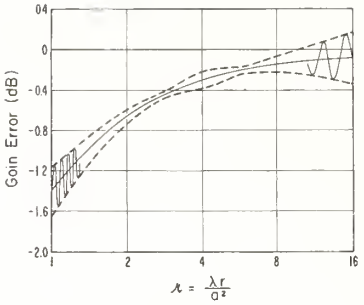


Fig. 1. Deviations in measured gain as a result of multipath interference and insufficient antenna separation. (Estimated values for a representative example.)

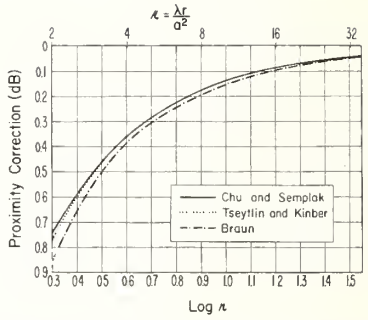


Fig. 2. Comparison of proximity corrections computed by methods provided by Chu and Semplak [4], Tseytlin and Kinber [45], and Braun [20]. The curves are for X-band horns of Slayton's design [13].

$h=a^2/\lambda$ (where "a" is the larger dimension of the rectangular aperture of the horns), and 20-dB power loss at the reflecting surface. The errors are graphed as a function of the normalized separation distance, $\kappa=\lambda r/a^2$. The periodic variations shown in Fig. 1 represent the effects of interference energy from the surface and from interhorn scattering.[6] (These variations are too closely spaced for accurate representation, and they are shown schematically in this figure.) The periodic variations are superposed on a monotonic variation (solid line in Fig. 1) that represents the error due to insufficient antenna separation.

For separation distances greater than $\kappa=8$, the periodic variations are caused almost entirely by interference energy from the specular surface; and for separation distances less than $\kappa=2$, the periodic variations are caused almost entirely by interference energy from interhorn scattering. Between $\kappa=2$ and $\kappa=8$, the variations are rather complicated and have the general appearance of two superposed sine waves of slowly varying period and amplitude. The two minima in the envelope curves are due to the first and second minima of the horn patterns.

It should be noted that the error in the measured gain due to insufficient antenna separation is significant even well beyond $\kappa=16$. This unfortunate fact is related to the rather large aperture "phase error" intrinsic to most horns ([55], p. 186; [20]). Also, it should be noted that the multipath interference level is significant at all separation distances. While it is technically feasible to construct an antenna range that has a lower level of multipath interference than the example considered here, it is very difficult or costly to do so.[7] In general then, high-accuracy horn-gain measurements will always involve both a method for analyzing or discriminating against multipath interference and a method for determining "proximity corrections."[6]

III. Proximity Corrections

A. Methods for Computing Proximity Corrections

Fields in the near-field region of simple apertures can be calculated without much difficulty, and such calculations are fairly numerous in the literature. However, the more difficult problem of power transfer from an aperture to a nearby aperture of comparable size has been calculated much less often. Apparently, the first such calculation was by Braun [20] (published in 1953).[8] Braun's calculation was for the case of two identical pyramidal horns, oriented as in the usual gain measurement, and it provided a simple scheme for computing proximity corrections for separations as small as $\kappa=2$. Recently, three similar calculations for pyramidal horns have appeared: Tseytlin and Kinber [45], Chu and Semplak [4], and Hamid [26].[9] These calculations are somewhat more rigorous than Braun's, and it is worth noting that the results of Tseytlin and Kinber [45] are not restricted to identical horns.

The computation schemes provided by Braun [20], Tseytlin and Kinber [45], and Chu and Semplak [4] are quite simple, but each paper provides graphs or tables that cannot be reproduced here due to space limitations. Corrections computed by each method are compared in Fig. 2 for X-band horns [13] operating at 10 GHz, and the curves[10] of this figure will be qualitatively correct for other horns of similar geometry. Since all three calculations are based on the same fundamental approximations, the deviations between the curves of Fig. 2 result from the different mathematical techniques used. The deviation shown by Braun's curve is due to an averaging technique that was not employed in the later calculations. The deviation between Chu and Semplak's curve and Tseytlin and Kinber's curve is small for the example given here and is due, at least in part, to a minor error in Tseytlin and Kinber's calculation (pointed out on p. 14 of their paper).

[6] Refer to "Terms" in the Introduction (Section I-D).

[7] Assuming that the range is to be used for frequencies as low as 1 GHz.

[8] Jakes [10] had previously noted the unusually large gain error for horns caused by receiver proximity. He found that, for optimum horns, this error could be corrected by measuring the horn separation from the apexes of the horns rather than from their aperture planes. No theoretical basis was given, but the correction was accurate to about ± 0.1 dB at $\kappa=2$. It is interesting that formulas developed by Hamid ([26], p. 118) from the geometrical theory of diffraction imply a certain amount of theoretical basis for Jakes' correction scheme.

[9] The author regrets that he was not aware of Hamid's work [26] in time to prepare an evaluation for the discussion here, which will be limited to the calculations by Braun [20], Tseytlin and Kinber [45], and Chu and Semplak [4].

[10] For accurate computations, the graphs by Braun [20] and Tseytlin and Kinber [45] should be enlarged and the tabular data by Chu and Semplak [4] should be graphed on large paper. This was done in computing the curves shown in Fig. 2.

It should be emphasized that a number of questionable assumptions are common to all three calculations, and the agreement between Tseytlin and Kinber [45] and Chu and Semplak [4] indiates only that these calculations will have about the same degree of inaccuracy. It is obvious that these calculations are valuable, but their limitations are not so apparent. They should be used with restraint until definite limits of accuracy can be established for them.

B. The Limits of Accuracy: An Estimate

For some later discussion, it is essential to know the basic approximations involved in the proximity corrections. These approximations are introduced now because they provide insight regarding the limits of accuracy for the corrections.

1) The Kirchhoff approximation is used in determing the field radiated by a horn ([55], p. 165; [48], ch. 9; [56], pp. 109–118). Among other things, the Kirchhoff approximation assumes that, in the general direction of maximum radiation, the contributions to the radiation field from currents and fields outside the aperture are negligible compared to the contribution from the aperture field. (This approximation is known to be accurate for large apertures, but most standard horns have small apertures.)

2) An approximation is used for the aperture field of a radiating horn. The field used is the field that would exist if the horn were extended indefinitely and excited in the dominant mode only. (It should be noted that this approximation also omits the effects of interhorn scattering.)

These are commonly used approximations, but it has apparently not been possible to compute definite limits for the associated errors. Even though the proximity corrections are ratio quantities that involve essentially the same assumptions in both the numerator and denominator [20], it would be optimistic to tacitly assume a high degree of accuracy for the corrections. Unfortunately, it seems that only Braun [20] has presented direct verification (by comparing gain measurements made for both small and large separation distances), and his verification was limited to about ± 0.1 dB. Tseytlin and Kinber [45] used the experimental results provided by Braun [20] and Jakes [10], and after considerable speculation they also estimated a verification of about ± 0.1 dB. Chu and Semplak [4] compared a corrected gain measurement made at a small separation distance with the calculated gain value. They found that the values differed by only 0.04 dB, but this comparison is not very conclusive since the gain calculation involves the same basic approximations as the proximity correction calculation.[11]

From the preceding discussion, it is evident that definite (i.e., neither conservative nor optimistic) error limits for the proximity corrections cannot be assigned at this time. It is felt that a reasonable error estimate is provided by assigning limits of ± 10 percent of the decibel value of the proximity corrections.

[11] If the conclusions reached by Slayton [13] and by Jull and Deloli [12] are correct (see Fig. 3), then this agreement would appear to be coincidental.

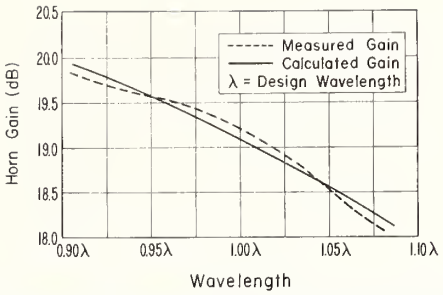


Fig. 3. Schematic representation of periodic variations in the gain of a small horn (as suggested by existing experimental data). λ is the design wavelength for the horn.

C. Some Unresolved Ambiguities

A report by Slayton [13] and a report by Jull and Deloli [11] (or [12]) appear to provide the only examples of precise measurements of gain as a function of frequency. In both ([13], [11]) it was concluded that the measured gain showed periodic variations somewhat as illustrated[12] in Fig. 3. (The measured-gain curve represents corrected gain values for measurements made at about $r = 3$.) In both reports these variations were considered to be "real," and it was suggested that the variations might be due to currents outside the horn aperture or higher-order modes in the aperture field. Thus, the variations in measured gain could represent interference, as a function of frequency, between the radiation from the dominant-mode aperture field and the radiation from these additional sources. However, it should be noted that this interference would also be a function of position; and because of approximations 1) and 2), the proximity corrections do not correct for this effect. Therefore, there is some ambiguity, if not error, in assuming that the measured gain values of Fig. 3 have been adequately corrected. It is reasonable to suppose that gain values measured using large separation distances might display variations significantly different from the variations shown in Fig. 3.

There are several other possible explanations for the deviations shown in Fig. 3. The possibility of some type of mismatch error deserves attention because the discussion will reveal some important considerations regarding experimental procedure. Because of approximation 2), the proximity corrections are based on the primary field only, and the effects of interhorn scattering must be eliminated by experimental technique. To measure the power that would be received from the primary field alone, the effect of interhorn scattering can be eliminated in the same manner as any other multipath interference component. It is also necessary to eliminate the effects of interhorn scattering (and any other scattering) when measuring the reflection coefficients of the horns or when adjusting any tuners incorporated with the horns.[13] This requirement is most accurately ac-

[12] The experimental data ([13], [11]) are only adequate to establish the magnitude of these variations. Figure 3 is based on a curve suggested by Jull and Deloli [11] and they emphasized that the period attributed to the variations is somewhat speculative.

[13] A horn that is matched (i.e., "reflectionless") for free-space transmission may appear to be mismatched when situated for the gain measurement, but the horn is actually properly matched with respect to the measurement scheme that is being discussed here.

complished by pointing the horns skyward.[14] If, instead, the horns are situated for the gain measurement when the reflection coefficients of the horns are measured, the resulting error in the measured gain will usually be small. However, the error resulting from this practice when adjusting tuners to provide matched horns can be much more serious. For instance, if the scattered energy entering the transmitting antenna is "tuned out" to achieve "reflectionless" power transmission, the effect will be to return this energy, minus ohmic losses, to the radiation space. Furthermore, the tuning process will, in general, change the relative phase between the radiated waves and the scattered waves leaving the antenna.[15] The result of these complications on the measured gain value depends, among other things, on the details of the tuning procedure and the particular technique used to eliminate the effects of multipath interference at the receiving load. Since the details of the matching procedure are not stated by Slayton [13] or Jull and Deloli [11], it is not possible to make a definite analysis; but some speculative analysis indicates that the variations illustrated in Fig. 3 may actually represent errors due to improper matching procedure. The limits of this possible error are easy to estimate, and the estimated limits of about ± 0.1 dB for $\rho = 3$ are in good agreement with Fig. 3.

In addition to being approximate, the proximity corrections correct only for the curvature and nonuniform power density of the primary field of the transmitting horn at the receiving horn. Actually, it is fortunate that they do not include the effects of interhorn scattering since they would then be much more difficult to compute (at least for $\rho < 6$). Furthermore, the reflection coefficients of the horns would necessarily have to be measured or tuned while situated for the gain measurement, and this requirement would greatly complicate the problem of eliminating the effects of the other multipath transmission components that are usually present. For instance, if the receiver is moved to introduce shifts in the relative phase between the multipath components, the reflection coefficients of the antennas would have to be remeasured or retuned for each position of the receiver.

IV. Multipath Interference

A. General Considerations

To date, horn-gain measurements with estimates of error less than ± 0.1 dB have all been made using small separation distances ($\rho < 4$), and these error estimates are questionable because they are dependent on the assumption of high accuracy for the proximity corrections. Until this assumption has been verified, large separation distances must be used for high-accuracy measurements of horn gain. Unfortunately, multipath interference (MPI) usually becomes a very serious problem when using large separation distances. The problem of MPI should be the primary consideration when designing the measurement.

There are a number of techniques that can be used to effectively eliminate MPI (for instance, using nanosecond pulses), but these techniques involve complications that are undesirable even if they could be developed to provide adequate precision. At the present time, the most satisfactory method for handling MPI is probably the following: by careful design, the antennas are located on a range so that the receiver can be moved sufficiently to produce reversals in the relative phase between the component voltage at the load due to the primary field and the component voltages due to the scattered fields.[16] Presumably, the MPI level will be small, and the "primary" (or "signal") power can be determined by simply averaging out the resulting periodic variations in the received power. Though simple in concept, this approach is complex in detail, and accurate results can be obtained only by careful design of the measurement. All of the significant sources of interference must be anticipated, and the antenna range, equipment location, and receiver movement must be designed to produce power variations that can be analyzed with precision and without ambiguity. In general, measurement designs should be based on attaining a low, analyzable level of interference rather than attaining a minimum level of interference. Special precaution is advisable when considering the use of baffles or screening fences since these will produce a lower level of interference only at the expense of increased complexity of interference, and it is by no means certain that the measurement of primary power will be made simpler or more precise by using these devices. Another common technique that is of questionable value is the practice of using the minima of the antenna patterns to reduce the MPI level at the receiving load. If sources of interference are located along these directions, it is obvious that the corresponding interference voltages may change radically as the receiver is moved and that it may be difficult or impossible to accurately analyze the resulting power perturbations.

The remainder of this section will be devoted to presenting some terms, concepts, and simple formulas that are very useful in designing and evaluating gain measurements involving significant MPI. This material is also applicable to other multipath transmission problems such as anechoic chamber evaluations and radar cross-section measurements.

B. Terms and Concepts

The measurement technique described above has been used extensively, and loosely, in evaluating anechoic chambers. In this connection, a few terms such as "free space VSWR" and "reflectivity level" have been used because of certain analogies to transmission-line measurements. While these terms are not devoid of meaning, they are ambiguous enough to be rather confusing. Standing waves result from wave reflections from symmetrical boundaries, and they represent only a special case of interference. The usual trans-

[14] An alternative procedure is to employ the same techniques used to eliminate multipath interference from the received power.

[15] It is pertinent to note that the scattering cross section for commonly used [13] horns is approximately equal to the receiving cross section.

[16] For simplicity of discussion, only receiver movements will be considered, but it may sometimes be necessary or desirable to employ transmitter movements or movements of surrounding objects.

mission-line terminology can be easily generalized by considering separately each wave of a multipath transmission to an arbitrary probe. The resultant voltage at the load of the probe is then the phasor sum of the separate voltages, as in (7) (where V_0 is the signal or reference voltage). The relations and definitions that follow are self-evident:

$$V_L = V_0 + V_1 + V_2 + \cdots \tag{7}$$

$$V_L = V_0(1 + K_1 + K_2 + \cdots) \tag{8}$$

$$V_L = V_0(1 + \kappa_1 e^{j\delta_1} + \kappa_2 e^{j\delta_2} + \cdots) \equiv V_0 F \tag{9}$$

$$P_L = P_0 F F^* \tag{10}$$

$$K_n \equiv \frac{V_n}{V_0} \equiv \left|\frac{V_n}{V_0}\right| e^{j\delta_n} \equiv \kappa_n e^{j\delta_n}, \qquad n = 1, 2, \cdots.$$

(In the following definitions, and elsewhere, identifying subscripts are implied when not stated explicitly.)

K_n = interference coefficient (for component "n")

$\kappa = |K|$ = interference voltage level

δ = phase difference (relative to V_0)

$k = 20 \log \kappa$ = interference power level

$\rho = \dfrac{1 + \kappa}{1 - \kappa}$ = voltage interference ratio (VIR)

$p = 20 \log \rho$ = power interference ratio (PIR)

P_0 = primary power, signal power, or reference power.

If the receiving and scattering characteristics of the probe are well known, and if extensive probe movements can be employed, it may be possible to determine all of the multipath field components. (Such an analysis is usually attempted when performing an evaluation of an anechoic chamber.) Furthermore, if the antenna, or "tip," of the probe is sufficiently small and isotropic, the interference pattern of the electromagnetic field can be measured directly since the voltage interference parameters at the load of the probe will be numerically equal to the field interference parameters at the tip of the probe. (A waveguide "VSWR" measurement can be interpreted as being such a measurement.) It should be noted also that the reference voltage does not necessarily have to be the signal voltage. For instance, in radar cross-section measurements, the signal voltage would result from the wave scattered by the target while the reference voltage might be derived either from a reference target or directly from the generator.

The PIR is a particularly useful concept. This is apparent if it is observed that, when the interference voltages are all small,

$$10 \log (P_{L\max}/P_{L\min}) \approx p_1 + p_2 + p_3 + \cdots. \tag{11}$$

The quantities κ, k, ρ, and p can be related in a convenient graph, as shown in Fig. 4. The dotted lines in this figure indicate how an interference voltage level of 0.01 is related

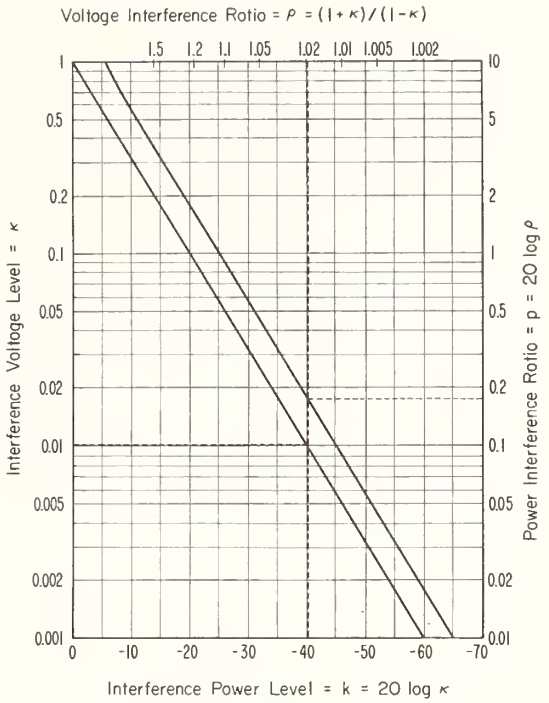


Fig. 4. Graph of multipath interference parameters.

to the other quantities. The following approximations are valid to within 5 percent for the indicated limits:

$$\rho \approx 1 + 2\kappa \qquad (\text{for } \kappa < 0.1) \tag{12}$$

$$p \approx 17.4\kappa \qquad (\text{for } \kappa < 0.3). \tag{13}$$

C. Measurement Design and Evaluation

Some important design considerations can be precisely stated if (10) is written in explicit form. Assuming that the interference voltages are all small,

$$P_L \approx P_0[1 + 2\kappa_1 \cos(2\pi \mathscr{D}_1 + {}^1\delta) + 2\kappa_2 \cos(2\pi \mathscr{D}_2 + {}^2\delta) + \cdots] \tag{14}$$

$$\mathscr{D}_n \equiv \lambda^{-1}(D_n - r) \qquad n = 1, 2, \cdots. \tag{15}$$

In (15), D_n is the length of the transmission path between the transmitting and receiving antennas for component "n" of the multipath field. In (14), the δ's represent the phase differences that are caused by phase shifts in the antennas and at the scattering surfaces. For a simple and accurate determination of P_0, the receiver movement must result in perturbations in P_L that are essentially a sum of sinusoidal variations. Thus, 1) the sum of the interference voltage levels should be less than about 0.03, and 2) during each perturbation cycle corresponding to component "n", κ_n and ${}^n\delta$ must remain essentially constant. Because fairly large receiver movements may be necessary, the precautions noted in Section IV-A are pertinent with respect to this second requirement.

Predicting the interference level for a gain measurement is

largely a matter of guesswork; therefore, the following estimation formulas for measurement design will emphasize simplicity rather than accuracy. A common source of interference is a "reflection" image of the transmitter, as in Fig. 5. For identical antennas the resulting interference power level $k_{I\theta}$ can be estimated by

$$k_{I\theta} \approx 2(\text{pattern level in dB})_\theta + (\text{surface loss in dB})_\theta \quad (16)$$

because the "inverse square" losses will be approximately equal. The number of perturbations $\mathscr{P}_{Ix}$ resulting from a linear receiver movement along the x-axis can be obtained from (15):

$$\mathscr{P}_{Ix} = \Delta\mathscr{D}_I = \lambda^{-1}[(r_I - r)_{x,0,0} - (r_I - r)_{0,0,0}]. \quad (17)$$

Then by expanding r_I and r in terms of h, s, and x, it is easy to show that

$$\mathscr{P}_{Ix} \approx 2\left(\frac{h}{s}\right)\frac{x}{\lambda}. \quad (18)$$

By repeating this procedure, it can be shown that

$$\mathscr{P}_{Iy} \approx \left(\frac{h}{s}\right)^2\frac{y^2}{s\lambda} \quad (19)$$

and

$$\mathscr{P}_{Iz} \approx 2\left(\frac{h}{s}\right)\frac{z}{\lambda}. \quad (20)$$

Approximations (18), (19), and (20) are accurate to within 10 percent if $h/s < \frac{1}{4}$ and x, y, or $z < \frac{1}{2}h$.

The successive scatterings between two antennas will produce a series of multipath components, but only the first component of this series will be significant for two-antenna gain measurements if $\mathscr{N} > 2$. By making some gross assumptions, it can be shown ([55], p. 590) that the corresponding interference voltage level κ_s can be estimated from $\kappa_s \approx (A_s/\lambda r)^2$, where A_s is the scattering cross section of the antennas. This approximation can be transformed into a very simple and useful formula by defining a scattering efficiency $f = A_s/A_p$, where A_p is the physical aperture area. Then, for rectangular apertures (dimensions a and b),

$$\kappa_s \approx \left(\frac{ab}{\lambda r}\frac{A_s}{A_p}\right)^2 = \left(\frac{b}{a}\frac{f}{\mathscr{N}}\right)^2. \quad (21)$$

Using $b/a = \frac{3}{4}$, which is typical for most horns, and using a scattering efficiency of $\frac{1}{3}$, which is typical for many antennas,

$$\kappa_s \approx \frac{1}{16\mathscr{N}^2}. \quad (22)$$

If the receiver is moved towards the transmitter, the power variations will have a period of $\lambda/2$. From (11), (13), and (22), the peak-to-peak variations in the decibel value of the received power will be approximately

$$p_s \approx \frac{1}{\mathscr{N}^2}. \quad (23)$$

A choice of $\frac{1}{3}$ for the scattering efficiency was made so that (22) and (23) would be in agreement with experimental data for Slayton's 20-dB horns. These formulas appear to be accurate to within about ± 20 percent for $\mathscr{N} > 2$.[17]

The same procedure used to obtain (22) can also be followed to obtain a formula for estimating the voltage that will appear in the transmission line of the transmitter due to interantenna scattering. If the transmitting antenna is reflectionless for free-space transmission, this voltage will cause an apparent reflection Γ_s that can be estimated from

$$|\Gamma_s| \approx \frac{1}{8\mathscr{N}^2}. \quad (24)$$

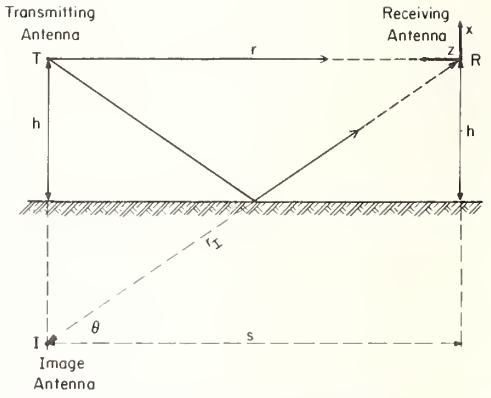


Fig. 5. Simple antenna range showing multipath interference from a reflection image of the transmitting antenna. Receiver movements used to evaluate multipath interference are related by rectangular coordinates located as shown.

V. Conclusions

If the signal power P_0 has been measured, then the gain can be determined by using (6) and applying a proximity correction. Referring to (6), if $G_T = G_R = G$, and if the factor containing the reflection coefficients is designated as M^2, it follows that

$$10\log G = 10\log M + 10\log\left(\frac{4\pi r}{\lambda}\right) + 5\log\left(\frac{P_0}{iP_L}\right) + C, \quad (25)$$

where C is the proximity correction in decibels. A thorough error analysis should include some discussion of all of the conditions listed for (6) in Section II-A.

Until a high degree of accuracy has been verified for the necessary proximity corrections, horn-gain measurements

[17] If the procedure used to derive (22) and (23) is repeated for circular apertures, the same formulas will result if "a" in $\mathscr{N} = \lambda r/a^2$ refers to the diameter of the aperture.

accurate to within ± 0.1 dB should not be attempted using separation distances less than about $r = 4$. Unfortunately, gain measurements using separations larger than $r = 4$ are usually very difficult because of complex multipath interference, and it is obvious that considerable effort is warranted in establishing more accurate limits of error for these corrections. Measurements are in progress at the National Bureau of Standards, Boulder, Colo., that should improve the verified accuracy of these corrections. These measurements are also designed to resolve the ambiguities discussed in Section III-C.[18] At the present time a national calibration service does not exist for antenna gain above 1 GHz, and these measurements are preliminary to establishing this service.

The problem of measuring the polarization parameters of pyramidal horns has not been discussed because these antennas are thought to be linearly polarized to a high degree in the direction of maximum gain. This is a common assumption, but it may deserve some investigation because of the possibility of higher-order aperture modes and the possibility of significant sources of radiation exterior to the aperture.

Appendix

A Working Formula for High-Accuracy Gain Measurements

Gain measurements by the two-antenna method involve determining the ratios of powers delivered to the receiving load. The generator and load are coupled more or less directly, and then the antenna path is "inserted" or "substituted" for the initial transmission device. In other words, a final two-port waveguide[19] junction (or simply "two-port") is substituted for an initial two-port waveguide junction [17]. Usually, the two-antenna method involves an insertion-loss measurement—that is, the initial two-port is an essentially perfect waveguide connector or adaptor. If the load and generator cannot be ideally coupled, then the gain measurement involves a substitution-loss measurement. The derivation presented here closely parallels a derivation by Beatty ([18]; [58], Session II, Lecture 3) for a nonideal initial two-port but where $\Gamma_G = \Gamma_L = 0$. The same method can be easily applied to the unsimplified case of substitution loss by using a formula provided by Beatty [17], but the resulting equation implies considerable computational effort. It will be assumed here that it is easier to arrange an ideal coupling for the initial two-port than it is to carry out the computations required for the general case.

The following insertion-loss equation can be written by inspection from Beatty [17]:

$$\frac{{}^{i}P_L}{{}^{f}P_L} = \frac{1}{\eta}\left[\left|\frac{1 - \Gamma_G\Gamma_1}{1 - \Gamma_G\Gamma_L}\right|^2 \frac{1 - |\Gamma_L|^2}{1 - |\Gamma_1|^2}\right] \tag{26}$$

where η is the efficiency of the final two-port and Γ_1 is the reflection coefficient of the generator side of the final two-port. Assuming that negligible scattered energy enters either antenna, and assuming that the incident waves at the receiving antenna do not differ significantly from uniform plane waves, then the final load power can be expressed in terms of the antenna parameters as

$${}^{f}P_L = SA_{R\text{mis}} = \frac{{}^{f}P_1 G_T}{4\pi r^2} A_{R\text{mis}} \tag{27}$$

where ${}^{f}P_1$ is the power accepted by the final two-port (i.e., the power accepted by the transmitting antenna) and $A_{R\text{mis}}$ is the effective receiving area of the receiving antenna as terminated by the mismatched load. $A_{R\text{mis}}$ is related to the receiving area A_R for a conjugately matched load as follows ([18]; [43]; [58], Session II, Lecture 3):

$$A_{R\text{mis}} = A_R \frac{(1 - |\Gamma_R|^2)(1 - |\Gamma_L|^2)}{|1 - \Gamma_R\Gamma_L|^2}. \tag{28}$$

Since ${}^{f}P_L/{}^{f}P_1 = \eta$, and $A_R = (\lambda^2/4\pi)G_R$, substitution of (28) into (27) yields

$$\eta = G_T G_R \left(\frac{\lambda}{4\pi r}\right)^2 \frac{(1 - |\Gamma_R|^2)(1 - |\Gamma_L|^2)}{|1 - \Gamma_R\Gamma_L|^2}. \tag{29}$$

Because of the assumption of negligible scattered energy, Γ_1 can be taken as Γ_T, and substitution of (29) into (26) yields the desired formula,

$$G_T G_R = \frac{|1 - \Gamma_R\Gamma_L|^2 |1 - \Gamma_G\Gamma_T|^2}{(1 - |\Gamma_R|^2)(1 - |\Gamma_T|^2)|1 - \Gamma_G\Gamma_L|^2} \left(\frac{4\pi r}{\lambda}\right)^2 \left(\frac{{}^{f}P_L}{{}^{i}P_L}\right). \tag{30}$$

Acknowledgment

The author is particularly endebted to Dr. R. C. Baird and R. W. Beatty of the National Bureau of Standards for many helpful suggestions.

Bibliography

Except for material that directly relates to the text, only a few recent papers and reports are listed here. For more extensive bibliographies concerning measurement techniques and measurements, [5], [57], and [59] are suggested; for more extensive bibliographies concerning electromagnetic theory or measurement theory, [5], [26], and [56] are suggested.

The following papers and reports are primarily concerned with measurement techniques or actual measurements.

[1] C. L. Andrews, "Diffraction patterns in a circular aperture measured in the microwave region," *J. Appl. Phys.*, vol. 21, pp. 761–767, August 1950.

[2] W. L. Barrow and F. D. Lewis, "The sectoral electromagnetic horn," *Proc. IRE*, vol. 27, pp. 41–50, January 1939.

[3] H. Brueckmann, "Antenna pattern measurement by satellite," *IEEE Trans. on Antennas and Propagation*, vol. AP-11, pp. 143–147, March 1963.

[18] These measurements are being made at *X*-band frequencies, and it should be possible to "scale" the results for use at lower frequencies.

[19] The term "waveguide" is used in the general sense to include all transmission lines.

385-582 O - 70 - 7

[4] T. S. Chu and R. A. Semplak, "Gain of electromagnetic horns," *Bell Sys. Tech. J.*, pp. 527–537, March 1965.
[5] W. A. Cumming, "Radiation measurements at radio frequencies: A survey of current techniques," *Proc. IRE*, vol. 47, pp. 705–735, May 1959.
[6] C. C. Cutler, A. P. King, and W. E. Kock, "Microwave antenna measurements," *Proc. IRE*, vol. 35, pp. 1462–1471, December 1947.
[7] R. J. Garbacz, "Determination of antenna parameters by scattering cross-section measurements," *Proc. IEE (London)*, vol. 111, pp. 1679–1868, October 1964.
[8] D. C. Hogg and R. W. Wilson, "A precise measurement of the gain of a large horn-reflector antenna," *Bell Sys. Tech. J.*, vol. 44, pp. 1019–1030, 1965.
[9] K. Iizuka, "Photoconductive probe for measuring electro-magnetic fields," *Proc. IEE (London)*, vol. 110, pp. 1747–1754, October 1963.
[10] W. C. Jakes, Jr., "Gain of electromagnetic horns," *Proc. IRE*, vol. 39, pp. 160–162, February 1951.
[11] E. V. Jull and E. P. Deloli, "The precise calibration of a horn-reflector antenna for radio astronomy," Nat'l Research Council of Canada, Radio and Elec. Engrg. Div., Ottawa, Rept. ERB-637, June 1963.
[12] ——, "An accurate absolute gain calibration of an antenna for radio astronomy," *IEEE Trans. on Antennas and Propagation*, vol. AP-12, pp. 439–447, July 1964.
[13] W. T. Slayton, "Design and calibration of microwave antenna gain standards," Naval Research Lab., Washington, D. C., Rept. 4433, November 1954.
[14] B. J. Strait and D. K. Cheng, "Microwave magnetic-field measurements by a modulated scattering technique," *Proc. IEE (London)*, vol. 109, pt. B, p. 33, 1962.
[15] A. M. Vural and D. K. Cheng, "A light modulated scattering technique for diffraction field measurements," *J. Res. NBS*, vol. 68D, pp. 355–362, April 1964.

The following papers and reports are primarily concerned with electromagnetic theory or measurement theory.

[16] W. L. Barrow and L. J. Chu, "Theory of the electromagnetic horn," *Proc. IRE*, vol. 27, pp. 51–64, January 1939.
[17] R. W. Beatty, "Insertion loss concepts," *Proc. IEEE*, vol. 52, pp. 663–671, June 1964.
[18] ——, "Discussion of errors in gain measurements of standard electromagnetic horns," NBS Tech. Note 351, March 1967.
[19] G. Borgiotti, "Calculation of directivity and equivalent area in terms of reaction," *IEEE Trans. on Antennas and Propagation (Communications)*, vol. AP-12, pp. 785–786, November 1964.
[20] E. H. Braun, "Gain of electromagnetic horns," *Proc. IRE*, vol. 41, pp. 109–115, January 1953.
[21] ——, "Calculation of the gain of small horns," *Proc. IRE (Correspondence)*, vol. 41, pp. 1785–1786, December 1953.
[22] ——, "Some data for the design of electromagnetic horns," *IRE Trans. on Antennas and Propagation*, AP-4, pp. 29–31, January 1956.
[23] J. Brown, "A theoretical analysis of some errors in aerial measurements," *Proc. IEE (London)*, vol. 105, pt. C, pp. 343–351, February 1958.
[24] ——, "A generalized form of the aerial reciprocity theorem," *Proc. IEE (London)*, vol. 105, pt. C, pp. 472–475, April 1958.
[25] T. S. Chu, "An approximate generalization of the Friis transmission formula," *Proc. IEEE (Correspondence)*, vol. 53, pp. 296–297, March 1965.
[26] M. A. K. Hamid, "Near field transmission between horn antennas," Dept. of Elec. Engrg., University of Toronto, Toronto, Canada, Research Rept. 43, April 1966.
[27] R. F. Harrington, "Theory of loaded scatterers," *Proc. IEE (London)*, vol. 111, pp. 617–623, April 1964.
[28] M. K. Hu, "Near-zone power transmission formulas," *IRE Nat'l Conv. Rec.*, pt. 8, pp. 128–135, 1958.
[29] E. V. Jull, "An investigation of near-field radiation patterns measured with large antennas," *IRE Trans. on Antennas and Propagation*, vol. AP-10, pp. 363–369, July 1962.
[30] A. F. Kay, "Near-field gain of aperture antennas," *IRE Trans. on Antennas and Propagation*, vol. AP-8, pp. 586–593, November 1960.
[31] J. B. Keller, "Diffraction by an aperture," *J. Appl. Phys.*, vol. 28, pp. 426–444, April 1957.
[32] ——, "Geometrical theory of diffraction," New York University, New York, N. Y., Rept. EM-175, April 1962.
[33] D. M. Kerns and E. S. Dayhoff, "Theory of diffraction in microwave interferometry," *J. Res. NBS*, vol. 64B, January–March 1960.
[34] B. Y. Kinber and V. B. Tseytlin, "Measurement of antenna parameters in the field of a plane wave generated by a collimator," *Radio Engrg. and Electronic Phys.*, vol. 10, pp. 1021–1031, July 1964.
[35] ——, "Measurement error of the directive gain and of the radiation pattern of antennas at short range," *Radio Engrg. and Electronic Phys.*, vol. 9, pp. 1304–1314, September 1964.
[36] B. Y. Kinber, "Inclusion of multiple diffractions by the scattering matrix method," *Radio Engrg. and Electronic Phys.*, vol. 9, pp. 1315–1323, September 1964.
[37] D. Midgley, "A theory of receiving aerials applied to the reradiation of an electromagnetic horn," *Proc. IEE (London)*, vol. 108, pt. B, pp. 645–660, November 1961.
[38] J. H. Richmond, "A reaction theorem and its application to antenna impedance calculations," *IRE Trans. on Antennas and Propagation*, vol. AP-9, pp. 515–520, November 1961.
[39] J. Robieux, "Near-zone power transmission formulas," *Proc. IRE (Correspondence)*, vol. 47, p. 1161, June 1959.
[40] F. Schwering, "On the range of validity of Fresnel-Kirchhoff's approximation formula," *IEEE Trans. on Antennas and Propagation (Communications)*, vol. AP-10, pp. 99–100, January 1962.
[41] T. Soejima, "Fresnel gain of aperture aerials," *Proc. IEE (London)*, vol. 110, pp. 1021–1027, June 1963.
[42] C. W. Steele, "A nonresonant perturbation theory," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-14, pp. 70–74, February 1966.
[43] A. F. Stevenson, "General theory of electromagnetic horns," *J. Appl. Phys.*, vol. 22, pp. 1447–1460, December 1951.
[44] C. T. Tai, "On the definition of the effective aperture of antennas," *IRE Trans. on Antennas and Propagation (Communications)*, vol. AP-9, pp. 224–225, March 1961.
[45] V. B. Tseytlin and B. Y. Kinber, "Measurement of the directive gain of horn antennas at a short distance," *Radio Engrg. and Electronic Phys.*, vol. 10, January 1965.

Books

[46] A. Z. Fradin, *Microwave Antennas*, transl. from Russian. New York: Pergamon, 1961.
[47] H. Jasik, *Antenna Engineering Handbook*. New York: McGraw-Hill, 1961.
[48] D. S. Jones, *The Theory of Electromagnetism*. New York: Pergamon (distributed by McMillan), 1964.
[49] D. M. Kerns and R. W. Beatty, *Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis*. New York: Pergamon, to be published in 1967.
[50] R. W. P. King, *Electromagnetic Engineering*. New York: McGraw-Hill, 1945.
[51] R. W. P. King and T. T. Wu, *The Scattering and Diffraction of Waves*. Cambridge, Mass.: Harvard University Press, 1959.
[52] J. D. Kraus, *Antennas*. New York: McGraw-Hill, 1950.
[53] S. A. Schelkunoff, *Electromagnetic Waves*. New York: Van Nostrand, 1943.
[54] S. A. Schelkunoff and H. T. Friis, *Antennas—Theory and Practice*. New York: Wiley, 1952.
[55] S. Silver, *Microwave Antenna Theory and Design*, M.I.T. Radiation Lab. Ser., vol. 12. New York: McGraw-Hill, 1949.

Miscellaneous

[56] *Electromagnetic Theory and Antennas*, E. C. Jordan, Ed. New York: Pergamon (distributed by McMillan), 1963. (A collection of papers in two volumes.)
[57] "IEEE Test Procedure for Antennas, Number 149 (Revision of 48 IRE 2S2) January, 1965," *IEEE Trans. on Antennas and Propagation*, vol. AP-13, pp. 437–466, May 1965.
[58] NBS Staff Members, "High frequency and microwave field strength" (unpublished NBS report), presented at the Precision Measurement Seminar, Boulder, Colo., 1966.
[59] *Progress in Radio Science (URSI) (Proc. XVth General Assembly)*, Secretary General of U.R.S.I., 7 Place Émile Danco, Bruxelles 18, Belgique.

NBS Field-Strength Standards and Measurements (30 Hz to 1000 MHz)

FRANK M. GREENE, MEMBER, IEEE

Abstract—**A description is given of the CW field-strength standards and associated measurement instrumentation and techniques, used for the calibration of both commercial and military field-strength meters in various frequency bands of the overall range from 30 Hz to 1000 MHz, which have been developed at the National Bureau of Standards during the past 25 to 30 years. The techniques are applicable only for evaluating the strength of steady-state, ac fields varying sinusoidally in time, and are not intended for use in broadband applications of any kind.**

Two principal types of field-strength standards and a prototype near-zone field-strength meter are described: 1) magnetic-field-strength standards (30 Hz to 30 MHz), 2) electric-field-strength standards (30 to 1000 MHz), and 3) near-zone electric-field-strength meter and interim field-strength standards (150 kHz to 30 MHz).

I. General Introduction

THIS PAPER presents a description of the various CW field-strength standards and associated measurement instrumentation and techniques developed during the past 25 to 30 years at the National Bureau of Standards. These are used for the calibration of both commercial and military field-strength meters in various frequency bands of the overall range from 30 Hz to 1000 MHz.[1] The techniques used are applicable only for evaluating the strength of steady-state, ac fields varying sinusoidally in time, and are not intended for use in broadband applications of any kind.

Two principal types of field-strength standards and a prototype near-zone field-strength meter are described. These are as follows.

1) *Magnetic-Field-Strength Standards* used over the frequency range 30 Hz to 30 MHz for the calibration of CW field-strength meters employing small-loop receiving antennas.

2) *Electric-Field-Strength Standards* used over the frequency range 30 to 1000 MHz for the calibration of CW field-strength meters employing half-wavelength self-resonant dipole receiving antennas.

3) *Near-Zone Electric-Field-Strength Meter and Interim Field-Strength Standards* used over the frequency range 150 kHz to 30 MHz for the evaluation of hazards of high-level electromagnetic radiation to ordnance devices and other uses.

The techniques used for calibrating field-strength meters must duplicate, insofar as possible, the conditions normally existing when these meters are used later for making actual measurements. Some of the principal factors that need to be considered are:

1) the spatial distribution of the amplitude and phase of the received wave over the aperture of the measuring antenna [1]–[3];

2) interaction between the measuring antenna and its image in the ground, in a nearby conducting plane, or in a nearby radiator [4];

3) perturbation of the field being measured by the RF transmission line connected between the measuring antenna and the receiver [5].

It is desirable to have *at least* two independent methods of measuring any physical quantity. From a direct intercomparison of the results of such measurements, then, it is possible to obtain a better estimate of the uncertainty existing in the methods. Basically, there are two independent methods or techniques by which the desired component of field strength can be evaluated. These are called 1) the *Standard-Antenna Method*, and 2) the *Standard-Field Method* [6]. The former consists of measuring the voltage induced in a standard receiving antenna by the component of the field being evaluated, and computing the value of the field in terms of this voltage, the dimensions and form of the standard antenna, and its orientation with respect to the field vector. The second method consists of generating and computing the desired component of a standard field in terms of the dimensions and form of a transmitting antenna, its current distribution, the distance from the transmitting antenna to the point at which the field component is being evaluated, and the effect, if any, of the ground.

The procedures described in this paper are used to establish standards of electric or magnetic field strength at a single amplitude level, or at most at a relatively few levels. The complete calibration of a field-strength meter is then made by extending these standard-field values to other levels, using secondary procedures which involve the attenuator calibration and the output-meter linearity determination of the instrument being calibrated. These procedures are then repeated, in part, for each specific frequency for which a calibration is to be carried out. A further discussion of these secondary procedures is given elsewhere [7], [8] and so will not be included in this paper.

All time-varying quantities used in this paper are expressed in terms of their rms values. Rationalized MKS units are used throughout.

Manuscript received March 28, 1967.

The author is with the National Bureau of Standards, Boulder, Colo.

[1] Field-strength measurements above 1000 MHz are discussed in a paper by R. R. Bowman elsewhere in this issue.

II. Magnetic-Field-Strength Standards (30 Hz to 30 MHz)

A. Introduction

A description will be given in this section of the magnetic-field-strength standards and measurement techniques used for the calibration of field-strength meters employing loop antennas in the frequency range 30 Hz to 30 MHz. As previously stated, these standards involve the use of single-frequency CW fields only.

In order that the calibration conditions duplicate, insofar as possible, those existing when the meters are used later for making measurements, the following specific factors need to be considered for the type of meters used in this frequency range.

1) The type of field (propagation mode and polarization).

2) Any change in the local $|E/H|$ ratio at the measurement site from the free-space plane-wave value caused by any of the following:

a) local ground nonhomogeneities, or discontinuities
b) multipath propagation
c) measuring in the near-field of a transmitting antenna
d) reflections or field distortion caused by local foreign objects.

3) Any difference in field distortion caused by the presence of the field-strength meter in the standard calibrating field and in the unknown field measured later.

B. Standard-Field Method

In this frequency range, NBS field-strength calibration services at the present time are limited to field-strength meters using small loop-type receiving antennas [7], [8]. Such antennas are either operated balanced with respect to ground or are electrostatically shielded. In either case their response is proportional to the average normal component of magnetic field strength incident on the loop. An accurately known quasi-static magnetic field can be produced by the *Standard-Field Method* for calibrating the antennas of such field-strength meters. A small, single-turn, circular, balanced transmitting loop is used to produce the standard field. The radius of the wire with which the loop is wound is assumed to be negligible compared to the loop radius r. The loop current I is assumed to be constant in amplitude and phase around the loop.

The loop antenna being calibrated is positioned coaxially with respect to the transmitting loop and is spaced an axial distance, d meters, from it. The geometry of the calibrating setup is shown in Fig. 1. The two loop antennas can also be positioned in the same plane. While coaxial loops are normally used for calibration purposes, co-planar loops may be useful under certain conditions (e.g., in the calibration of some types of ferrite-core loop antennas [9]).

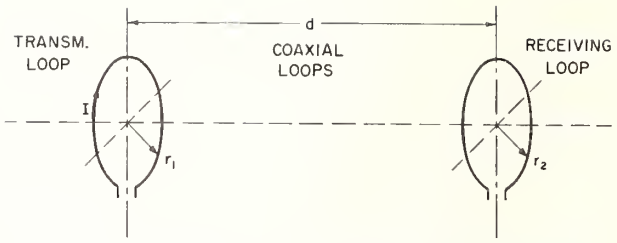


Fig. 1. Coaxial arrangement of transmitting loop antenna (radius r_1 meters), and receiving loop antenna (radius r_2 meters), spacing d meters, in the NBS calibration setup.

The rms value of the normal component of the magnetic field H, averaged over the area of the receiving loop, is given (in amperes per meter) by the following equation to within 0.2 percent [10], provided $\beta R_0 \leq 1.0$, and $r_1 r_2 / R_0^2 \leq 1/16$:

$$|H_{av}| \cong \frac{IS_1}{2\pi R_0^3}\left[1 + \frac{15}{8}\left(\frac{r_1 r_2}{R_0^2}\right)^2 + \frac{315}{64}\left(\frac{r_1 r_2}{R_0^2}\right)^4 + \cdots\right](1 + \beta^2 R_0^2)^{\frac{1}{2}}, \tag{1}$$

where

$$R_0 = [d^2 + r_1^2 + r_2^2]^{\frac{1}{2}} \tag{2}$$

and

I = transmitting loop current, rms amperes
S_1 = area of transmitting loop, square meters ($S_1 = \pi r_1^2$)
r_1 = radius of the transmitting loop, meters
r_2 = radius of the receiving loop, meters
d = axial spacing between the loops, meters
β = wavelength constant ($\beta = 2\pi/\lambda$)
λ = free-space wavelength, meters
f = frequency, hertz.

The infinite series contained in (1) converges extremely rapidly, more so than most other series solutions found in the literature. It is seldom necessary to use even the first correction term of the series, since if d/r_1 and d/r_2 are each greater than 4, the correction is less than 1.0 percent. Equation (2) for evaluating R_0 is also in an easier form to use than normally found. These factors tend to make this an easy formula to use for computing standard-field values.

For the conditions specified, higher-order terms in βR_0 are negligible and have been omitted from (1). Since the first correction term in the infinite series will contribute less than 1 percent, (1) can be further simplified to

$$|H_{av}| \cong \frac{IS_1(1 + \beta^2 R_0^2)^{\frac{1}{2}}}{2\pi R_0^3} \text{ amperes per meter.} \tag{3}$$

1) Calibration of Field-Strength Meters: The loop current is measured by means of a vacuum thermocouple previously calibrated with direct current. It is located sym-

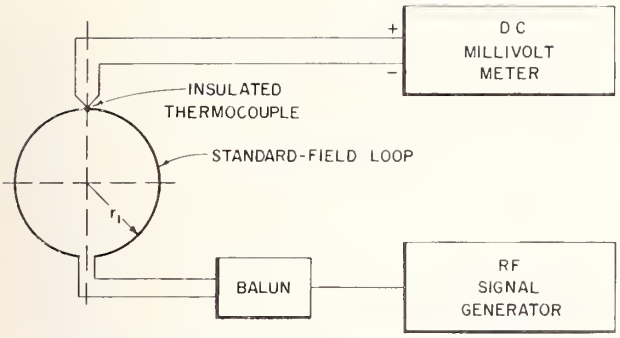


Fig. 2. Method used to excite the NBS standard transmitting loop antenna, and to measure the RF loop current.

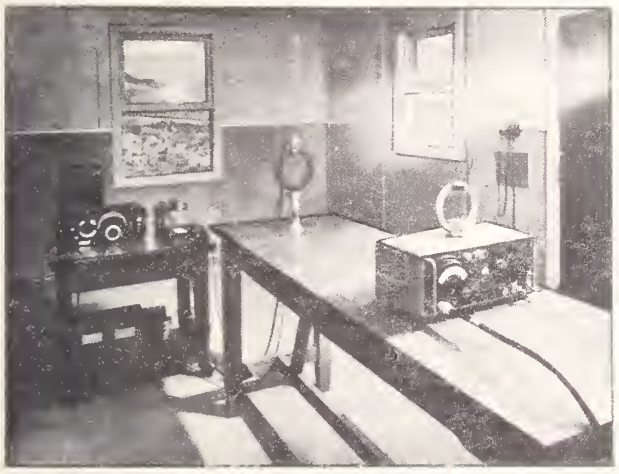


Fig. 3. The NBS field-strength calibration site showing: (left) RF generating and measuring instrumentation; (center) the standard transmitting loop antenna; and (right) a field-strength meter under calibration.

metrically at the top of the loop winding, and its dc output is measured with a dc millivoltmeter or a precision slide-wire potentiometer as shown in Fig. 2. The RF substitution error of this type of thermocouple, when used for this purpose in a balanced circuit with the thermocouple at essentially ground potential, has been found to be less than 1 percent at frequencies even as high as 100 MHz [11].

A clear space must be used for the calibration such that a distance of at least two or three times the spacing d between loops exists from either loop to the nearest sizable metallic objects and to the ground. The effect of these objects in distorting the field can be determined by placing metal objects of similar size in the vicinity of the loops and noting the effect on the value of the field at the receiving loop. The calibration setup at NBS, which is located in a small frame building remote from the main laboratory, is shown in Fig. 3. The use of metal and electrical wiring was kept to a minimum in its construction.

The value of the quasi-static magnetic field H produced by the standard loop is expressed in terms of the equivalent electric component E (in volts per meter) that would exist in a free-space, plane-wave radiation field. For this case, the magnitudes of these two field components are related by the plane-wave impedance ($\eta = \sqrt{\mu/\varepsilon} \cong 120\pi$), such that $E \cong 120\pi H$. Using this relationship, (3) can be rewritten (in volts per meter) as

$$|E| \cong \frac{60\pi r_1^2 I}{R_0^3} (1 + \beta^2 R_0^2)^{\frac{1}{2}}. \tag{4}$$

The value of field given by (4) is essentially independent of frequency up to about 5 MHz. At higher frequencies, the frequency correction term under the radical (the induction-field component) begins to become appreciable for the spacing in the NBS standard ($d \cong 1.25$ meters). The value of equivalent field strength E used for calibration is of the order of 0.1 volt per meter ($r_1 = 0.1$ meter, $I = 0.1$ ampere).

Portable VLF standard-field calibrating instrumentation has also been recently developed at NBS using the above principles [12]. With this equipment a very-low-frequency field-strength meter can be periodically calibrated during a sequence of measurements in the field.

2) *Factors Affecting the Accuracy of Calibration and Use:*

Loop-antenna current distribution: The accuracy of the value of the standard field given by (3) or (4) is affected by the constancy of the current I as well as its phase around the circular transmitting loop. The uniformity depends upon the circumference of the loop in wavelengths, but usually does not become a problem except at frequencies above 5 or 10 MHz [13]. In order to keep the decrease in average current from exceeding 1 percent, the loop circumference should not exceed $\lambda/16$ at the highest calibrating frequency [10].

An interesting error analysis was recently made on the *standard-field* method of loop calibration [14]. The author estimated the uncertainty to be from 2.5 to 4.0 percent, after analyzing possible errors in evaluating the various parameters involved.

Local field distortion: An important factor that can affect the accuracy of calibration and use of a loop-type field-strength meter is the distortion of a field in the immediate vicinity of the instrument by its enclosing metallic case. If the loop antenna of the field-strength meter is mounted a distance less than its own diameter above the case, the field-strength meter must be calibrated and later used with a specific relative angular orientation between the loop antenna, the case, and the incident field being measured. Failure to observe this precaution may result in a measurement error of as much as 5 to 10 percent, depending upon the height of the loop above the case [15], [16]. An example of the variation in this error as a function of orientation between the loop antenna and case is shown in Fig. 4. This is the measured error that would result for one particular make of field-strength meter whose antenna was very close to the case. An even larger error may be encountered when a loop antenna is mounted above the metallic body of an automobile, as, for example, when making mobile field-strength measurements [17].

The field distortion just described is, in general, different when the field-strength meter is being calibrated in the near-zone quasi-static field given by (3) or (4) than when used

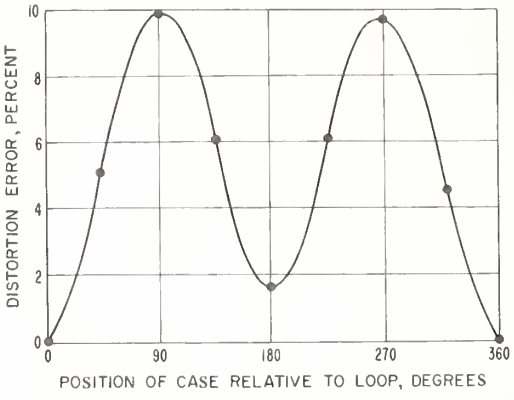


Fig. 4. Field-distortion error versus the orientation of a loop-antenna-type field-strength meter. The orientation angle is measured between the plane of the loop and the long dimension of the case, with the position of the loop remaining fixed relative to the incident magnetic field.

later to measure a uniform plane wave. In the former case, the instrument is within one or two meters of the standard transmitting loop, in a field that is varying rapidly as $1/d^3$. In the latter case, the meter is usually one or more kilometers from the transmitting antenna, in a radiation field that is varying much more slowly and approximately as $1/d$. There is therefore a large difference, in these two cases, in the percentage change in field strength along the longitudinal dimension of the field-strength meter. Any resulting difference in local field distortion in the two instances can be interpreted as a calibration error for those field-strength meters with their loop antenna mounted close to the case. While any such error is probably small, in most instances, this is something that should probably be investigated further.

Use of loop-type field-strength meters: The response of a small receiving loop antenna that is either operated balanced with respect to ground, or that is electrostatically shielded, is proportional to the average normal component of magnetic field strength incident on the loop. It can be thought of as a magnetic probe. However, the calibration and subsequent measurements of field strength are expressed in terms of the electric component E that would exist if the measurement were being made in free space (in this case, $E \cong 120\pi H$).

When such a field-strength meter is used to make measurements near the ground, the indicated value of the electric component of field strength may not always be valid. If there are serious local nonhomogeneities, or discontinuities in the electrical properties of the ground (ε_r and σ), the free-space relationship between total E and total H does not hold. The same will be true if measurements are made in a multipath field [18] resulting from any type of reflection. Likewise, measurements made with such an instrument in the near field of a transmitting antenna ($d \leq \lambda$) will be of questionable value. In this case, the ratio $|E/H|$ may range from a value much less than that in free space, to one much greater, depending upon the type of transmitting antenna (i.e., whether basically a magnetic dipole, or basically an electric dipole), distance, polarization, etc. This is true in near-field *antenna* measurements, as well as in very-low-frequency *propagation* measurements, where the near field may extend many miles from the transmitting antenna [19].

However, in all of these instances, the value of the magnetic component of the field H (in amperes per meter) can be correctly determined if the value of equivalent electric field strength E (in volts per meter), as indicated by the meter, is divided by 120π. If it is a knowledge of the electric component of field strength that is desired in these cases, it must be determined directly using a meter whose response is proportional to the electric component of the field.

C. Standard-Antenna Method

1) General: This is an independent method that can be used for evaluating magnetic field strength at frequencies above a few megahertz. Its use has been limited mainly to the higher frequencies because of the lack of a balanced-voltage standard of suitable sensitivity. Its principal use here is for intercomparison with the standard-field method to determine the agreement between the standards. The normal component of the magnetic field can be evaluated in terms of the rms voltage V_0 that it induces in a standard circular receiving loop antenna of known dimensions from the following (based on the relation $V = -N(d\phi/dt)$)

$$|H_{av}| = \frac{V_0 \cdot 10^7}{4\pi S_2 N \omega} \text{ amperes per meter,} \tag{5}$$

where

S_2 = area of the receiving loop, square meters ($S_2 = \pi r_2^2$)
N = the number of loop turns
r_2 = radius of the receiving loop, meters
$\omega = 2\pi f$.

It is necessary that the receiving loop antenna be electrically small ($2\pi r_2 \ll \lambda$), if the induced voltage is to be equal to the voltage appearing at the antenna terminals. In effect, this means that both the self-inductance and the distributed capacitance should be small, or the operating frequency of the loop should be low compared to its lowest self-resonant frequency.

The distributed capacitance will cause the terminal voltage V' of the loop to be higher than the induced voltage V_0 due to partial resonance. It can be shown from elementary circuit theory [20] that for the case where the distributed capacitance can be replaced by an equivalent *lumped* capacitance, the relative rise in the terminal voltage will be given quite accurately by (6) so long as $f_0/f_r < 1/2$:

$$\frac{V'}{V_0} \cong \left[1 - \left(\frac{f_0}{f_r}\right)^2\right]^{-1}, \tag{6}$$

where f_0/f_r is the ratio of the operating frequency of the loop antenna to its lowest self-resonant frequency. It can be seen that the ratio f_0/f_r must be less than 0.1 for the error in terminal voltage to be less than 1 percent.

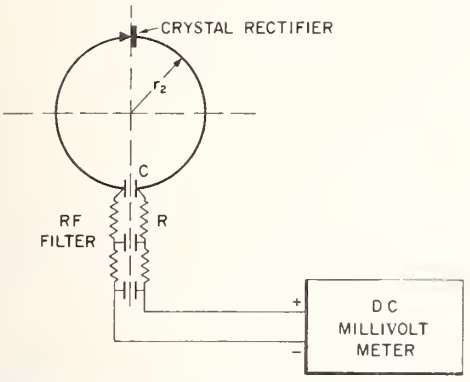


Fig. 5. Method used to measure the induced RF voltage in the NBS standard receiving loop antenna.

2) Measuring the Antenna Voltage: The standard receiving loop antenna is similar in size to the standard-field transmitting loop, and uses a single turn having a radius of approximately 0.1 meter. The induced voltage is measured by means of a calibrated balanced silicon-crystal [21] voltmeter mounted at the top of the loop turn. The dc output voltage of the crystal voltmeter is measured across an insulated gap at the bottom of the loop by means of a sensitive dc millivoltmeter. There is an RF bypass capacitor across the gap, and the dc leads are filtered close to the loop to prevent any RF voltage that might be induced in them from reaching the crystal diode and being rectified. A schematic diagram of the loop and associated measuring circuitry is shown in Fig. 5.

The crystal is calibrated in terms of an accurate RF voltage standard by removing it from the loop and placing it is a special Type N coaxial mount. The voltage standard is derived from an accurate self-balancing type of thermistor power bridge.

The standard receiving loop is unshielded and must be operated balanced with respect to ground to avoid the electric-dipole response that would otherwise exist [22]. It is wound with No. 14 AWG copper wire, and is air-supported to minimize its distributed capacitance. The silicon semiconductor diode used has a low value of shunt capacity (0.5 pF or less). Subminiature glass diodes selected from types 1N832-A, 1N830-A, or 1N82-A or equivalent are satisfactory. Corrections obtained from (6) were applied to measurements made with this loop.

D. Uncertainty of the NBS Standards

The *Standard-Antenna Method* of evaluating magnetic field strength based on (5) has been intercompared with the *Standard-Field Method* based on (3) or (4) at several frequencies up to 30 MHz. The agreement obtained between the two methods was within 3 percent at frequencies below 5 MHz, and within 5 percent at frequencies from 5 to 30 MHz. The basic setup used for these intercomparisons is the same as that shown in Fig. 1. Several axial spacings d were used between the coaxial loops in the range from 0.5 to 2 meters.

III. Electric-Field-Strength Standards (30 to 1000 MHz)

A. Introduction

A description will be given in this section of the electric-field-strength standards and measurement techniques used for the calibration of VHF-UHF field-strength meters employing half-wavelength self-resonant dipole antennas in the frequency range 30 to 1000 MHz. The procedures described are generally carried out only for horizontal polarization for reasons of convenience and maximum accuracy. The techniques are applicable for evaluating steady-state, uniform-plane-wave, CW fields only.

It is necessary here also that the calibration conditions be essentially the same as those under which measurements will be made later. Some of the principal factors involved for the type of field-strength meters used in this frequency range are: 1) the spatial distribution of amplitude and phase in the wave under measurement, 2) the influence of the ground on the self-impedance of the measuring antenna, and 3) interaction between any RF transmission line used and the field being measured.

The *standard-antenna* and *standard-field* methods are utilized to evaluate field strength in this frequency range, as discussed in the previous section for the lower frequencies. The techniques are very similar except for obvious differences in the antennas, and in the voltage- and current-measuring instruments.

B. The Standard-Antenna Method

1) Effective Length of the Antenna: The magnitude of the electric component of field strength existing at a given point in space may be determined in terms of the voltage V_{oc} induced in a standard receiving dipole immersed in the field, together with certain factors involving the antenna geometry. It will be assumed in the following that the measuring dipole is oriented parallel to the electric-field vector E and that the voltage induced is referred to the center terminals. The relationship used is (in volts per meter)

$$|E| = \frac{V_{oc}}{L_{\text{eff}}}, \tag{7}$$

where L_{eff} is the effective length of the antenna in meters.

Assuming a sinusoidal current distribution on a half-wavelength or shorter dipole, its effective length, when measuring a uniform plane wave, is [23]

$$L_{\text{eff}} = \int_{l-}^{l} \frac{I(x)}{I(0)}\,dx = L\left[\frac{\tan\left(\frac{\beta l}{2}\right)}{\beta l}\right] \text{ meters}, \tag{8}$$

where $I(x)/I(0)$ is the relative current distribution along the antenna (in the transmitting mode), and L is the physical length in meters ($L = 2l$). The effective lengths of a half-wavelength dipole and an electrically short dipole are, respectively,

$$(l = \lambda/4), \qquad L_{\text{eff}} = \left(\frac{2}{\pi}\right)L, \text{ meters}, \tag{9}$$

$$(l \ll \lambda/4), \qquad L_{\text{eff}} \cong \left(\frac{1}{2}\right)L, \text{ meters}. \tag{10}$$

Equation (9) is only valid and meaningful for the uniform plane wave case, while the approximation (10) is valid when measuring any complex electric field, if $l \leq 0.03\ \lambda$.

The current distribution on an antenna is not exactly sinusoidal except for an infinitely thin filament. For cylindrical antennas of radius a and half-length l the departure from a sinusoidal distribution becomes progressively greater as the ratio $2\ l/a$ decreases (i.e., as the antenna gets fatter [2], [24]). This results in a small proportionate *increase* in the effective length.

In order to operate a half-wavelength dipole at self-resonance, it is necessary to shorten its physical length somewhat from a full half-wavelength due to the capacitive end effect. This results in a corresponding *decrease* in the effective length [23]. This *decrease* and the previous *increase* tend to cancel each other.

2) Description of the Standard Receiving Antenna: The standard receiving dipoles are shortened the required amount from a physical half-wavelength to make them self-resonant at their operating frequency. The antenna tubing or rod stock used at NBS varies in diameter from 3/8 inch (9.52 millimeters) for the 30-MHz dipoles down to 1/16 inch (1.59 millimeters) or less for the 1000-MHz antennas. The required percentage shortening for self-resonant operation varies from 3.8 percent at 30 MHz to 5.3 percent at 1000 MHz. The percentage increase in effective length, due to the finite radius of these antennas, varies from 5.5 percent at 30 MHz to 8.2 percent at 1000 MHz.

The induced voltage is measured directly by means of a relatively high-impedance balanced voltmeter connected across the gap at the center. In this manner the necessity for a separate measurement of the antenna input impedance is eliminated, which greatly simplifies the problem. The induced voltage is rectified by a low-capacitance point-contact silicon-crystal diode [21] built into the gap as shown in Fig. 6. The crystal output is filtered by means of a balanced resistance-capacitance network, and the dc output voltage is measured on a sensitive dc millivoltmeter. The central portion of one of the standard receiving-antenna assemblies minus the antenna rods is shown in Fig. 7. The input impedance of the resistance-capacitance filter is approximately 10 000 ohms in shunt with 0.05 pF.

3) Influence of the Ground: VHF-UHF field-strength measurements generally will be in error if made at receiving antenna heights other than that used when the field-strength meter was calibrated. An error will likewise exist if the electrical ground constants (ε_r and σ) at the site chosen to make measurements are appreciably different from those existing at the time or place of calibration. This results from a change of the antenna-input impedance Z_A with height above ground (or with changing ground conditions), caused by the interaction between the measuring antenna and its image in the ground. The measurement error is therefore also a function of the antenna load impedance Z_L.

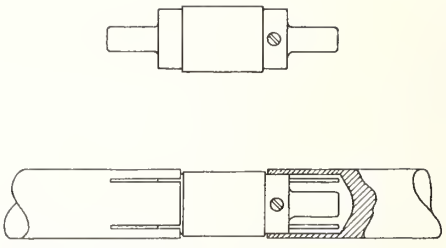


Fig. 6. Method used to mount the silicon-crystal diode in the gap at the center of the NBS half-wavelength standard receiving dipole.

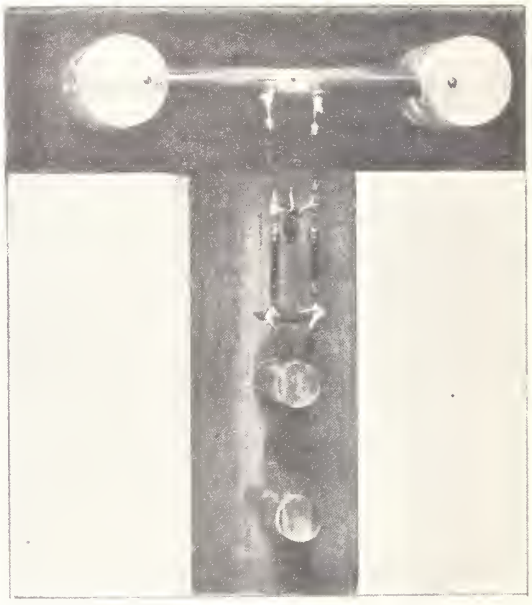


Fig. 7. View of the central portion of an experimental standard receiving dipole minus the antenna rods, showing the crystal diode and resistance-capacitance filter network.

If a field-strength meter is calibrated with its receiving antenna at a height greater than two or three wavelengths above the ground, the calibration can be said to have essentially a "free-space" value. That is, the effect of the ground on the calibration will be minor at such an antenna height.

The percentage difference between the *true* and *indicated* values of field strength, when making measurements at other antenna heights, can be as large as 10 to 15 percent (for horizontal polarization) at heights appreciably less than one wavelength, as can be seen from Fig. 8 [25]. The error will decrease with increasing antenna height, and will be generally less than 5 percent for heights greater than one wavelength over average ground ($\varepsilon_r \cong 15$, $\sigma/\varepsilon\omega \ll 1$).

This error will also decrease as the antenna load impedance is increased, since any fluctuations in Z_A will become small in comparison. The resulting error will approach zero as Z_L approaches infinity (other factors remaining the

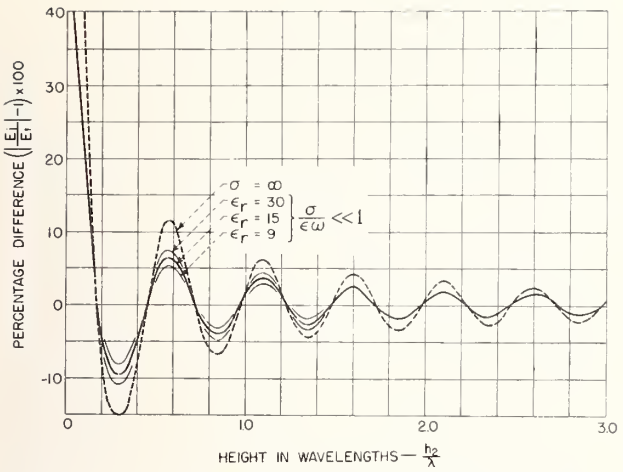


Fig. 8. The theoretical effect of the ground on the accuracy of electric-field-strength measurements using a horizontally polarized, half-wavelength, self-resonant dipole terminated in a matched load ($Z_L = Z_A \cong 73$ ohms). The percentage difference between the true E_t and indicated E_i values of field strength (the measurement error) is a function of the electrical ground constants, ε_r and σ; the receiving-antenna load impedance Z_L, and the antenna height (in wavelengths) above ground h_2/λ.

same). If the field-strength meter can be designed to provide a sufficiently large value of Z_L, the calibration will have essentially a "free-space" value [25]. That is, the calibration will be largely independent of the height of the antenna above ground, the antenna polarization, and any changes in the electrical constants of the ground itself.

C. Standard-Field Method

1) General: A predetermined value of field strength may be established at a given point in space in terms of the current distribution in a transmitting antenna, the effect of the ground, and the geometry involved. Horizontally polarized transmission over plane-homogeneous earth having finite values of relative dielectric constant ε_r and conductivity σ will be considered. The geometry of the transmission system is shown in Fig. 9.

The rms value of the electric component of field strength produced by a horizontal transmitting dipole, in its equatorial plane, at distances greater than about 2λ over plane homogeneous earth is [26]–[28] (in volts per meter)

$$E \cong -j\frac{60\pi L_{\text{eff}} I}{\lambda}\left[\underbrace{\frac{e^{-j\beta R_1}}{R_1}}_{\substack{\text{Direct}\\ \text{wave}}} + \underbrace{\frac{\Gamma e^{-j\beta R_2}}{R_2}}_{\substack{\text{Ground-}\\ \text{reflected}\\ \text{wave}}} + \underbrace{\frac{(1-\Gamma)}{R_2}A(R)e^{-j\beta R_2}}_{\substack{\text{Surface}\\ \text{wave}}}\right], \tag{11}$$

in which

L_{eff} = effective length of the antenna, meters
λ = free-space wavelength, meters
R_1 = direct-ray path length, meters
R_2 = ground-reflected-ray path length, meters
$\Gamma = \rho e^{-j\phi}$ = complex plane-wave reflection coefficient (horizontal polarization)
$A(R)$ = complex surface-wave attenuation factor
I = rms current in amperes at the center of the transmitting antenna
$\beta = 2\pi/\lambda$
$j = \sqrt{-1}$.

The first term of (11) represents the field strength that would exist if the transmitting and receiving antennas were located in free space. The remaining terms take into account the presence of the earth. The second term represents the ground-reflected wave which, when added vectorially to the first term of (11), comprises the space wave. The third term of (11) represents the surface wave associated with RF currents actually flowing in the ground. The total field given by (11) is usually referred to as the ground wave. Propagation via the ionosphere or troposphere, which is subject to changing solar or meteorological conditions, is ignored here. The magnitude of the surface wave is usually negligibly small for horizontally polarized transmission at the antenna heights and frequencies involved here [29]. In using (11) for establishing standard electric fields, then, the presence of the surface wave will be neglected entirely.

2) Description of the Standard Transmitting Antenna: The transmitting antenna used is identical to the standard receiving antenna previously described. The RF current flowing at its center can be measured either by means of a calibrated VHF vacuum thermocouple, or a small bead-type thermistor and associated bridge. A thermocouple is satisfactory for use in the lower part of the frequency range, but the thermistor is preferable and can be used over the entire frequency range of interest here with not over a 1- or 2-percent error [30]. The antenna and thermistor, respectively, terminate two sections of 72-ohm balanced, unshielded, twin-lead transmission line, each one-quarter wavelength long, as shown in Fig. 10. A balanced feed-line is connected in common with their sending ends, and a special broadband balun [31] is used to transform the unbalanced output of the RF generator to the balanced transmission system. A suitable thermistor bridge is used to complete the current-measuring instrumentation as shown.

The transfer impedance of a quarter-wavelength lossless line is equal to its characteristic impedance, and is *not* a function of the terminating impedance. If the characteristic impedances of the two quarter-wavelength sections used are identical, the currents flowing at their receiving ends will also be identical for a given voltage applied to their sending ends. The magnitude of the RF current flowing at the center of the transmitting antenna is thus indicated by the thermistor bead (or thermocouple), $\lambda/2$ distant, and is independent of the magnitude and phase angle of the two terminations, namely, the antenna input impedance on the

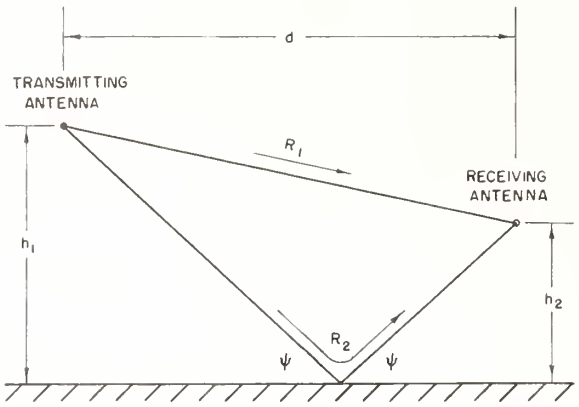


Fig. 9. Ray-path diagram showing direct ray along R_1, and ground-reflected ray along R_2 making an angle $\psi = \tan^{-1}(h_1 + h_2)/d$ with the earth.

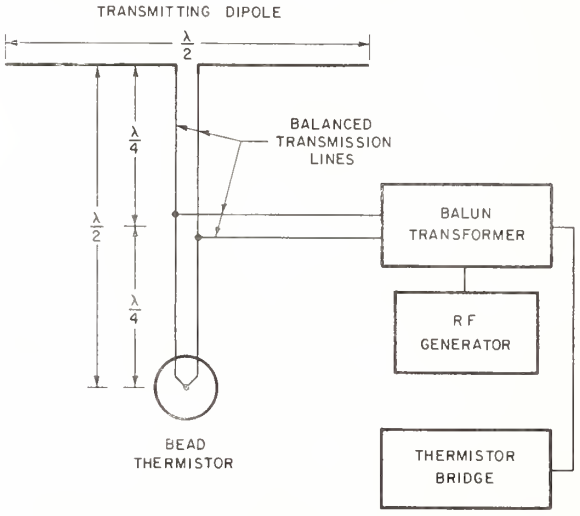


Fig. 10. Diagram showing the method used to measure the RF current at the center of the NBS transmitting dipole.

one hand, and that of the thermistor bead, or thermocouple [32] on the other.

D. Propagation Tests

1) General: An intercomparison can be made between the *standard-antenna* and *standard-field* methods in order to obtain a better estimate of the uncertainty existing in the two standards. The value of field strength established using the standard-field method based on (11) is simultaneously measured using the standard-antenna method based on (7), and the results intercompared. This can be accomplished by either of two different methods.

The first, called the *standing-wave* method, involves the complex plane-wave reflection coefficient Γ of the ground at a selected test site. This method is based on a vertical-incidence technique that enables one to measure Γ, and to make the intercomparison between the two standards essentially simultaneously in the same setup. In the second, or *field-averaging* method, a vertical-incidence technique is also employed. Using a similar setup, the ground-reflected

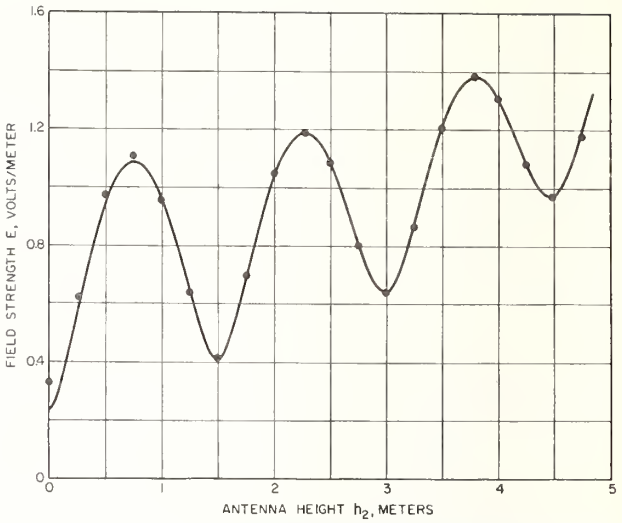


Fig. 11. Variation of the theoretical and measured values of horizontally polarized electric-field strength $|E|$ in volts per meter, versus the height h_2, in meters, of the horizontal standard receiving dipole (as used in the *standing-wave method*).

wave is, in effect, averaged out, yielding directly the free-space field without the need to know the ground-reflection coefficient.

2) Standing-Wave Method: In this method the horizontal transmitting dipole is mounted at the top of a light, wooden, ladder mast at a *fixed* height of several wavelengths above the ground. In the space beneath, the direct wave and the wave reflected from the ground at vertical incidence combine to produce a standing wave, as shown in Fig. 11. The receiving dipole is mounted parallel to the transmitting dipole, but on a sliding carriage running vertically on the ladder rails. This permits measurement of the resulting variation in field strength with height above ground. The reflection coefficient of the ground for vertical incidence is determined from the resulting standing-wave ratios. The resulting value of Γ is then substituted in (11) from which the magnitude of the standard field can be calculated for any height above the ground. The results are then compared with corresponding measured values, as determined using (7), to determine the agreement between the standards.

A separate analysis has been made recently of the errors involved in measuring the reflection coefficient of the ground [33] using the *standing-wave* method. The error was reported to vary from approximately 5 to 17 percent over the range of the various parameters involved. However, the effect of this error on the final accuracy in evaluating the electric-field strength is considerably less, as can be determined from (11).

3) Field-Averaging Method: In this method, both the transmitting and receiving dipoles are mounted parallel to each other on the sliding carriage. A fixed spacing of approximately two wavelengths is used between the two dipoles, so that essentially far-field conditions will exist. The resulting standing-wave pattern that is present, as the carriage is varied in height, is shown in Fig. 12. Since the spacing between the two antennas is fixed, the *direct* wave,

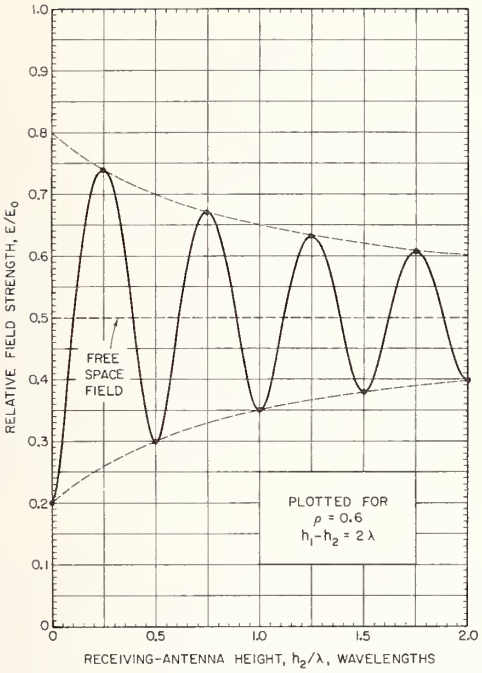


Fig. 12. Variation of the relative horizontally polarized electric-field strength E/E_0 versus the height h_2/λ, in wavelengths, for a fixed vertical spacing between the horizontal standard transmitting and receiving dipoles (as used in the *field-averaging method*).

or *free-space* field, given by the first term of (11) will remain constant in magnitude. As the carriage is moved up or down the mast, the resulting value of field strength will oscillate around this free-space value. By properly averaging the measured data based on (7), the free-space field can be determined directly, without the need for knowing the value of the ground-reflection coefficient Γ.

This is a variation of a method employed by the Federal Communications Commission in connection with one phase of their overall standards program, which includes the design [34] and calibration [35] of VHF-UHF field-strength meters.

E. Uncertainty of the NBS Standards

Error analyses have been made of the *standard-antenna* and *standard-field* methods at several frequencies in the range 30 to 1000 MHz. Results of these studies as well as direct intercomparisons between the two methods indicate the uncertainty in the standards to be approximately 1.0 dB. It is hoped that further development planned for the near future will reduce this uncertainty to 0.5 dB.

IV. Near-Zone Electric-Field-Strength Meter and Interim Field-Strength Standards (150 kHz to 30 MHz)

A. Introduction

Prototype instrumentation has recently been developed by NBS for the Defense Atomic Support Agency to measure the electric-field components of complex, high-level, near-zone electromagnetic fields [36]. This equipment is capable of measuring both the magnitude and direction of elliptically polarized CW electric fields having strengths in the range from 0.1 to 1000 volts per meter, at frequencies from 150 kHz to 30 MHz. These field-strength meters are intended primarily for use in evaluating the hazards of high-level electromagnetic radiation to electro-explosive ordnance devices on shipboard, or at other military installations. Interim field-strength standards were also developed at NBS for calibrating these *near-zone* meters, but will also serve for the calibration of certain types of *far-zone* instruments.

The design of the NBS meters is based on the use of a novel form of telemetry, employing a completely nonmetallic electrical transmission line to avoid perturbing the field being measured. The high line loss involved necessitates miniaturizing the RF portions of the receiving and calibrating instrumentation and placing them and the associated battery supplies *inside* the measuring antenna. The field information contained in the detected dc-AF output of the receiver is transmitted over the line to a remote readout unit, where the strength of the electric-field component parallel to the axis of the antenna is read directly.

B. The New "Semiconducting" Plastic Transmission Line

1) Errors Caused by Metallic Lines: In the past, electric field-strength meters have usually made use of a long *metallic* RF transmission line to connect the measuring antenna with its receiver, usually located at a point remote from the antenna. Such metallic lines often cause large measurement errors, especially when measuring near-zone fields having complex spatial distributions. In these cases, it is difficult or impossible to orient the line so that it is everywhere normal to existing electric-field components. Thus, not only may the line perturb the field being measured, but unwanted RF currents induced on the line can be coupled into the antenna and contribute to the total response of the field-strength meter.

2) The Use of a Nonmetallic Line: The previous difficulties were avoided by making use of a special nonmetallic balanced transmission line in which the conductor RF loss was purposely made extremely high compared to that of the usual copper line. If the conductors are made of sufficiently high-resistance material, the line can be made essentially "transparent" to the surrounding field. This can be achieved if the volume resistivity of the conductor material used is of the order of a million times or more higher than that of copper. This is roughly midway (on a logarithmic scale) between the volume resistivity of metals, on the one hand, and that of insulators such as glass or mica on the other. Such materials can therefore be said to be "semiconducting."[2]

3) Characteristics of the Conductor Material: The material used in the NBS lines is basically polytetrafluoroethylene (PTFE), rendered "semiconducting" by uniformly

[2] The term "semiconducting" will be used in this section to denote a very low level of conductivity when referring to the *nonmetallic* transmission line developed at NBS.

dispersing finely divided carbon black (approximately 30 percent by weight) throughout the plastic while it is in the semifluid state during its manufacture. The parallel-line conductors are in the form of 0.03-inch (0.76-millimeter) diameter monofilaments in place of the usual copper conductors. This material has a volume resistivity of approximately 3.0 ohm·centimeter (compared to 1.7×10^{-6} ohm centimeter for copper), giving a resistance of approximately 20 000 ohms per lineal foot (65 600 ohms per meter), or a loop resistance of approximately 1.2 megohms for a line 30 feet (9.14 meters) in length. The use of a transmission line of this type has been found to reduce the perturbation of the surrounding field by more than *two* orders of magnitude below that existing in the case of a copper line. This renders any effect of the line on the field wholly negligible in most instances.

C. Description of the Near-Zone Field-Strength Meters

1) Principles of Operation: Two prototype near-zone field-strength meters were developed for the frequency ranges 150 to 250 kHz and 18 to 30 MHz, respectively. These meters make use of a tuned RF type of receiver to amplify the voltage induced in the receiving antenna by the component of the electric field being measured. A tunable CW reference oscillator is included to permit resetting the overall receiver gain periodically during a sequence of measurements in the field. A block diagram of the basic circuitry is shown in Fig. 13. The meters employ the same basic method of resetting the receiver gain that has been used in CW field-strength meters in this country for many years [37].

The receiver portion of the field-strength meter consists basically of an RF input capacitive step attenuator, a fixed-tuned bandpass RF amplifier (with manual gain control), and a diode detector. The RF output level of the oscillator is monitored by its own detector. When resetting, the receiver gain is simply adjusted until the dc outputs of the two detectors are identical, as read on the remote indicator. In this method, the exact level of the CW oscillator is not important, and does not have to be known. However, it *is* important that the oscillator monitoring detector remain stable.

2) Design—The Dipole Antenna: The dipole of the near-zone field-strength meter is approximately 12 inches (30 centimeters) in overall length. The miniaturized solid-state receiver and its battery supply are contained in one half, and the solid-state CW reference oscillator and its battery supply are in the other. The balanced "semiconducting" transmission line enters the interior of the hollow dipole through a two-section balanced RC filter to help preserve the electrical symmetry of the dipole and to minimize common-mode RF pickup. The response of the dipole to the cross-polarized component of the electric field is 40 dB or more below the principal response.

The dipole is mounted on a hollow fiber-glass shaft which is perpendicular to the dipole axis. This shaft is in turn supported by a reinforced polyfoam ring, so that manual adjustments in both azimuth and elevation can be readily made in the orientation of the dipole. A view of the complete field-strength meter is shown in Fig. 14. The transmission line runs from the dipole, through the hollow shaft on the right, to the remote indicator unit shown in the foreground.

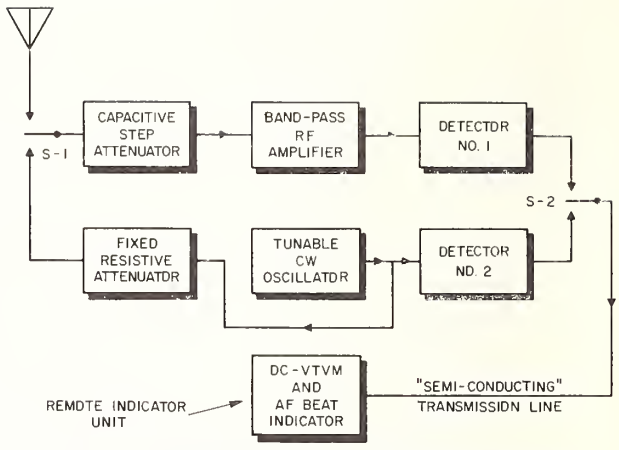


Fig. 13. Block diagram showing the basic layout of the NBS near-zone electric-field-strength meter.

Fig. 14. View of the complete near-zone electric-field-strength meter.

A measuring dipole must be short both physically and electrically if meaningful measurements are to be made in the near zone of a transmitting antenna. The physical length of the NBS dipole is about 12 inches (30 centimeters), and the electrical length about 0.03 wavelength at the highest operating frequency. Under these conditions, it has been determined at NBS that the effective length of the measuring dipole, when measuring a near-zone field having a highly complex spatial distribution, will not differ by more than 2 or 3 percent (in the worst case) from the effective length when immersed in a uniform plane wave. Likewise, any interaction error [4], resulting from coupling between such a short antenna and its image (in a nearby conducting plane,

or in a nearby radiator) will be small provided the loading is light and that measurements are not made at distances less than 2 or 3 times the length of the measuring dipole from these objects.

D. Development of the Interim Field-Strength Standards

This phase of the NBS program for the development of instrumentation for measuring near-zone electric fields involved the establishment of interim calibration standards and associated techniques for use over the frequency range from 150 kHz to 30 MHz. These were needed in order to properly evaluate the performance of the field-strength meters during their development, and to provide a preliminary calibration of the completed models. It is also planned to eventually use these standards for other purposes including the calibration of electric-field-strength meters having vertical-rod antennas.

Two different antenna configurations were selected to establish independent standard calibrating fields. These were 1) a thin cylindrical monopole over a large metallic ground plane, and 2) a relatively large parallel-plate capacitor system. An electrically short transfer (dipole) probe and field indicator were also developed for use in directly intercomparing these two types of standard fields. The present uncertainty of these standards is believed to be less than ± 2 dB. It is hoped that further development effort will reduce this to ± 1.0 dB or less and will also extend the usable frequency range [5].

V. Summary and Conclusions

A description has been given in this paper of the various CW field-strength standards and associated measurement instrumentation and techniques developed at the National Bureau of Standards during the past 25 to 30 years. Two of the standards are used as the basis of the calibration service maintained by NBS for both commercial and military CW field-strength meters of various types. This service is available over the frequency range 30 Hz to 1000 MHz. These standards and their respective uncertainties are as follows.

1) *Magnetic-Field-Strength Standards*, used over the frequency range 30 Hz to 30 MHz for the calibration of CW field-strength meters employing small-loop receiving antennas. The limit of uncertainty of the standards is believed to be ± 3 percent from 30 Hz to 5 MHz, and ± 5 percent from 5 to 30 MHz.

2) *Electric-Field-Strength Standards*, used over the frequency range 30 to 1000 MHz for the calibration of CW field-strength meters employing half-wavelength, self-resonant, dipole receiving antennas. The limit of uncertainty of these standards is believed to be ± 1.0 dB over this frequency range. It is hoped that further development effort in the near future will reduce the uncertainty to ± 0.5 dB.

In addition to the above, a prototype *Near-Zone Electric-Field-Strength Meter* and *Interim Field-Strength Standards* were recently developed by NBS for the Defense Atomic Support Agency. These are intended for use over the frequency range 150 kHz to 30 MHz for the evaluation of hazards of high-level electromagnetic radiation to ordnance devices and other uses. The limit of uncertainty of these interim standards is believed not to exceed ± 2.0 dB over this frequency range. It is hoped that further development effort will reduce the uncertainty to ± 1.0 dB and will also extend the usable frequency range.

References

[1] S. A. Schelkunoff, *Electromagnetic Waves*. Princeton, N. J.: Van Nostrand, 1943, pp. 476–479.

[2] S. A. Schelkunoff and H. T. Friis, *Antennas, Theory and Practice*. New York: Wiley, 1952, pp. 298–299.

[3] W. L. Weeks, *Electromagnetic Theory for Engineering Applications*. New York: Wiley, 1964, pp. 332–333.

[4] R. A. Bartfeld et al., "Study of methods of improved measurement for electromagnetic field components," University of Pennsylvania, Final Rept. on Contract NBy-3200, Moore School Rept. 62-19, pp. 16–57, July 31, 1962.

[5] F. M. Greene, "A new near-zone electric-field-strength meter," NBS Tech. Note 345, November 15, 1966.

[6] "IRE standards on radio wave propagation: Methods of measuring field intensity," 1942 (Supplement to *Proc. IRE*, vol. 30, pt. 2, July 1942).

[7] F. M. Greene, "Calibration of commercial radio field-strength meters at the National Bureau of Standards," NBS Circular 517, December 1951.

[8] H. E. Taggart, "Field strength calibration techniques at the National Bureau of Standards," *IEEE Trans. on Electromagnetic Compatibility*, vol. EMC-7, pp. 163–169, June 1965.

[9] "IRE standards on radio receivers: method of testing receivers employing ferrite core loop antennas," *Proc. IRE*, vol. 43, pp. 1086–1088, September 1955.

[10] F. M. Greene, "The near-zone magnetic field of a small circular-loop antenna," *J. Res. NBS*, vol. 71C, October–December 1967, to be published.

[11] G. F. Gainsborough, "Experiments with thermocouple milliammeters at very-high radio frequencies," *J. IEE* (*London*), vol. 91, pt. III, pp. 156–161, September 1944.

[12] A. G. Jean, H. E. Taggart, and J. R. Wait, "Calibration of loop antennas at VLF," *J. Res. NBS*, vol. 65C, pp. 189–193, July–September 1961.

[13] G. Glinski, "Note on circular loop antennas with non-uniform current distribution," *J. Appl. Phys.*, vol. 18, pp. 638–644, July 1947.

[14] V. S. Buzinov, "Equipment for testing frame-type field-strength measuring sets by means of a reference induction field," *Meas. Techniques*, pp. 484–487, December 1961 (English transl. of *Izmeritel'naya Tekhnika*, pp. 46–48, June 1961).

[15] H. Diamond, K. A. Norton, and E. G. Lapham, "On the accuracy of radio field-intensity measurements at broadcast frequencies," *J. Res. NBS*, vol. 21, pp. 795–818, December 1938.

[16] F. M. Greene [7], p. 3.

[17] J. H. DeWitt, Jr., and A. C. Omberg, "The relation of the carrying car to the accuracy of portable field-intensity-measuring equipment," *Proc. IRE*, vol. 27, pp. 1–4, January 1939.

[18] R. Mittra and C. O. Stearns, "Field strength measurements in a multipath field using linear and circular probing," *Radio Sci.*, vol. 2, pp. 101–110, January 1967.

[19] J. R. Wait, "Influence of source distance on the impedance characteristics of ELF radio waves," *Proc. IRE* (*Correspondence*), vol. 48, pp. 1338–1339, July 1960.

[20] H. Diamond, K. A. Norton, and E. G. Lapham, *op. cit.*, p. 811.

[21] H. C. Torrey and C. A. Whitmer, *Crystal Rectifiers*, M.I.T. Radiation Lab. Ser. New York: McGraw-Hill, 1948, pp. 333–360.

[22] H. Whiteside and R. W. P. King, "The loop antenna as a probe," *IEEE Trans. on Antennas and Propagation*, vol. AP-12, pp. 291–297, May 1964.

[23] F. M. Greene and M. Solow, "Development of very-high-frequency field-intensity standards," *J. Res. NBS*, vol. 44, pp. 527–547, May 1950.

[24] R. W. P. King, *The Theory of Linear Antennas* (with charts and tables

for practical applications). Cambridge, Mass.: Harvard University Press, 1956.
[25] F. M. Greene, "Influence of the ground on the calibration and use of VHF field-intensity meters," *J. Res. NBS*, vol. 44, pp. 123–130, February 1950; see also *Proc. IRE (Australia)*, vol. 14, pp. 58–64, March 1953.
[26] K. A. Norton, "The propagation of radio waves over the surface of the earth and in the upper atmosphere—Part II," *Proc. IRE*, vol. 25, pp. 1203–1236, September 1937.
[27] ——, "The calculation of ground-wave field intensity over a finitely conducting spherical earth," *Proc. IRE*, vol. 29, pp. 623–639, December 1941.
[28] C. R. Burrows, "Radio propagation over plane earth—field strength curves," *Bell Sys. Tech. J.*, vol. 16, pp. 45–75, January 1937.
[29] K. Bullington, "Radio propagation at frequencies above 30 megacycles," *Proc. IRE*, vol. 35, pp. 1122–1136, October 1947.
[30] G. M. Strizhkov and B. E. Rabinovich, "Measurements of current by means of thermistor bridges at frequencies up to 1000 MC," *Meas. Techniques*, pp. 797–799, October 1959 (English transl. of *Izmeritel'naya Tekhnika*, pp. 38–40, October 1959).
[31] N. Marchand, "Transmission-line conversion transformers," *Electronics*, vol. 17, pp. 142–145, December 1944.
[32] F. M. Greene and M. Solow, *op. cit.*, p. 538.
[33] A. A. Kotovich, "Errors in measuring the reflection coefficient of ground," *Meas. Techniques*, pp. 299–302, November 1961 (English transl. of *Izmeritel'naya Tekhnika*, pp. 34–37, April 1961).
[34] W. K. Roberts, "A new wide-band balun," *Proc. IRE*, vol. 45, pp. 1628–1631, December 1957.
[35] ——, "Avenues for improvement in the design and calibration of VHF-UHF noise and field strength meters," *IRE Trans. on Radio Frequency Interference*, vol. RFI-1, pp. 6–12, May 1960.
[36] F. M. Greene, "A new near-zone electric-field-strength meter," *J. Res. NBS*, vol. 71C, pp. 51–57, January–March 1967.
[37] J. V. Cosman, "Portable field-intensity meter," *Communications*, vol. 18, pp. 22–23, September 1938.
[38] R. P. Harrison and E. A. Lewis, "A method for accurately measuring the vertical electric field strength of a propagating VLF wave," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-14, pp. 89–97, March/June 1965.
[39] E. C. Jordan, *Electromagnetic Waves and Radiating Systems*. Englewood Cliffs, N. J.: Prentice-Hall, 1950, pp. 320–325.
[40] R. W. P. King and T. T. Wu, "Currents, charges, and near-fields of cylindrical antennas," *J. Res. NBS (Radio Science)*, vol. 69D, pp. 429–446, March 1965.
[41] E. Weber, *Electromagnetic Fields*. New York: Wiley, 1950, pp. 333–338.
[42] S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*. New York: Wiley, 1965.
[43] A. A. Kotovich, "Errors in standard devices for measuring field strength in the 30 to 600 Mc band," *Meas. Techniques*, pp. 570–575, September–October 1958 (English trans. of *Izmeritel'naya Tekhnika*, pp. 61–65, September–October 1958).
[44] R. L. Smith-Rose, "Radio field-strength measurement," *Proc. IEE (London)*, vol. 96, pt. III, pp. 31–36, January 1949.
[45] K. A. Norton, "The physical reality of space and surface waves in the radiation field of radio antennas," *Proc. IRE*, vol. 25, pp. 1192–1202, September 1937.

Reprinted from the

PROCEEDINGS OF THE IEEE

Volume 55, No. 6, June 1967

(pages 970-981)

FIELD STRENGTH CALIBRATION TECHNIQUES AT THE NATIONAL BUREAU OF STANDARDS

Harold E. Taggart
National Bureau of Standards
Boulder, Colorado

Summary—The techniques and instrumentation utilized by the National Bureau of Standards to calibrate field strength meters and standards in terms of absolute field strength at frequencies from 10 kHz to 1 GHz are described. Two basic techniques are used: (1) the standard-field method for loop antennas at frequencies from 30 kHz to 30 MHz and (2) the standard-antenna method for horizontally-polarized dipole antennas at frequencies from 30 to 1000 MHz. The accuracy and the limitations of each technique are discussed. Present loop antenna calibration uncertainties vary from 3 to 5 percent depending on the frequency. Dipole antenna uncertainties are presently 12 percent. Various other antenna calibration techniques that have been studied are mentioned, pointing out their limitations. The instrumentation used to calibrate the field strength receiver characteristics is described. These measurements include the calibration of the signal attenuators, calibration of the overall linearity of the receiver, and calibration of the receiver as a two-terminal rf voltmeter.

Future plans of the National Bureau of Standards field strength and noise calibration services are discussed. Improved uncertainties and extended services are planned for antennas from 30 Hz to 1000 MHz. New noise calibration services are planned; these include the calibration of random noise generators at selected frequencies from 30 to 1000 MHz, and the calibration of impulse generators in terms of their spectral density.

Introduction

A field strength calibration service has been available from the National Bureau of Standards since 1932. The initial calibrations were limited in frequency range, primarily to the broadcast band. Since that time the available field strength calibration services have been expanded and improved to include services from 10 kHz to 1000 MHz. It is the purpose of this paper to describe the calibration techniques currently used by NBS, the working standards that are used, the limits of accuracy, and the services that are available to industry and other governmental agencies.

Two general techniques of calibrating antennas will be discussed: (1) The Standard-Field Method and (2) The Standard-Antenna Method, the use of each being dependent upon the type of antenna being calibrated and the frequency. The standard-field method is used to calibrate loop antennas from 10 kHz to 30 MHz, and the standard-antenna method is used to calibrate horizontally-polarized dipole antennas from 30 to 1000 MHz. The instrumentation used to calibrate the field strength receiver characteristics will be illustrated; these measurements include the calibration of the signal attenuators, calibration of the overall linearity of the receiver, and calibration of the receiver as a two-terminal rf voltmeter.

Other measurement techniques that have been investigated will be mentioned showing inherent advantages, disadvantages, and accuracy limitations. Long range plans for field strength calibration services at NBS are briefly discussed; these will include anticipated improvements in uncertainty and extensions of frequency range.

Present Calibration Services

The present field strength calibration services include the calibration of loop and half-wave dipole antennas at frequencies from 10 kHz to 1000 MHz. All measurements are made in terms of rms cw signals. Loop antennas can be calibrated at frequencies from 10 kHz to 30 MHz, and horizontally-polarized dipole antennas at frequencies from 30 to 1000 MHz. Also available is a calibration service for the field strength receiver; this includes calibration of signal attenuators, overall linearity of the receiver, and calibration of the receiver as a two-terminal rf voltmeter.

Loop Antenna Calibrations

The standard-field method is used to calibrate loop antennas from 10 kHz to 30 MHz. Basically, this technique involves placing the antenna in a known field referred to as a standard field, thus the name standard-field method, and determining a calibration factor or antenna coefficient from the magnitude of the known field and the field strength meter reading. The standard-field method is favored over the standard-antenna method for calibrating loop antennas, because it is an accurate, convenient, and simple technique to use, and calibrations can

IEEE TRANSACTIONS ON Electromagnetic Compatibility
Vol. EMC-7, No. 2, June 1965

be performed rapidly.

Loop antennas are calibrated in terms of a free space, quasi-static magnetic field produced by a single-turn, unshielded, balanced transmitting loop of known radius carrying a known current. The magnitude of the field produced by a single-turn circular loop is given by the following equation:[1,2]

$$|E| \cong \frac{60\,\pi r_1^2 I}{(d^2 + r_1^2 + r_2^2)^{3/2}} \sqrt{1 + \left(\frac{2\pi d}{\lambda}\right)^2} \qquad (1)$$

where $|E|$ = equivalent free-space field strength in rms volts per meter,

r_1 = radius of the transmitting loop in meters,

r_2 = radius of the receiving loop in meters (if the receiving loop is rectangular, use the radius of a circle having same area),

d = axial spacing in meters between the coaxial loops,

I = transmitting loop current in rms amperes, and

λ = free space wavelength in meters.

The actual value of the quasi-static field, H, produced by the loop is expressed in terms of the equivalent electric field, E, that would exist in a free-space radiation field bv the relationship E = ZH where Z is the impedance of free space, 376.7 ohms.

The magnitude of the field is a function of frequency as indicated by equation (1); however, the induction field component under the radical which is the frequency correction term has a negligible effect at frequencies below approximately 5 MHz. Below 5 MHz this term can be omitted without appreciable error.

The loop antenna configurations are illustrated by Figure 1. The transmitting and receiving loops are positioned coaxially to each other at a spacing of 1 to 2 meters. The spacing is determined by the desired magnitude of the calibrating field and the frequency. Equation (1) is only valid for determining the equivalent free-space electric field strength when r_1, r_2, and d are electrically small compared to λ. The loop spacing should be a minimum of four times the radii of the larger of r_1 or r_2 for equation (1) to be valid within one percent.

To meet the free-space requirements of equation (1), the calibration site should be in an area that is relatively free of any sizeable metallic objects that might influence or distort

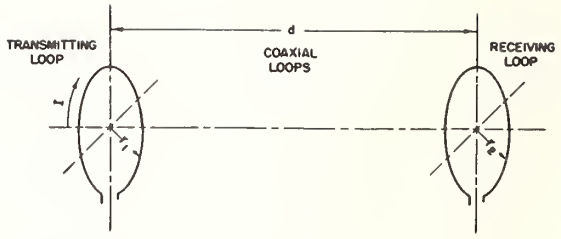


Fig. 1 — Simplified Loop Calibration Set-Up

the calibrating field. Normally, if the calibrating area is cleared of metallic objects for about two or three times the loop spacing, d, there should be no appreciable effect. The calibrating site should, therefore, be in an area relatively free of any objectionable metallic objects such as overhead power lines, shielded rooms, or steel reinforced walls. A non-metallic building with no overhead wiring makes a satisfactory calibration site. Figure 2 shows a typical loop calibration set-up at the NBS loop calibration site.

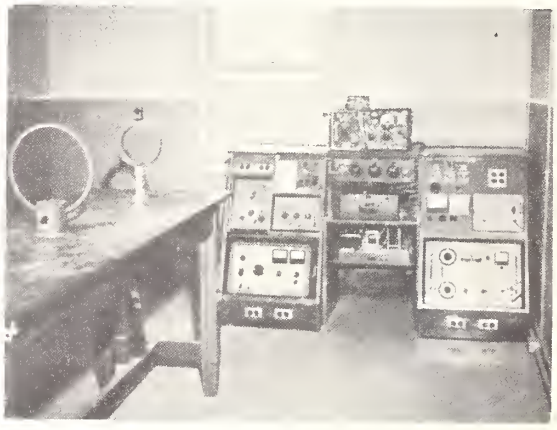


Fig. 2 — Typical Loop Calibration Set-Up

Another point of consideration is the position of the loop antenna relative to the receiver case when the loop antenna is mounted directly to the case. The receiver case will distort the measured field, but the distortion may not be the same for the quasi-static field (calibrating field) as the radiation field. To minimize this source of error, it is advisable to specify the position of the loop relative to the receiver case if the proximity of loop is within one loop diameter of the receiver case.

Loop Antenna Calibration Accuracies

The standard-field method of calibrating loop antennas has been compared with the standard-antenna method, and good agreement between the two obtained. A standard receiving loop antenna used at 10 MHz consisted of a single turn, approximately 20-cm in diameter, with a point contact silicon crystal diode built into the center of the loop and an R-C filter network connected at the terminals as illustrated in Figure 3.

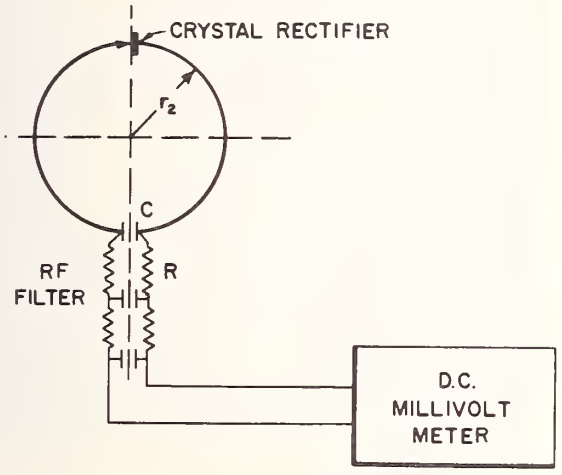


Fig. 3 — Circuit Diagram of a Standard Receiving Loop Antenna

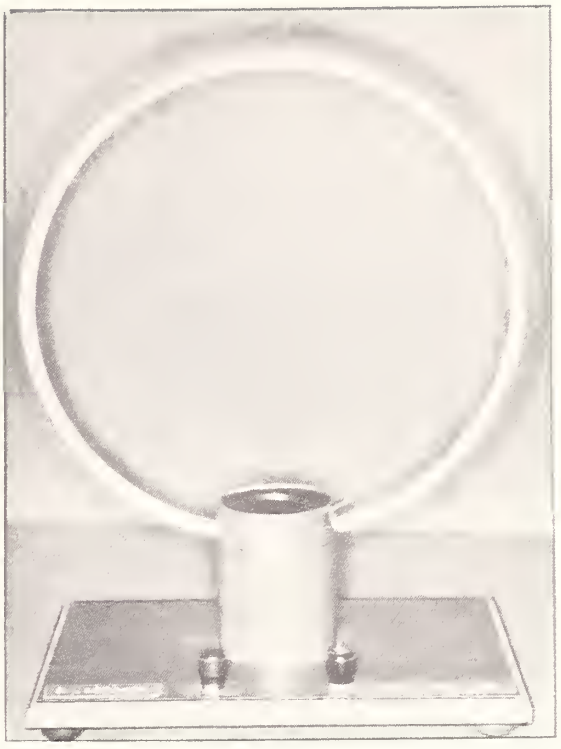


Fig. 4 — A 10 MHz Standard Receiving Loop Antenna

A 10 MHz receiving loop antenna is shown in Figure 4. The dc output of the crystal is measured with a precision, high impedance millivoltmeter. By knowing the rf-dc voltage characteristics, the open-circuit antenna voltage, V_{oc}, can be determined. The magnitude of the electric field strength can be calculated by the relationship:

$$|E| = \frac{V_{oc}}{\ell_{eff}} \quad ,$$

where ℓ_{eff} is the effective length of a single-turn loop in meters. The effective length of a single turn loop may be determined by the following equation:

$$\ell_{eff} = \frac{2\pi A}{\lambda} \quad ,$$

where A = loop area in square meters,
λ = free-space wavelength in meters.

Another technique that can be employed to calibrate balanced loop antennas is the loop-injection method. A low impedance voltage source (approximately 0.1 ohm) is used to inject a known voltage in series with the loop. By calculating the effective length of the loop and knowing the series voltage, the antenna coefficient can be determined. This method is limited to antennas of certain physical characteristics and to lower frequencies. The upper frequency limit is a few megahertz. Agreement between this method and the standard-field method is approximately 10 percent.

The diameter of the NBS transmitting loop is approximately 20 centimeters, and the loop current is normally 100 milliamperes. The current is measured by means of a vacuum thermoelement mounted in the top-center of the loop. This dc calibrated thermoelement is believed to measure rf current to well within one percent at frequencies up to 30 MHz. The output of the thermocouple is measured with a high

385-582 O - 70 - 8

quality millivoltmeter or digital voltmeter. The loop input is balanced by means of a high quality balun transformer. The balance error of this transformer with respect to ground does not exceed 2 percent. The uniformity of the loop current is also a function of the loop circumference. If the loop circumference is no greater than $\lambda/16$, the loop current will be essentially uniform. It is apparent that the current distribution presents problems only at the higher frequencies; it may be possible to drive the loop from an unbalanced source at the lower frequencies without serious accuracy degradation.

The magnitude of the NBS standard fields used to calibrate loop antennas is usually between 20 and 200 millivolts per meter. At the present time it is believed that standard fields can be established with an uncertainty of 3 percent at frequencies from 10 kHz to 5 MHz and 5 percent or less at frequencies from 5 to 30 MHz.

Dipole Antenna Calibrations

Dipole antennas are calibrated in terms of a horizontally-polarized radiated field at frequencies from 30 to 1000 MHz. The standard-antenna method is utilized for these calibrations. A field is radiated by a transmitting antenna. In the radiated field a self-resonant half-wave dipole antenna of known characteristics (called a standard-antenna herein) capable of measuring the magnitude of this radiated field is placed at some predetermined location, and the field measured. The transmitting level is carefully monitored to insure no transmitting level changes, and the antenna being calibrated is then substituted in place of the standard-antenna. The antenna coefficient is determined by the relationship of the two measurements. The standard-antenna method is used in place of the standard-field method to calibrate dipole antennas because of the inconvenience of establishing known fields using the standard field method.

The standard receiving antenna consists of a self-resonant half-wave dipole antenna with a relatively high impedance (approximately 10 K ohms) balanced voltmeter built into the center gap of the antenna. The balanced voltmeter is composed of a selected point-contact silicon crystal diode as the detector and an R-C filter network so that the dc output of the crystal can be measured with any high quality high impedance millivoltmeter. Figure 5 shows an experimental receiving dipole antenna used from 300 to 600 MHz, showing the silicon crystal diode and the R-C filter network.

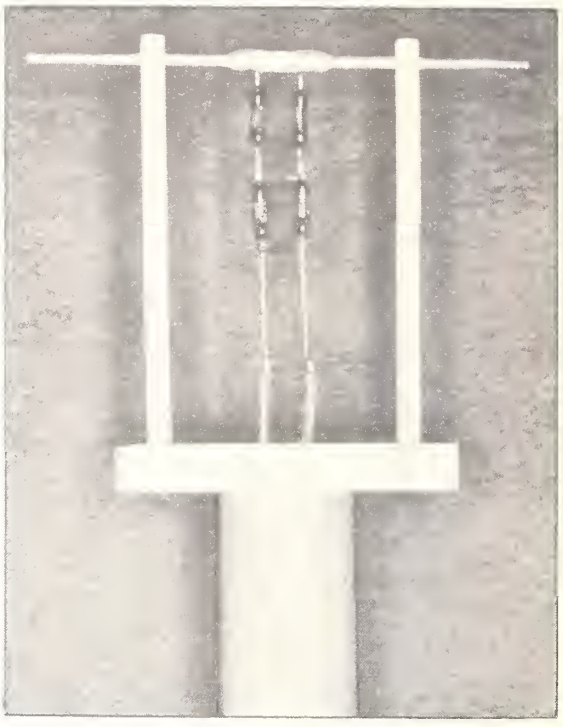


Fig. 5 - An Experimental Receiving Dipole Antenna

The crystal diodes mounted into the center gap of the receiving antenna are carefully selected to optimize characteristics such as frequency insensitivity, temperature stability, voltage sensitivity, rf impedance, and shunt capacity. Several types have been used. For the lower frequencies (30 to 400 MHz) a cartridge type mixer diode has been found to be satisfactory; the frequency characteristics of this diode are flat to within 1 percent at frequencies up to 400 MHz and within 10 percent at 1000 MHz. Above 400 MHz it is necessary to apply frequency corrections to achieve maximum accuracy. At the higher frequencies (400 to 1000 MHz) glass encapsulated, subminiature, crystal diodes are normally used because of their improved frequency response and lower shunt capacity. Crystals of this type have a frequency response that is flat to within approximately 5 percent to 1000 MHz and a shunt capacity of 0.3 pf or less. The frequency response of these crystal diodes was evaluated in a special coaxial mount.

The rf voltage calibrations of the crystals

are made in terms of a balanced rf voltmeter that has been referenced to an NBS unbalanced voltage standard. It is believed that the balanced voltmeter is accurate to within 1 or 2 percent at the levels and frequencies used. The crystal diodes are normally calibrated when mounted into the center gap of the antenna. Figure 6 shows two views of the balanced voltmeter; the crystal diode is attached to the voltmeter by means of the spring clips. The

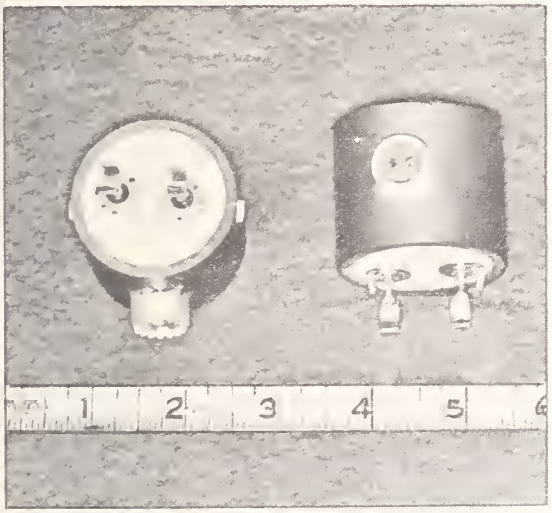


Fig. 6—Balanced Voltmeters Used to Calibrate Crystal Dipole Antennas.

calibration is accomplished by applying a known balanced voltage to the crystal and then reducing the voltage in known increments by means of an accurate step attenuator. From these data a curve showing the rf input voltage and the dc output voltage of the crystal can be drawn. Figure 7 illustrates a typical calibration curve of one of these crystals. The calibrating frequency is normally 50 MHz. If necessary, frequency corrections are applied when crystals are used at the higher frequencies.

The electric component of a field can be expressed in terms of the induced antenna voltage, V_i, and the effective length of the antenna, ℓ_{eff}. The effective length of half-wave dipole (assuming sinusoidal current distribution) can be calculated from the relationship:

$$\ell_{eff} = \frac{\lambda}{\pi} \tan \frac{\pi \ell}{\lambda} \quad ,$$

where ℓ is the antenna half-length in meters.

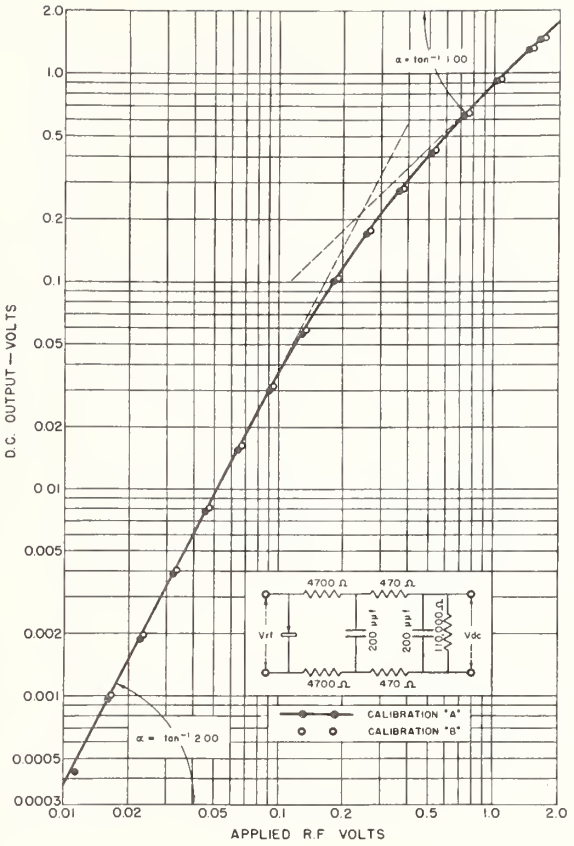


Fig. 7 — Typical Calibration Curves of the Crystal Voltmeter, Showing the Variation in dc Output Voltage Versus Applied rf Voltage for the Circuit Shown.

The induced antenna voltage, V_{oc}, is measured with a reasonable high impedance balanced voltmeter; therefore, the loading effect on the antenna is negligible and its terminal voltage is essentially the same as the induced voltage (referred to the center terminals). The magnitude of the field can be determined by the relationship:

$$|E| = \frac{V_{oc}}{\ell_{eff}} \quad ,$$

where $|E|$ = electric field strength in volts per meter,

V_{oc} = open-circuit antenna voltage in volts, and

ℓ_{eff} = effective antenna length in meters.

The loading effect of the voltmeter at the center gap of the antenna may be determined by the following relationship:

$$\left| \frac{V_L}{V_{oc}} \right| = \left| \frac{Z_L}{Z_A + Z_L} \right| ,$$

where V_L = load voltage,

V_{oc} = open-circuit voltage,

Z_L = load impedance, and

Z_A = antenna input impedance.

In order to achieve a self-resonant, half-wave, dipole antenna of finite thickness, it is necessary to shorten the overall length of the antenna. The shortening is a function of the antenna rod diameter and frequency. Figure 8 is a curve of the percent shortening versus the characteristic impedance, K_o, of a cylindrical half-wave antenna. The percent shortening may be determined by the following equation: [3]

$$\% \text{ shortening} = \frac{2708}{K_o} ,$$

where $K_o = 120 \left[\log_e \left(\frac{\lambda}{d} \right) - 1 \right]$, and

d = antenna rod diameter in centimeters.

The percent shortening of the NBS antenna is usually between 4 and 6 percent, depending on the length to diameter ratio. The transmitting antenna is a self-resonant, half-wave antenna identical in physical dimensions to the receiving antenna. The transmitting antenna is fed with a balanced input by means of a balun transformer.

All antenna measurements are performed at a nearby calibration site which meets the requirements previously mentioned for antenna calibrations. The internal characteristics of the receiver are determined in the laboratory. Figure 9 is a view of the calibration consoles

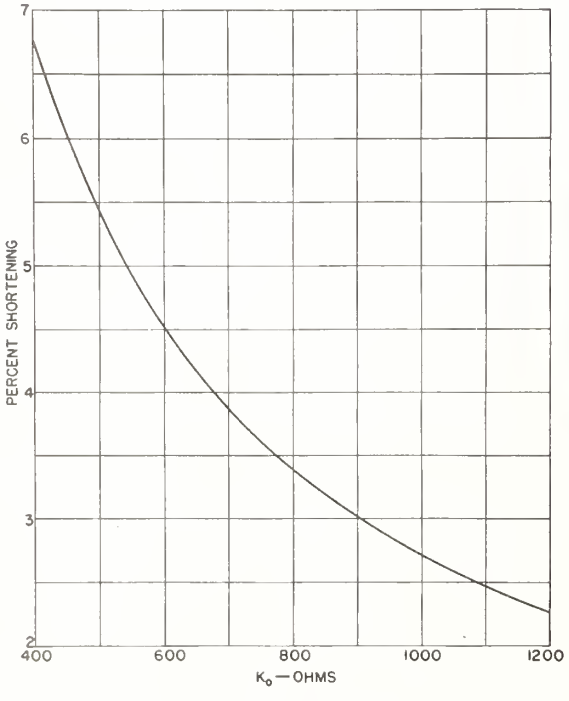


Fig. 8 — Percentage Shortening of a Cylindrical Half-Wave Dipole Required for Self-Resonant Operation Versus the Average Characteristic Impedance, K_o.

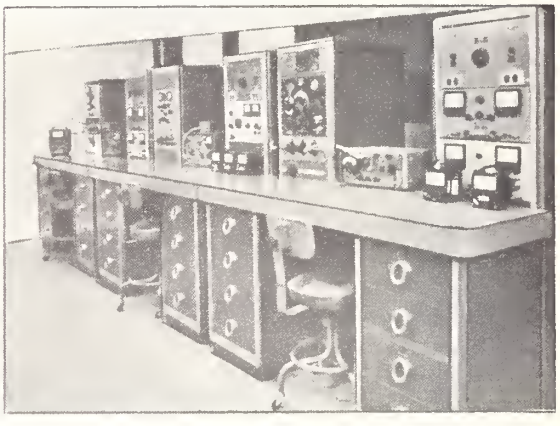


Fig. 9 — Calibrating Consoles Used to Measure Field Strength Receiver Characteristics.

used to determine attenuator ratios, overall linearity, and receiver sensitivity. These consoles consist of various signal generators, precision step attenuators, rf micropotentiometers, and frequency meters to perform measurements from dc to 1000 MHz. Attenuation and linearity measurements are performed by direct substitution, employing resistive type standard attenuators. Voltage measurements are made in terms of rf micropotentiometers.

Accuracy of Dipole Calibrations

The standard-antenna method of calibrating

dipole antennas has been compared with the standard-field method at selected frequencies to 1000 MHz in order to determine the uncertainty. The agreement between the two methods[4] varied from 5 percent at the lower frequencies to 10 percent at the higher frequencies. The establishment of accurately known standard-fields at these frequencies is difficult and requires a large flat area (several acres) without interfering objects such as trees or power lines. This type of location is often not readily available nearby. The standard-antenna method of calibrating dipole antennas is easier to perform, and the calibrating site requirements are less stringent.

Other techniques that have been investigated and compared with the above-mentioned techniques include the measurement of the receiving antenna current, and the determination of the electric field strength in terms of the antenna current, dipole impedance, load impedance, and effective length. Thermoelements are used to measure the antenna current. The measurements using this technique are in reasonably good agreement with the standard-antenna method, but the technique has certain inherent disadvantages and is somewhat difficult to accomplish.

Another comparison was made by calibrating the antennas in terms of receiver accuracy as a two-terminal rf voltmeter, antenna balun transformer characteristics, and the assumption that the antenna was ideal. This method is easy to perform, because the measurements can be performed in the laboratory, but the results are not very encouraging. Agreement with the standard-antenna method was no better than 15 to 25 percent at some of the higher frequencies.

Antennas submitted for calibration are normally calibrated at a height of 10 feet above the ground; the presence of the ground will have an effect on the antenna impedance and thus an effect on the antenna coefficient. Therefore, it is recommended that the antenna be used at a height of 10 feet or more. If used at lower heights it is advisable to use additional corrections.[5] It has been found that the presence of average ground does not effect the antenna impedance more than 5 percent for antenna heights of 0.65 wavelength or greater.

At the present time a calibration service for horizontally-polarized half-wave dipole antennas is offered at frequencies from 30 MHz to 1000 MHz with a calibration uncertainty of 12 percent. Although the calibrations are performed with horizontal polarization, it is believed to be substantially the same for vertical polarization.

Future Plans

Long range planning includes improvements in the uncertainties of the present field strength calibration services, as well as the extension of the frequency range, and range of magnitudes available. Also planned are the eventual introduction of new calibration services in both field strength and related areas.

Efforts are continuing to improve the uncertainty of loop antenna calibrations over the frequency range 10 kHz to 30 MHz, and eventual extension of the frequency range downward from 10 kHz to the vicinity of 30 Hz. Improvements are anticipated in the uncertainties of dipole antenna calibrations over the frequency range 30 to 1000 MHz. Other new areas are being investigated for possible calibration services; these areas include vertical rod antennas and electric and magnetic probes for high-level, near-field measurements.

Long range plans in related fields include some new services for noise measuring devices. These plans include a service for the calibration of random noise generators in the 30 to 1000 MHz frequency range, and eventually expanding this service to include measurements of effective input noise temperature and the calibration of impulse generators.

Acknowledgment

The author wishes to acknowledge the work of Frank M. Greene and others in the High-Frequency Field-Strength Standards Project as referenced herein for development of the basic measurement techniques and standards used for the calibration of antennas, and to express appreciation for the constructive criticism of this paper.

References

1. Greene, F. M., Calibration of Commercial Radio Field Strength Meters at the National Bureau of Standards, NBS Circ. 517, 1951.
2. Jean, A. G., H. E. Taggart, and J. R. Wait, Calibration of Loop Antennas at VLF, NBS J. Research, Vol. 64, No. 3, 1961.
3. Schelkunoff, S. A., Theory of Antennas of Arbitrary Size and Shape, Proceedings I. R. E., Vol. 29, No. 9, 1941.
4. Greene, F. M., and M. Solow, Development of Very-High-Frequency Field Intensity Standards, NBS J. Research, Vol. 44, 1950.
5. Greene, F. M., Influence of the Ground on the Calibration and Use of VHF Field Intensity Meters, NBS J. Research, Vol. 44, 1950.

Addendum: Taggart, H. E., Field Strength and RFI Standards at the National Bureau of Standards, 1968 IEEE EMC Symposium Record, IEEE 68C12-EMC, July 1968.

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 71C, No. 1, January–March 1967

A New Near-Zone Electric-Field-Strength Meter*

Frank M. Greene **

Institute for Basic Standards, National Bureau of Standards, Boulder, Colo. 80302

(May 17, 1966)

The National Bureau of Standards has recently completed the development of prototype instrumentation for measuring the electric-field components of complex, high-level, near-zone electromagnetic fields. The instrumentation is intended for use in evaluating hazards of high-level electromagnetic radiation to electroexplosive ordnance devices at military installations. The measuring range is from 0.1 to 1000 V per meter, at frequencies from 150 kHz to 30 MHz, with a present uncertainty of less than ±2 dB.

The design of the NBS meters is based on the use of a novel form of telemetry, which apparently has not been fully exploited heretofore. This involves the use of a completely nonmetallic electrical transmission line over which the field information is transmitted from the measuring antenna to a remote readout unit. The line is essentially "transparent" to the field being measured, and reduces the perturbation of the field two orders of magnitude below that normally experienced when using a metallic line. The high r-f line loss involved necessitates miniaturizing the r-f portions of the receiving and calibrating instrumentation and placing them and their associated battery supplies *inside* the measuring antenna. The design and performance of the meters are discussed in some detail.

Key Words: Device, electroexplosive ordnance; field, near-zone electromagnetic; hazards, electromagnetic radiation; line, nonmetallic electrical transmission; line, semiconducting plastic transmission; meter, electric field-strength; telemetry, novel form of.

1. Introduction

The National Bureau of Standards, under the sponsorship of the Defense Atomic Support Agency, has recently completed the development of prototype instrumentation for measuring the electric-field components of complex, high-level, near-zone electromagnetic fields. The instrumentation is capable of measuring both the magnitude and direction of elliptically polarized CW electric fields having strengths in the range from 0.1 to 1000 V per meter at frequencies from 150 kHz to 30 MHz. This represents one step in the effort to develop accurate and meaningful measuring and calibrating instrumentation, for use in evaluating hazards of electromagnetic radiation to electroexplosive ordnance devices. It is well known that high-level fields of powerful nearby radio transmitters may cause premature detonation of missile and rocket type weapons during storage or loading operations on shipboard, or at other military installations. It was from the standpoint of improving both the weapons reliability, as well as the safety of operating personnel and equipment, that the program was undertaken.

The design of the NBS meters is based on the use of a novel form of telemetry, employing a completely nonmetallic electrical transmission line to avoid perturbing the field being measured. This approach has apparently not been fully exploited, heretofore. The extremely high r-f loss of the line used attenuates any r-f currents induced on the line by the surrounding field and essentially eliminates reradiation from the line, or unwanted coupling between the line and the measuring antenna. The high line-loss, however, necessitates miniaturizing the r-f portions of the receiving and calibrating instrumentation and placing them and the associated battery supplies *inside* the measuring antenna. The field information contained in the detected DC-AF output of the receiver is transmitted over the line to a remote readout unit, where the strength of the electric-field component parallel to the axis of the antenna is read directly.

Interim electric-field standards and calibration techniques were also develped at NBS to evaluate the performance of the field-strength meters during their development, as well as to provide a tentative calibration of the completed instruments. The present uncertainty of the standards is believed to be less than ±2 dB, but further development effort is expected to reduce this to less than ±1 dB.

2. The New "Semiconducting" Plastic Transmission Line

2.1. Errors Caused by Metallic Lines

In the past, electric field-strength meters have usually made use of a long *metallic* r-f transmission

*For complete design details the reader is referred to "A New Near-Zone Electric-Field-Strength Meter"—Frank M. Greene, National Bureau of Standards Technical Note No. 345, November 15, 1966. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. Price 35 cents.

**Radio Standards Engineering Division, National Bureau of Standards, Boulder, Colo. 80302.

line to connect the measuring antenna with its receiver, usually located at a point remote from the antenna. Such metallic lines often cause large measurement errors, especially when measuring near-zone fields having complex spatial distributions. In these cases, a neutral-field path, along which the line can be placed, either seldom exists, or the orientation of such a path, if it *does* exist is usually not known in advance of making the measurements. Under these conditions, it is difficult or impossible to orient the line so that it is everywhere normal to existing electric-field components. Thus, the line not only may perturb the field being measured, but unwanted r-f currents induced on the line can be coupled into the antenna and contribute to the total response of the field-strength meter.

2.2. The Use of a Nonmetallic Line

The NBS near-zone meters avoid this difficulty by making use of a special nonmetallic balanced transmission line in which the conductor r-f loss has been purposely made extremely high compared to that of the usual copper line. If the conductors are made of sufficiently high resistance material, the line can be made essentially "transparent" to the surrounding field. This is to say that the r-f currents induced on the transmission line by the field will be negligibly small, resulting in insignificant reradiation, or that any r-f energy propagating along the line will be heavily attenuated because of the extremely high loss of the line. It has been found that this can be achieved if the volume resistivity of the conductor material used is of the order of a million times or more higher than that of copper. This is roughly midway (on a logarithmic scale) between the volume resistivity of metals, on the one hand, and that of insulators such as glass or mica on the other. Such materials can therefore be said to be "semiconducting."

2.3. Description of the "Semiconducting" Line

The material used in the NBS lines is basically polytetrafluoroethylene (PTFE), rendered "semiconducting" by uniformly dispersing finely divided carbon black (approximately 30 percent by weight) throughout the plastic while it is in the semifluid state during its manufacture. Tests at NBS have indicated the PTFE material to be more stable electrically than other types tried, such as silicone rubber, polyethylene, carbon-coated multifiber glass, etc. Other types of conductor materials (including high-resistance metallic alloy wire and thin deposited metallic films) were examined, but were not found to be practical for this use, either because of unsuitable electrical characteristics or excessive cost.

Several parallel-conductor balanced transmission lines were made at NBS using the semiconducting PTFE material. The line conductors are in the form of 0.03 in diam monofilaments in place of the usual copper conductors. They are coated with a thin nylon film (approximately 0.005 in thick) to further improve their strength and electrical stability. Each of the filaments is in turn enclosed in a vinyl-coated woven fiber glass sleeve approximately 0.1 in O.D. Two such insulated conductors are encased in a polyvinyl chloride outer jacket, for added strength and protection, to form the completed transmission line. Military Type JJ–048 or PJ–291 twin-conductor plugs are used at the two ends for connecting to the dipole antenna, and to the remote indicator unit, respectively. The overall length of the lines is approximately 30 ft.

2.4. Electrical Characteristics of the Line

The semiconducting PTFE material used in the line has a volume resistivity of approximately 3.0 Ω-cm (compared to 1.7×10^{-6} Ω-cm for copper), giving a resistance of approximately 20,000 Ω per lineal foot, or a loop resistance of approximately 1.2 MΩ for a line 30 ft in length. The measured mutual capacitance between the two line conductors is approximately 10 pF per foot, or roughly 300 pF for the entire line. The transmission loss, to the desired field-strength information at dc and low af, is negligible if a high-impedance d-c a-f vacuum-tube voltmeter is used as the readout device at the receiving end of the line. The line attenuation increases as the square-root of frequency, reaching a theoretical as well as a measured value of approximately 53 dB per foot at a frequency of 30 MHz, for the differential mode of transmission. The common-mode attenuation with the line lying flat on a metal ground plane turns out to have essentially the same value. The use of a high-loss, parallel-conductor, balanced transmission line of this type has been found to reduce the perturbation of the surrounding field by more than *two* orders of magnitude below that existing in the case of a copper line. This renders any effect of the line on the field wholly negligible in most instances. A type of "flexural" noise is generated on the transmission line when it is suddenly flexed or moved physically. The common-mode component of this "noise" may be relatively high, having a peak value of several volts on open circuit. However, the differential-mode component is appreciably smaller and does not present a problem in the NBS meters at the level of d-c transmission used (see table 1).

3. Description of the Near-Zone Field-Strength Meters

3.1. Principles of Operation

Two prototype near-zone field-strength meters were developed for the Defense Atomic Support Agency, and cover the frequency ranges 150 to 250 kHz, and 18 to 30 MHz, respectively. These meters make use of a tuned r-f type of receiver to amplify the voltage induced in the receiving antenna by the component of

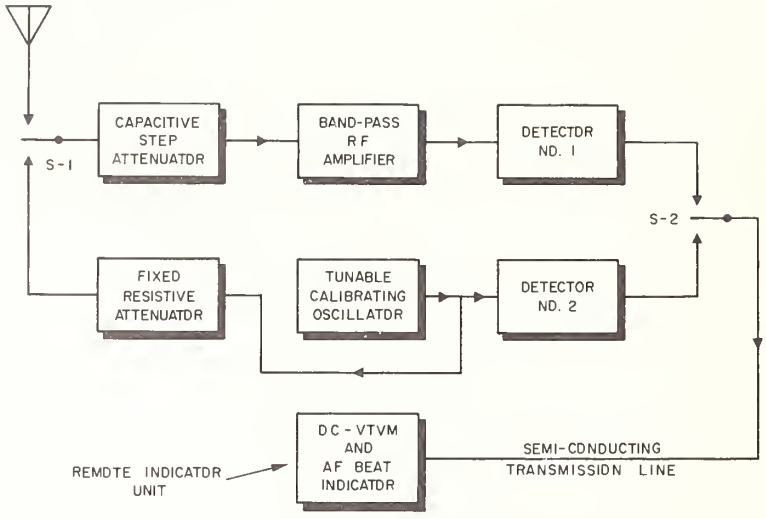


FIGURE 1. *Block diagram of the NBS field-strength meter.*

the electric field being measured. A tunable CW oscillator is included to permit calibrating the overall receiver gain periodically during a sequence of measurements in the field. A block diagram of the basic circuitry is shown in figure 1. The meters employ the same basic method of receiver-gain calibration that has been used in CW-type field-strength meters in this country for many years. The receiver portion of the field-strength meter consists basically of an r-f input capacitive step attenuator, a fixed-tuned bandpass r-f amplifier (with manual gain control), and a diode detector. The r-f output level of the oscillator is monitored by its own detector. When calibrating, the receiver gain is simply adjusted until the d-c outputs of the two detectors are identical. In this method, the exact level of the calibrating signal is not important, and does not have to be known. However, it *is* important that the oscillator monitoring detector remain stable.

TABLE 1. *Distribution of nominal operating rf-voltage levels (in the TRF field-strength meter)*

E(V/m)	α(dB)	V_i(volts)	V_0(volts)	V_1(volts)	V_2(volts)	V_3(volts)
0.1–1.0	0	0.015–0.15	0.003–0.03	0.003–0.03	0.3–3.0	0.75–7.5
1.0–10	20	0.15–1.50	0.03–0.30	0.003–0.03	0.3–3.0	0.75–7.5
10.–10^2	40	1.50–15.0	0.30–3.00	0.003–0.03	0.3–3.0	0.75–7.5
10^2–10^3	60	15.0–150.	3.00–30.0	0.003–0.03	0.3–3.0	0.75–7.5

E(V/m) ≡ Electric field strength, volts per meter
α(dB) ≡ Measuring dipole attenuator setting
V_i ≡ Induced voltage in measuring dipole
V_0 ≡ Output voltage of measuring dipole
V_1 ≡ Input voltage to rf amplifier
V_2 ≡ Output voltage of rf amplifier
V_3 ≡ Output voltage of amplifier detector (DC).

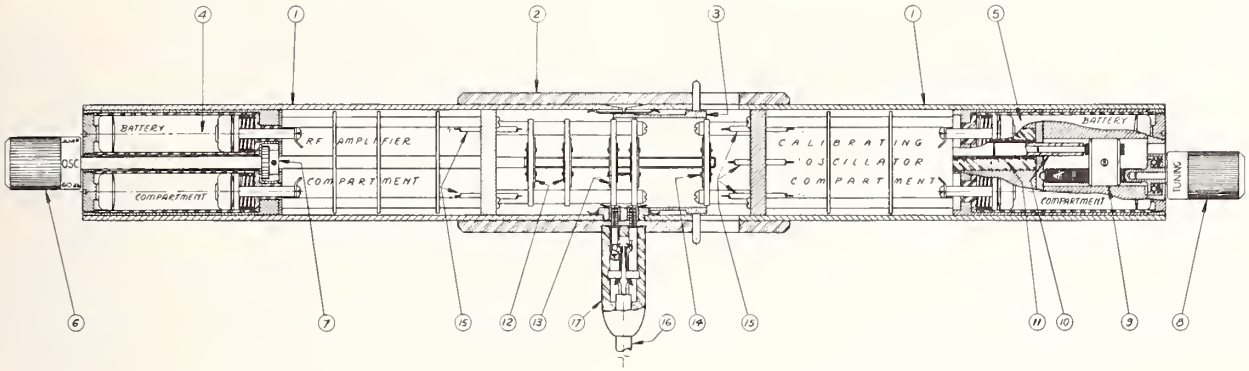
A list of the various ranges of r-f voltage over which the principal components operate is shown in table 1 for the range of electric field strength from 0.1 to 1000 V per meter. An explanation of how the ranges were calculated is given in the following:

Effective length of measuring dipole..., $L_{\text{eff}} \cong 0.15$ m
Induced voltage in dipole....., $V_i = E \cdot L_{\text{eff}}$
Dipole voltage-transfer ratio, $V_0/V_i = C_a/(C_a+C_s) \cong 0.20$
$C_a \equiv$ Internal capacitance of dipole....., $C_a \cong 3$ pF
$C_s \equiv$ Residual gap and circuit capacitance, $C_s \cong 12$ pF
RF amplifier nominal voltage gain....., $V_2/V_1 \cong 100$.

3.2. Electrical Design

The Dipole Antenna. A cutaway view of the dipole assembly is shown in figure 2, and a view of an assembled unit is shown in figure 3. The dipole is made of 1/16 in wall copper tubing 1 3/8 in O.D. The length is approximately 13 in with a 1/32 in insulated gap at the center. The miniaturized solid-state receiver and its battery supply are contained in the left half, and the CW calibrating oscillator and its battery supply are in the right half. The capacitive step attenuator, the DPDT line-transfer switch, and the oscillator ON-OFF control comprise the five rotary switch wafers seen in the central portion of the dipole assembly. The balanced transmission-line plug connects to two insulated stainless steel pins at the center of the dipole. The line enters the interior of the dipole through a two-section balanced RC filter to help preserve the electrical symmetry or balance of the dipole with respect to the transmission line, and to attenuate any differential, as well as common-mode r-f pickup that might exist on the line. The symmetry is important if the antenna is to respond only to the component of the electric field parallel to its axis. The symmetry of the NBS dipole is such that its response to the cross-polarized component of the electric field is 40 dB or more below the principal response.

It can be shown that if the length of the measuring dipole does not exceed 0.03 wavelength, its effective



1. DIPOLE ANTENNA CYLINDER,
2. CENTER-SUPPORT SLEEVE,
3. DIPOLE-SHORTING SLEEVE WITH CONTACT FINGERS,
4. BATTERY COMPARTMENT (AMPLIFIER),
5. BATTERY COMPARTMENT (OSCILLATOR),
6. R.F. ATTENUATOR CONTROL,
7. OFFSET-SHAFT GEARS,
8. TUNING CONTROL, CALIBRATING OSCILLATOR,
9. OSCILLATOR TUNING MECHANISM,
10. FERRITE TUNING SLUG
11. R.F. INDUCTOR,
12. ATTENUATOR WAFER SWITCHES,
13. DETECTOR D.C.-TRANSFER WAFER SWITCHES,
14. OSCILLATOR ON-OFF WAFER SWITCH,
15. MINIATURE COAXIAL LINES (TYP),
16. D.C. TRANSMISSION LINE, BALANCED,
17. PLUG CONNECTOR, TWIN, FEMALE.

AMPLIFIER-END VIEW

OSCILLATOR-END VIEW

FIGURE 2. *Cutaway drawing of the measuring dipole.*

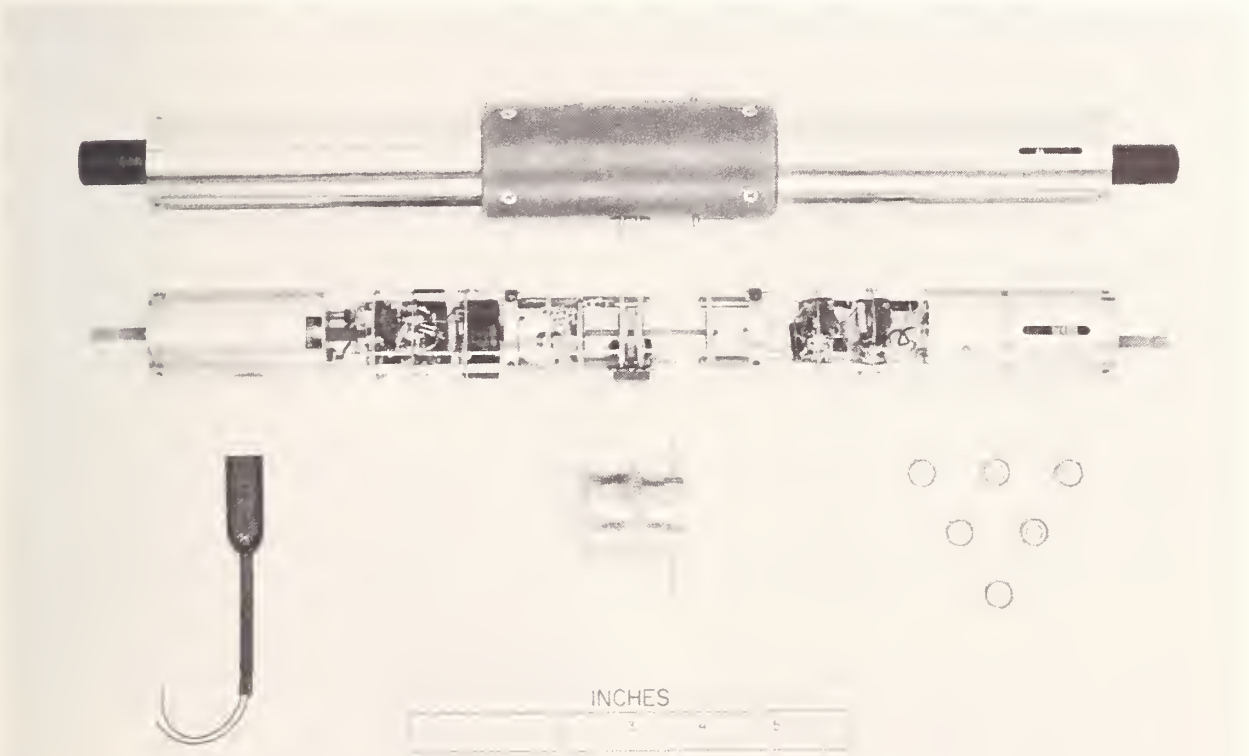


FIGURE 3. *View of the complete measuring dipole.*

length when measuring a field having a highly complex spatial distribution will not differ by more than 2 or 3 percent (in the worst case) from the effective length when immersed in a uniform plane wave. Likewise, the interaction error, resulting from coupling between such a short antenna and its image, will be small provided the loading is light and that measurements are not made closer than 2 or 3 ft to the ground or large metallic objects.

RF Capacitive Step Attenuator. This attenuator has four principal steps, 0, 20, 40, and 60 dB, with two additional 0 dB positions, calibrate A(OSC) and calibrate B(AMP). In the latter two positions, the calibrating oscillator is automatically turned ON, and the "semiconducting" transmission line can then be switched between the d-c outputs of the oscillator and amplifier detectors, while calibrating the amplifier gain. The attenuator is constructed using special subminiature rotary wafer switches. Miniature low-temperature coefficient ceramic capacitors are used in either an "L" or a "Pi" configuration for the various attenuator pads. The attenuator steps are accurate to ± 0.1 dB. The attenuator module is approximately 1 in O.D. by 1 in long.

Band-Pass Amplifier and Detector. In order to provide a high input impedance, so as not to appreciably load the dipole antenna or the capacitive step attenuator, a metal-oxide-silicon field-effect transistor (MOS–FET) is used in the first stage as an untuned input-impedance converter. This is followed by a six-pole, fixed-tuned, bandpass filter having a 3 dB pass-band of 150 to 250 kHz, and 18 to 30 MHz respectively in the two units and a 40 dB pass-band of 100 to 300 kHz and 12 to 36 MHz, respectively. The filter is followed by a 3-stage r-f amplifier. In the low-frequency unit, this is broadband resistance-coupled, using overall negative r-f feedback to improve gain stability with respect to temperature and supply-voltage variations. In the high-frequency unit, the r-f amplifier has 3 stages using wide-band synchronously tuned interstage transformers. The overall gain of both amplifiers can be manually adjusted to the required operating value of approximately 40 dB. The output detectors of both amplifiers employ two stable silicon diodes in a voltage-doubling circuit, to ease voltage-swing requirements on the final stage and consequently to reduce battery drain. A 12 V silver-oxide battery supply is used, having a useful life of from 10 to 20 hr. The complete amplifier module measures 1¼ in O.D. by 5 in long including its filter and battery supply.

Calibrating Oscillator and Detector. This unit is also transistorized. It consists of a permeability-tuned Clapp-Colpitts oscillator followed by a two-stage automatic-level-control amplifier to maintain the r-f output essentially constant over its tuning range. The tuning range of each of the two oscillators is the same as the 3 dB bandwidth of the band-pass amplifier with which it is used. The r-f output of the oscillator is adjustable to roughly 3.0 V and fed into a fixed 30 dB resistive attenuator pad. When calibrating the gain of the band-pass amplifier, the output of approximately 0.1 V from the pad is applied to the amplifier input through a small variable capacitor (adjusted to approximately 3 pF) which simulates the dipole probe capacitance. The detector circuit is identical to that used with the band-pass amplifiers and is used to monitor the 3.0 V oscillator level at the input terminals of the 30 dB pad. A 13.5 V silver-oxide battery supply is used, having a useful life of about 20 hr. A portion of the supply is Zener regulated and applied to the transistor base circuits for improved amplitude stability. The complete calibrating oscillator module measures 1¼ in O.D. by 5 in long including the battery supply.

The Remote Indicator Unit. This unit consists basically of a battery-operated, balanced d-c vacuum-tube voltmeter, with its d-c microammeter, on which the magnitude of the electric field strength is indicated. This VTVM has exceptionally good stability and linearity of response because of the large amount of negative d-c feedback employed in its design. It is used over a single 20 dB input range of approximately 0.75 to 7.5 V. The balanced input circuitry provides from 40 to 60 dB rejection of common mode input in the 60 to 400 Hz power-frequency range. In addition, a two-section balanced r-f filter provides at least 40 dB suppression of common and differential-mode input at the higher r-f frequencies. An output circuit in the VTVM is used to drive a transistorized audio amplifier and loud speaker. These units are used as a zero-beat indicator when adjusting the frequency of the calibrating oscillator to that of the field being measured. Separate, self-contained battery supplies are used with the VTVM and a-f amplifier. The complete remote indicator is housed in a drawn-aluminum case, using a tight-fitting panel to minimize r-f leakage into the unit. A small drawn-aluminum case is used to enclose the rear side of the d-c microammeter to reduce r-f entry through the meter face.

The Assembled Field-Strength Meter. The measuring dipole of the near-zone field-strength meter is mounted on a hollow fiber-glass shaft which is perpendicular to its axis. This shaft is in turn supported in plastic bearings in a circular reinforced polyfoam gimbal, approximately 21 in. in diameter, as shown in figure 4. This assembly is mounted on a sturdy wooden surveyor's tripod, in such a manner that manual adjustments in both azimuth and elevation can be readily made in the orientation of the dipole. A view of the complete field-strength meter is shown in figure 5. The balanced transmission line connects to the dipole through the hollow shaft on the right. The other end of the 30 ft line connects to the remote indicator unit, a closeup view of which is shown in figure 6.

3.3. Performance

The two near-zone field-strength meters underwent extensive field tests at the U.S. Naval Weapons Laboratory at Dahlgren, Va. Electric field strengths were

FIGURE 4. *View of the measuring dipole and mounting gimbal.*

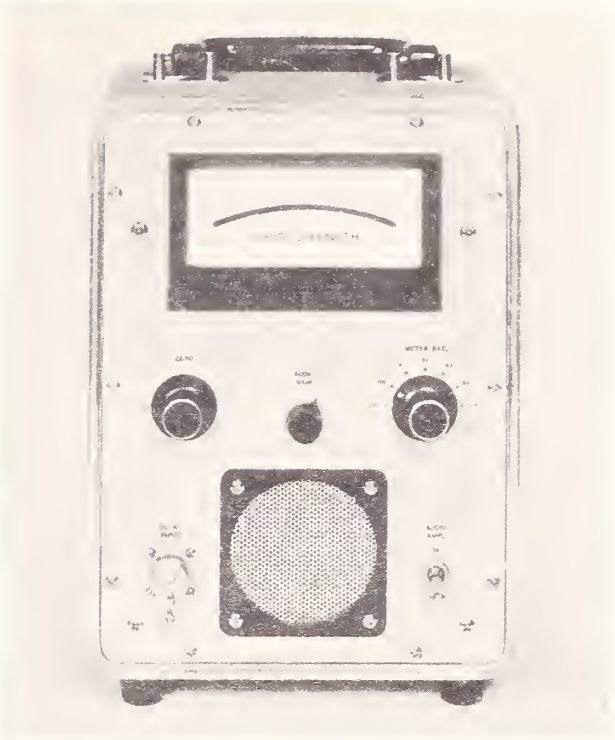


FIGURE 6. *View of the remote indicator unit.*

FIGURE 5. *The complete near-zone field-strength meter.*

measured without difficulty over the entire range from 0.1 to 1000 V per meter (CW-RMS). There was no observable susceptibility to out-of-band CW signals having levels as high as 200 V per meter, or to pulsed signals from nearby high-powered radar transmitters. No interaction was observed between the semiconducting transmission line and the field being measured. In general, the agreement was about ± 1 dB between the instrument readings as calibrated in the NBS standard field at Boulder and those obtained in the NWL standard field at Dahlgren. The uncertainty in the calibration accuracy of the instruments is believed to be less than ± 2 dB. Further development effort on the standard fields is expected to reduce this to less than ± 1 dB.

The work described in this paper was performed for the Field Command, Defense Atomic Support Agency, under DASA Order No. HD-1102-3171-0001 (June 20, 1963).

The author wishes to acknowledge the valuable assistance of William E. Jessen of NBS-BL in the assembly and performance evaluation of the instrumentation described herein. The assistance of Ezra B. Larsen in the standard-field development work is also gratefully acknowledged.

4. Bibliography

[1] Stoker, W. C., et al., An investigation of electromagnetic coupling devices for the measurement of noise fields, Final Report, Rensselaer Polytechnic Institute, Troy, N.Y., Signal Corps Contract No. DA 36–039 sc–207 (Sept. 1952).

[2] Abromavage, M. M., et al., Electromagnetic coupling to ordnance systems, Report No. N178–7604, August 1961, Jansky & Bailey Div., Atlantic Research Corp., Alexandria, Va.

[3] Bartfeld, R. A., et al., Study of methods of improved measurement for electromagnetic field components, Final Report, 26 May 1960 to 30 June 1962, University of Pennsylvania, Philadelphia, Pa., USNCEL Contract NBy–3200.

[4] Cowles, W. W., et al., Study of methods of implementing Poynting-vector measurements, Final Report 1 July 1962 to 31 December 1964, University of Pennsylvania, Philadelphia, Pa., USNCEL Contract NBy–32219.

[5] Harrison, R. P., and Lewis, E. A., A method for accurately measuring the vertical electric field strength of a propagating VLF wave, IEEE Trans. on I&M, **IM–14**, P1/2, pp. 89–97 (Mar.–June 1965).

[6] Clark, D. B., et al., Electromagnetic field-strength measurements in a quasi-absolute monopole field, Final Report August 1963 to October 1964, TR–R–361, U.S. Naval Civil Engineering Lab., Port Hueneme, Calif. (Mar. 1965).

[7] Cowdell, R. B., HERO ground-plane measurements and development of the near-field field-intensity meter, Final Report No. 2667–7191, 30 April 1965, and Addendum No. 1, July 1965, Genistron, Incorporated, College Park, Md.

(Paper 71C1–443)

ADDENDUM: Greene, F. M., The Near-Zone Magnetic Field of a Small Circular-Loop Antenna, J. Res. NBS 71C, No. 4, 319, Oct.-Dec. 1967. (Mathematical basis for NBS calibration standards.)

3.a. **Discussion of errors in gain measurements of standard electromagnetic horns,** R. W. Beatty, NBS Tech. Note 351 (March 1967).

In setting up a calibration service for measuring the gain of standard electromagnetic horns, one needs a reference horn in which one has developed a high degree of confidence. Although it is possible to calculate the gain of horns of certain design, confidence can be increased by carefully measuring the gain. This note examines a method for measuring the gain of two identical horns, listing the assumptions made in making such a measurement. The theory of 2-port waveguide junctions is applied to the analysis of the measurement technique. The method is shown to be essentially an attenuation measurement which has additional sources of error. Although these errors are not analyzed and evaluated in this note, the problem is perhaps more clearly stated than it was previously. The mismatch error in comparing two horns as receiving antennas is analyzed. Data is given on the aperture efficiency of standard horns which indicates that improvements in the design of such horns are feasible. It is concluded that, at present, an uncertainty limit of the order of tenths of decibels seems realistic, but hundredths of decibels seems unattainable until further refinements are made both in the standard horns themselves and in the measurement techniques.

3.b. **Theory of diffraction in microwave interferometry,** D. M. Kerns and E. S. Dayhoff, J. Res. NBS 64B (Math. & Math. Physics), No. 1, 1-13 (Jan.-Mar. 1960).

Microwave Michelson and Fabry-Perot interferometers are respectively considered as instances of: (1) A "reflection system", consisting of a radiating-receiving system and a reflecting object (e.g., a finite mirror); and (2) a "transmission system", consisting of a radiating system and a receiving system with an object (e.g., a Fabry-Perot etalon) interposed. The basic theoretical objective is the calculation of the amplitude and phase of the (time-harmonic) received signal in the systems considered. The electromagnetic field in space transmission paths is represented in terms of continuous angular spectra of vectorial plane waves, and the elements of the systems are described by means of suitable tensor scattering matrices (having both discrete and continuous indices). Needed scattering matrices are considered known; relationships to experimentally determinable data are outlined. The general case of either the reflection or transmission system is soluble formally in terms of a series of integrals stemming from the Liouville-Neumann series solution of certain integral equations. Formulas are obtained for models of the Michelson and Fabry-Perot instruments with arbitrary radiating and receiving characteristics. The theory and various features of the instruments considered, including Fresnel-region (or quasi-optical) behavior, are illustrated by means of examples obtained by choosing relatively simple and rather hypothetical analytical expressions for the radiating and receiving characteristics.

3.c. **Calibration of loop antennas at VLF,** A. G. Jean, H. E. Taggart, and J. R. Wait, J. Res. NBS 65C (Eng. & Instr.), No. 3, 189-193 (July-Sept. 1961).

A technique and the equipment used for the precise determination of field strength of signals received from VLF transmitters is described. The equipment, which is battery-operated, contains provisions for the reception of VLF signals and the generation of standard fields to an accuracy of 5 percent. Both the receiving and transmitting antennas are loops. The field strength is determined in terms of a quasi-static magnetic field with the two loop antennas positioned coaxially at a spacing of approximately two meters. Although the technique was developed for use at VLF, it can be used at higher frequencies for calibrating loop antennas, generators, and voltmeters, and for determining effective heights of antennas, or similar applications.

See also 6.A.

4. Radar and Baseband Pulses

Papers

Abstracts

The Measurement of Baseband Pulse Rise Times of Less than 10^{-9} Second

N. S. NAHMAN, SENIOR MEMBER, IEEE

Abstract—**This is a review paper dealing with the measurement of fractional nanosecond pulse rise time in which the following subjects are discussed: oscillographic systems, pulse comparison techniques, a basic instrumentation system, and the distortion of pulses by transmission lines. Extensive references are provided. Included in the discussion is a delineation of equivalent-time oscillographic sampling systems and a classification into three sampling categories: sequential, random, and multiple. Also considered are single transient oscillographic systems employing either traveling-wave-deflection structure cathode ray tubes or multiple sampling methods. In order to clearly present the rise-time limitations caused by TEM transmission lines, attention is given to the distortion incurred by pulses upon passing through such lines. Some suggestions and predictions relating to future work are presented.**

I. Introduction

THE PURPOSE of this paper is to describe the present-day capability in measuring baseband (i.e., video) pulse transition times of less than one nanosecond and to provide some insight into the probable course of future development.

At present the measurement techniques can be classified into three categories: 1) real-time oscillography, 2) sampling oscillography, and 3) pulse comparison. The first two techniques provide in principle a direct view of the pulse as a function of time while the third technique may or may not provide such a view.

Central to the development of fractional nanosecond pulse techniques has been the implicit establishment of a basic oscillographic instrumentation system which consists of three parts: 1) an abrupt-transition pulse generator, 2) interconnecting transmission circuits, and 3) the oscillographic indicator. This basic system can be calibrated by fundamental means so that the display *system* pulse waveform can be known with very little uncertainty. Note that the displayed waveform is the response due to the cascade of the three networks comprising the system and is not the response of the oscilloscope by itself.

The pulses under discussion here contain frequency components within the range from dc to above 20 GHz. Because of the large bandwidths required to transmit or delay such pulses from point to point, the pulses must be transmitted or delayed within low-loss uniform TEM-mode transmission lines. Furthermore, because transmission-line losses and nonuniformities are always present, the pulse shapes will be changed by transmission through such transmission lines.

The order of the topics to be presented is as follows: real-time oscillography, sampling oscillography, pulse com-

Manuscript received February 15, 1967; revised March 22, 1967.

The author is with the National Bureau of Standards and the University of Colorado, Boulder, Colo.

385-582 O - 70 - 9

parison techniques, a basic instrumentation system, transmission-line pulse distortion, and concluding remarks.

Finally, numerous references are cited, some of which include private oral or written communications between the author and other workers in this field. In order to provide recognition to these individuals, their communications are acknowledged in the references.

II. Real-Time Oscillography

The main utility of the real-time oscilloscope rests in its ability to display a nonrecurrent pulse, i.e., operate in the single-shot mode. To do so, the oscilloscope cathode ray tube (CRT) must possess sufficient beam energy and suitable phosphor characteristics so that the light from its trace is able to expose a photographic film producing a permanent record of the single-shot pulse waveform. It is also possible to employ a vacuum-demountable CRT in which the electron beam writes directly upon a photographic film within the vacuum space. In order to produce the high writing rates required to record fractional nanosecond pulses, design compromises are necessary which yield CRT's that are relatively deflection-insensitive when compared to conventional tubes.

To increase the deflection sensitivity in the signal (vertical) channel, a traveling-wave (TW), or in other words, slow-wave deflection structure is used. Simply put, the slow-wave deflection structure allows the vertical deflection signal to travel with the same velocity as the electron beam. Consequently, over the length of the deflection structure each particular value of the vertical signal travels along with a portion of the electron beam and acts on the beam for a longer period of time than would be the case with an ordinary parallel-plate deflection structure. Because of the increased action time the corresponding deflection is greater; hence, the deflection sensitivity is increased.

Since there are no amplifiers available possessing the bandwidth, impedance characteristics, and dynamic range suitable for amplifying fractional nanosecond pulses, the vertical deflection signals are fed directly into the traveling-wave deflection structure via transmission lines having the same characteristic impedance as the traveling-wave structure. The relatively low sensitivity of TW oscilloscopes has been capitalized upon for measurements on large-amplitude signals. In fact, TWCRT's have been designed for very low-deflection sensitivities to eliminate the need for wideband and high-dissipation attenuators in observing such signals [8].

Table I contains performance data which are representative of what is to be expected from TWCRT's; data on some other tubes are given elsewhere [8]–[13]. Of particular interest are entries under f_{co} and τ, the deflection response cutoff frequency (6 dB down) and the 10- to- 90-percent *system* rise time, respectively. For a given tube, no direct correlation can be made between f_{co} and τ because f_{co} relates solely to the performance of the TWCRT while τ represents the collective performance of the pulse generator, interconnecting transmission lines, and the TWCRT. However, upon considering the data for all of the tubes listed, the expected trend of τ being inversely proportional to f_{co} does appear.

A direct correlation between bandwidth and rise time could only be made if all the tubes could be tested in the same system; this of course is not possible because of the different impedance levels.

The Model-2 tube is a vacuum-demountable three-beam device which writes directly on photographic film. Six sets of deflection plates provide three independent oscillographic channels. The greater-than-10-GHz value for f_{co} is a theoretical value computed solely from transit-time distortion considerations; however, sinusoidal waveforms up to 10 GHz have been recorded, but an overall bandwidth study has not been made [7]. No information is available on fractional nanosecond systems using the Model-2.

In regard to some of the physical details of the traveling-wave deflection structures, the helical structures have a "*D*"-shaped cross section with the flat side facing and parallel to the electron-beam axis; balanced helical structures consist of two helices each unbalanced to ground but balanced to each other. The zig-zag structure is a planar structure that meanders regularly from side to side. The disk-coaxial structure is a disk-loaded coaxial line, i.e., disks are placed regularly along the center conductor.

Of major importance in realizing a constant deflection factor over the range of dc to several GHz is the collective uniformity of the transmission-line system consisting of the vertical deflection plate connectors, the transitions between the connectors and the traveling-wave deflector, the deflector, and the deflector-terminating impedance. The measurement technique employed in obtaining the f_{co} values in Table I (excepting the Model-2) measured the overall effectiveness of the vertical deflection transmission-line system.

Balanced deflection operation requires the use of a transmission-line phase inverter [14] to produce the required out-of-phase deflection signals. This will double the deflection factor, but the nonuniformities and high-frequency losses introduced by the inversion processes may impose a serious reduction in system bandwidth. The balanced-deflection TWCRT also is useful as a pulse comparator and a pulse coincidence indicator. The two pulses (of the same polarity) to be compared are introduced to separate vertical deflection plates. If the pulses are coincident in time and are of exactly the same shape, the trace deflection will be zero.

Up to this point nothing has been said about the horizontal channel and the generation of the time base. Inspection of Table I shows that the tubes listed have a horizontal deflection sensitivity within the approximate range of 0.3 to 6.5 volts per spot width or 94 to 1625 volts per inch. For a 1-ns/cm time base, horizontal sweep voltages ranging from 37 to 640 V/ns would be required. Consequently, it is apparent that the horizontal scan length may be limited by the dynamic range of the time-base generator. Typical time-base circuitry is given in the literature [2], [4], [8], [13], [15].

Representative of present-day development efforts is a TWCRT employing a serrated slow-wave structure, and a fiber-optic coupler; the development objectives are: $f_{co} > 2$

TABLE I

TRAVELING-WAVE CATHODE RAY TUBES

Tube No.	Vertical TW Deflector					Horizontal Plates,		Beam	
	Type	S_v	Z	f_{co}	τ	S_H	D	W	P
KR1/3341A	*B*-Helix	0.054	120*	1.6 [1]	300 [2]	0.3	0.002	2×10^{11}	24
KR3	*B*-Helix	0.054	100	1.0 [1]	400 [3]	0.3	0.002	2×10^{11}	21.4
KR5	*U*-Helix	0.93	50	>2.0 [3]	250 [3]	1.38	0.006	10^{12}	25
K1524P	*U*-Helix	1.04	50	2.1 [1]	370 [4]	3.64	0.013		25
T519-1	*U*-Zigzag	0.084	125	1.7 [5]	300 [6]	0.376	0.004		24
T519-2	*U*-Coax.**	0.74	125	—	125 [6]	0.376	0.004		24
Model-2	*U*-Disk-Coax.	0.2	50	>10† [7]	—	6.5	0.004	3×10^{11}	50

S_H—Horizontal sensibility, volts per spot width
S_V—Vertical sensibility, volts per spot width; in balanced structure, value given for unbalanced operation of one deflector
Z—Characteristic impedance, ohms
B—Balanced to ground
U—Unbalanced to ground
f_{co}—Frequency in GHz at which the deflection response is 6 dB down from dc value; generator impedance and traveling wave structure termination equal to Z
τ—10-to-90-percent rise time in picoseconds (10^{-12} second) of experimental system
D—Spot size, inches
W—Writing rate, spot widths per second
P—Total acceleration potential, kilovolts
[ ]—See References
*—First models had $Z=120$ ohms, in later models $Z=100$ ohms
†—Theoretical value based on transit-time considerations
**—Nonloaded coaxial line; no slow-wave properties.

gigahertz, $\tau=200$ picoseconds, $Z=50$ ohms, $S_v=0.375$ volts per spot width, $D=0.0015$ inch, $W=4\times10^{12}$ spot widths per second [16].

Future TWCRT's will employ fiber-optic systems. The fiber-optic system in conjunction with presently available photographic film in effect provides a faster film. The effective increased film speed can be traded off in design to enhance the deflection sensitivity or the writing rate [3]. An excellent discussion on the application of fiber optics to CRT design is given by Katzmann [17].

Also, there is a need for additional investigation into the analysis of wideband CRT's. In particular, attention should be given to the detailed theory of TEM-mode slow-wave structures, electron ballistics within the fields of wideband deflection structures, and the realization of uniform deflector transmission-line systems.

III. Sampling Oscillography

In contrast to the low sensitivity and single-shot capability of the real-time oscilloscope, the sampling oscilloscope possesses a high sensitivity and ordinarily requires a recurrent waveform. The shape of the recurrent waveform is determined by automatic point-by-point measurement, each individual value being measured at a different point on successive waveform recurrences. The successive measurements are displayed on a relatively narrowband oscilloscope. If smoothing networks are not used, the displayed waveform may appear as a discrete number of dots spaced uniformly in the horizontal direction (time axis) with a vertical displacement of each dot respectively proportional to the amplitude of the waveform. Two different display traces are possible in sampling oscillography; they result from the type of horizontal scanning used. Usually, discrete horizontal scanning produces a dotted trace while continuous scanning yields a smooth trace. However, low-pass filter networks in the sampled-data channel may be used to transform a dotted trace into a smooth one, and intensity modulation to transform a smooth trace to a dotted one.

In general, an ideal sampling oscilloscope may be thought of as a waveform converter that converts a rapidly varying waveform into a slowly varying waveform of exactly the same shape, i.e., uniformly stretched in time.

The samples need not be taken in a uniformly spaced time sequence; in fact, a random sequence has particular advantages. Furthermore, in principle a recurrent waveform is not required if multiple samples may be taken on a single occurrence. Consequently, sampling oscilloscopes may be classified into three sampling categories: 1) sequential, 2) random, and 3) multiple.

Before discussing the details of the three sampling methods a brief illustration of a sampling process [18] will be given in order to provide a basis for later discussion. Consider an unknown waveform source, a sampling-pulse generator, a capacitor, and a diode connected in a series loop as shown in Fig. 1. For simplicity assume that a) the respective generators have zero internal impedance and the same peak value, and b) the diode is ideal.

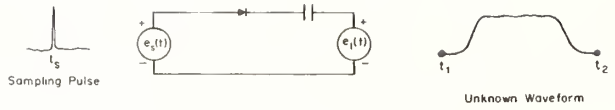


Fig. 1. Elementary sampling system.

First, consider that only the sampling generator is triggered. The sampling pulse appears at the diode anode, passes through it and charges the capacitor to the sampling-pulse peak value which is retained indefinitely; hence, the sampling-pulse amplitude may be determined at a later convenient time by measuring the capacitor potential.

Second, consider that the sampling pulse and the unknown appear simultaneously, i.e., $t_1 < t_s < t_2$. The unknown waveform applied by itself cannot produce a capacitor potential because the diode will not conduct for positive cathode-to-anode potential. However, when the unknown waveform is applied simultaneously with the sampling pulse, the effective charging amplitude of the sampling pulse is reduced by the instantaneous value of the unknown waveform. If the waveform is at its peak value, then the capacitor potential will remain at zero.

By successively varying the time position of the sampling pulse in relation to the unknown waveform and measuring the resultant capacitor potential, the unknown waveform may be determined. In this hypothetical system the capacitor must be discharged after each measurement. In physically realizable systems, the capacitor discharges through the reverse diode resistance and the loop circuit resistance.

From this illustration it is apparent that the sampling technique defers to a later time the actual measurement of the instantaneous values. In practice, this means that the instantaneous values are measured during the dead time between successive recurrences of the unknown waveform. Furthermore, many recurrences are necessary to determine the entire waveform; consequently, the time scale of the resultant waveform constructed from the samples directly depends upon the number of samples (recurrences).

In a constant-current sampling (charging) system, if the sampling pulse has a symmetrical triangular shape, the step response of the oscillographic system will have a 0-to-100-percent rise time equal to time duration of the sampling-pulse base, and the deflection frequency response will be of the form $(\sin x)/x$, x being proportional to frequency [19]. Some more detailed analyses and other systems are presented in other articles [20]–[24].

The discussion now returns to the three types of sampling systems. The first kind, the sequential system, takes the samples in order of increasing time referred to the unknown waveform or signal. The original sampling oscillographic system [25] is of the first kind and was designed to display periodic (as contrasted with aperiodic) signals. To scan the signal it employed phase modulation of a sine wave that drove the sampling-pulse generator. The scanning, being continuous, produced a smooth trace. Some later systems [26]–[28] also employed similar phase-modulation scanning techniques; however, the resultant displays were not all continuous due to intensity modulation, but all of these systems were suitable for viewing periodic waveforms or waveforms initiated by a command pulse generated within the oscillograph.

With the development of the first "signal-triggered" sampling oscilloscope [29], [30], sampling oscillography was extended to recurrent aperiodic waveforms. Subsequent development of the signal-triggered system has been relatively rapid [31]–[47] and has led to quite sophisticated instruments. Typical specifications for some currently available sampling oscilloscopes are given in Tables II and III.

The time delay T given in Table III has a significance in relation to the usable oscilloscope bandwidth when it is necessary to trigger the oscilloscope with a trigger derived directly from the signal. Figure 2 shows the method required to view aperiodic signals from nontriggerable sources. Notice the delay line placed in the vertical channel to delay the signal in order to allow for the time required to generate the sampling pulse and deliver it to the sampler. For real-time oscillography, operation in the signal-triggered mode also required a time delay in the signal channel; the time delay is based on the time required to start the sweep and intensify the electron beam. Introduction of the delay line in the vertical channel of the real-time or sampling system reduces the overall system bandwidth. The effects of such delay lines will be discussed in Section VI.

The second kind of sampling system is the random-sampling type which presents the signal samples in a random order but places each sample in its correct time position. Figure 3 illustrates the difference between the sequential presentation and the random presentation; the numbers represent the order of the successive samples. In the sequential system the pulse information is collected at uniform intervals from the beginning to the end of the pulse; in the random system the information is collected in a nonuniform manner and in no prescribed sequence. Two random-sampling oscillographs are reported in the literature [26], [50]. The earlier oscillograph employed the random-sampling method only as an antijitter circuit, while the later one fully utilized the method as an oscillographic technique which realized a bandwidth of greater than 250 MHz.

The term "random sampling" has also been used to describe the incoherent sampling employed in sampling voltmeters [51]. Sampling voltmeters [52] and real-time sampling oscilloscopes [42] employ incoherent sampling. Incoherent sampling does not provide direct (time-based) information regarding the shape of the periodic signal; it does provide a measurement of the peak-to-peak value and can follow low-frequency modulation of the periodic signal. Consequently, a real-time sampling oscilloscope can only provide a display of a modulation envelope. Incoherent sampling may be used to provide histographic (statistical) information of periodic waveforms.

It is to be emphasized that the term "random sampling" as used in this paper relates to equivalent-time sampling systems and denotes a nonsynchronous sampling technique through which the signal shape is preserved.

TABLE II
TYPICAL DUAL-CHANNEL SAMPLERS*

DFR	Sampling Head								
	Type	Z	VSWR	f_{co}	τ	Δ	N	DR	Kick-out
1–200	$R/B/2$	50 E	3 to 1	12.4	35 [48]	±5	7	±1	10/5
2–200	$R/B/4$	50 E	1.5 to 1	>3.5	100 [49]	±4	4	±1	—
2–200	$V/B/2$	50 I	2 to 1	>4	90 [6]	±5	4	±1	—

*—Data per channel
DFR—Deflection factor range, mV/cm
V—Sampler in vertical unit
R—Remote sampler
$B/2$—Bridge sampler per two diodes
Z—Sampler input impedance, ohms
E—Sampler channel externally terminated to allow feed-through operation
I—Sampler channel internally terminated
VSWR—Sampler voltage standing wave ratio, less than indicated value up to f_{co}
f_{co}—Frequency in GHz at which sampler output response is 3 dB down from dc value; generator impedance and sampler termination equal to Z
τ—10-to-90-percent rise time in picoseconds (10^{-12} second) of experimental system
Δ—Percent overshoot
N—Noise unsmoothed, millivolts peak to peak, observed on screen excluding 10 percent of random dots
DR—Dynamic range, volts
Kick-out—Pulse produced by sampler and sent back toward signal source, approximate amplitude in millivolts per duration in nanoseconds (10^{-9} second).

TABLE III
SAMPLER-TIME BASES

Time Base Range	External Trigger				
	Z	L	W	J	T
10 ps/500 μs	50	50	2	20	60 [48]
3 ps/100 μs*	50	10	2	30	80 [49]
10 ps/100 μs†	50	50	2	30	45 [6]

ps—Picoseconds (10^{-12} second) per centimeter
μs—Microseconds (10^{-6} second) per centimeter
*—Real-time sampling time base 1 ms/cm to 10 ms/cm in addition to range listed
†—Real-time sampling time base 0.2 ms/cm to 5 s/cm in addition to range listed
Z—Input impedance, ohms
L—Trigger pulse level, millivolts
W—Trigger pulse width, nanoseconds
J—Time-base jitter for L and W conditions. Less than tabulated value, in picoseconds (10^{-12} second)
T—Time delay in nanoseconds (10^{-9} second) before sample is taken in response to time-base trigger.

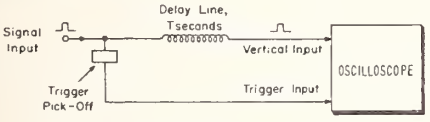


Fig. 2. Signal-triggered operation.

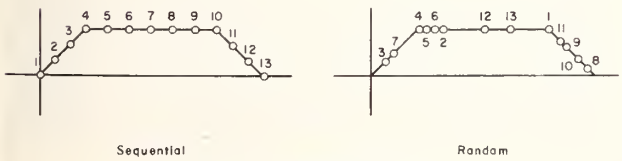


Fig. 3. Sampling displays.

Briefly, the random-sampling system operates as follows. When the signal enters the system it starts a linear ramp and also enters the sampler. When a sampling pulse occurs (at random) it samples the signal in the sampler and after a fixed delay stops the linear ramp. The value of the signal sample and the height reached by the linear ramp are both stored. The ramp height is related to the time position of the sample. Since the ramp started when the signal arrived and was subsequently stopped by the delayed sampling pulse, the time position of the sample is known in respect to the beginning of the signal [50].

There are two advantages to the random sampling system: 1) a signal channel delay line is not required in the signal-triggered mode and 2) synchronization is not required between the signal and the sampling pulse. On the other hand, there are two disadvantages: 1) time-base memory circuitry is required and 2) the signal information-collection rate may be very slow, depending upon the number of simultaneous occurrences of the signal and sampling pulse.

The third kind of system, the multiple-sampling system, has been physically realized and is undergoing further development. To obtain multiple samples from a single pulse several methods are in principle possible. Two such methods could be accomplished by 1) passing the pulse through a cascade of samplers, and 2) recirculating the pulse through a single sampler. To date, systems based upon the first method claim to have attained a bandwidth of 500 MHz for a cascade of 50 samplers [96]. Other systems are described elsewhere [53], [97]–[99].

An active effort is being made to construct a recirculating system by employing superconductive lines [54]. Central to this effort is the development of a superconductive switch which would admit the pulse to the recirculating system [55]. Experiments have been performed in which pulses

were recirculated in superconductive lines. One method [56] recirculated nanosecond pulses 100 times over a 300-foot superconductive line, while another method [57] observed a pulse after it had traveled a total distance of 140 000 feet along a superconductive line of 700 feet in actual length. Also, an active serial storage system has been realized which employs a superconductive line and a tunnel diode [100], [101].

Basic to present-day sampling technology is the physical realization of a sampler whose configuration provides 1) a uniform transmission line for the signal, and 2) sampling at a point. The most advanced sampling structure produced to date which approaches these objectives is decribed by Grove [47]. Future developments along such lines will probably incorporate integrated circuitry in the sampling gate in an effort to reduce the size and loading effects of the sampling circuit. Consideration of intrinsic frequency limitations for semiconductor (hot carrier) diodes [58] will assist in developing accurate baseband models of sampler circuits.

Because of the ease in providing an analog output from a sampling oscilloscope, it is possible to analyze the oscilloscope display by processing the analog output in a computing system. One method for converting a sampling oscilloscope analog output to digital form employed as a buffer unit a coincident core memory with a capacity of 4096 16-bit words [59]. Such computer attachments will provide expanded possibilities in reducing measurement data.

IV. Pulse Comparison Techniques

Pulse comparison techniques are basically time-coincidence measurements. A given recurrent pulse is applied simultaneously to the two channels of a time-coincidence detector. If the channel time delays are identical, the pulses in each channel will arrive in time coincidence, and the coincidence detector will give the maximum coincidence reading. If the time delay in one channel can be continuously varied, then the time position of the pulse coming out of the variable-delay channel may be shifted in relation to the pulse output of the constant-delay channel; the resultant detector reading will depend upon the degree of coincidence between the pulses.

The elementary sampling system circuit (Fig. 1) may be used as a pulse coincidence circuit. Consider the unknown pulse to be applied to each side of the diode capacitor combination. When the pulses are in time coincidence, no current will flow, and the capacitor C will not accumulate charge. If either pulse (but not both) is shifted by $t_2 - t_1$ seconds, the pulse duration, from the coincident time position, the capacitor will charge to the maximum possible value. For time shifts less than $t_2 - t_1$ the capacitor will accumulate less charge.

Coaxial comparison systems employing the simple diode detector (Fig. 1) have been used to determine the approximate time dimensions of fractional nanosecond pulses [60]–[65]. A typical system is shown in Fig. 4. It is apparent from the basic circuit arrangement that the method most

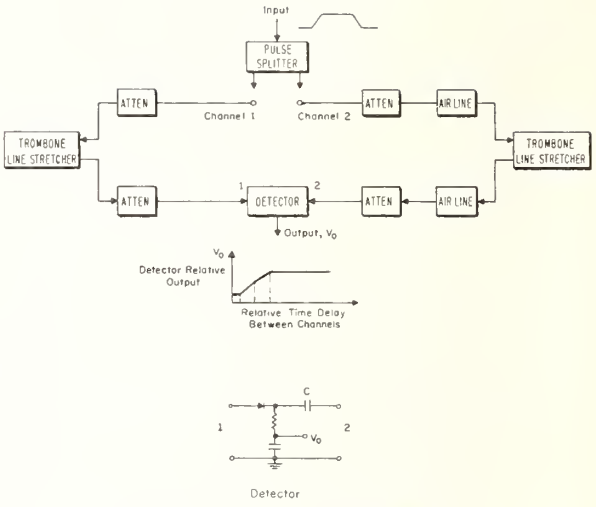


Fig. 4. Pulse comparison system.

certainly could be used to determine pulse width; however, it is not evident how the method could provide information about rise time, flat top, and fall time. Analysis [64] demonstrates for a trapezoidal pulse that the nonlinear voltage-current characteristics of the diode in the forward direction causes changes in the curvature of the coincidence detector output curve. The changes occur at relative times which are approximately equal to the 0-to-100-percent rise time, the flat top, and the fall time. It should be emphasized that this particular technique cannot provide the detailed measurements that are realized with a sampling oscilloscope. However, it can be used to evaluate sampling-pulse width and in that way assist in the development of sampling systems. Pulse comparison systems provided the first measurements of 100-MHz klystron pulses having 10-to-90-percent rise and fall times of 100 and 300 picoseconds, respectively [61], [64].

The diode detector, pulse comparison system has also been called [63] a "self-sampling" system because the signal samples itself. It is possible to build a rudimentary sampling system from a pulse comparison system using only passive components to create a sampling pulse in one channel from the input signal [66]. If the comparison detector is a true square-law detector, then a pulse comparison study for a given pulse would yield an output curve that would be the autocorrelation function of the pulse. If the two pulse inputs were not the same, then the output curve would be the correlation function between the two pulses. It is evident that comparison studies could be made between any two pulse shapes.

The pulse comparison technique using the diode coincidence detector or a square-law detector does not provide the exact shape of the pulse. In order to obtain more information about the pulse shape, the coincidence detector could be replaced by an amplitude-sensitive detector such as a tunnel diode [64], [65]. A pulse comparison system

using a tunnel-diode detector has been employed by Frye [6]. Such amplitude-discriminator techniques used in the dissimilar input form of the comparison technique should yield useful results. In regard to applying amplitude discriminators to the measurement of fractional nanosecond pulses, measurements [67] have been made with amplitude discriminators on pulses of the order of 20 to 30 picoseconds.

V. A Basic Oscillographic Instrumentation System

The development of oscillographic systems for fractional nanosecond pulse instrumentation has been a bootstrap process. Improvements in oscilloscopes led to improvements in pulse generators, transmission-line components, and other components which in turn engendered further oscilloscope development. Central to such development work was the use of the instrumentation system shown in Fig. 5. This system contains amplitude- and time-calibration circuitry which provides for calibrating the *displayed* system waveform.

Time-base calibration is accomplished by changing the length of the trigger channel transmission line. Upon shortening the trigger line the display moves to a later time position. If v is the velocity of propagation in the removed length x of trigger line, then the display will shift by x/v seconds. Small changes are accomplished with a line stretcher which provides a time shift of 10 picoseconds per 3 millimeters. The vertical deflection scale is calibrated by applying a known variable dc potential to the vertical input.

The calibrated *system waveform* can be employed as a pulse source for testing four-terminal networks inserted between the insertion connectors; the resultant display will be the convolution of the system waveform and the impulse response of the inserted four-terminal network. Consequently, the system may be employed to evaluate a theoretical model of a four-terminal network inserted into the vertical channel; this method has been used to investigate the validity of a transmission-line model [68].

The system waveform rise time is an upper bound that each component rise time at most may approach. The isolation of the individual component transient responses requires accurate models for all but one component in conjunction with the displayed response.

In many applications a calibrated system waveform is all that is really needed. Of course, comparisons between oscilloscopes ideally require an absolute reference pulse; but for comparisons between state-of-the-art high-speed oscilloscopes or other pulse-measurement instruments, most probably such a pulse will not exist. Consequently, the indirect methods of the calibrated system waveform will continue to be employed and refined.

Frequently it is necessary to demonstrate that the pulse response characteristics of a given oscillographic system have or have not changed with time. This can be done effectively if a regular system-calibration schedule is followed and if the system components are not altered. Sampling oscilloscopes are particularly amenable to such calibration as the display analog read-out provides a convenient means for recording the system response.

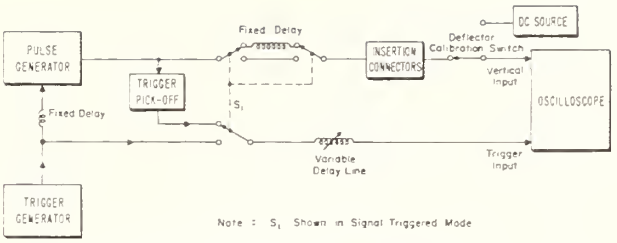


Fig. 5. Basic instrumentation system.

VI. Transmission-Line Pulse Distortion

Because large bandwidths are required to transmit or delay fractional nanosecond waveforms, low-loss uniform TEM-mode transmission lines are employed. Even though the losses are small and the nonuniformity very slight, pulses are altered upon passing through such lines [68]–[79], [95].

Considering for the moment strictly uniform transmission lines, the pulse distortion is due to skin effect losses in the conductors and polarization processes in the dielectric. The conductor loss per unit length is inversely proportional to the conductor surface area, while for a given characteristic impedance the dielectric loss per unit length is independent of the line dimensions. Accordingly, for coaxial lines of the same materials and characteristic impedance, the smaller the line diameter, the greater is the loss per unit length.

As the line diameter is increased, the metal losses decrease, and the size-independent dielectric loss appears as a greater portion of the total loss. To further reduce the total loss while maintaining the same outer diameter, the volume of the solid dielectric material within the dielectric space is reduced by removing or hollowing out the solid dielectric and increasing the center-conductor diameter to maintain the given characteristic impedance. The resultant semisolid dielectric contributes less dielectric loss per unit length while also the conductor loss per unit length is decreased due to the increased inner-conductor diameter. Consequently, the total line loss per unit length has been significantly reduced, but not without penalty.

The price paid for reducing the attenuation is twofold: the hollowing of the dielectric introduces additional nonuniformities, and the generally larger dimensions decrease the cutoff frequencies of the higher-order transmission modes.

The effects of loss and nonuniformity on the pulse response of a transmission line can be shown by the step response. In coaxial lines for which simple skin effect is the dominant distorting effect, the step response, after the initial delay of l/v seconds, is of the form $\operatorname{cerf}\text{-}(kl/\sqrt{t})$ [68] where cerf x is the complementary error function of argument x, and l, v, t, and k are the length, propagation velocity, time, and a constant, respectively (see Fig. 6). For lines in which plated conductors are used or in which small dielec-

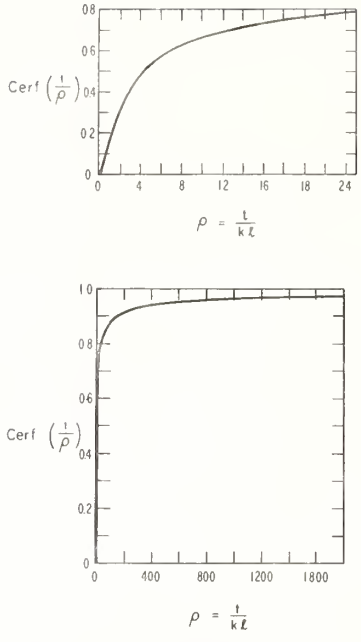


Fig. 6. Normalized step response considering skin-effect loss.

tric losses (compared to skin-effect losses) exist, the step response will depart from the error function response but will have a similar shape [69]. Inspection of Fig. 6 shows that the step response as a whole rises rapidly to the 50-percent level, then takes about 50 times as long to reach 90 percent, about 500 times as long to reach 95 percent, and then tails up even more slowly as the 100-percent level is approached. Typical 0-to-50-percent times of 100-foot lengths for RG-8/U and RG-58/U are of the order of 0.32 and 1.8 nanoseconds, respectively.

Semisolid dielectric cables have nonuniformities which introduce large amounts of phase shift which cannot be associated to the joule heating losses; consequently, the phase shift is nonminimum phase shift. Because of this nonminimum phase property, the step response cannot be predicted solely on the basis of attenuation data as both attenuation and phase information are required. Characteristic of these cables are step responses which ring at the leading edge while afterwards tail up slowly, similar to the response of uniform cables.

In order to improve the transmission properties of cables for pulse instrumentation two approaches have been pursued: 1) equalization of the transmission characteristics [48], [80]–[85] and 2) reduction of line losses by using superconductive metals and a cryogenic environment [57], [86]–[94].

By means of the first technique useful results can be achieved using simple reflective equalizers, e.g., constant 7-dB loss out to 7 GHz for 100 feet of $\frac{1}{2}$-inch styroflex cable [81]. For long lengths and matched impedances, efficient equalization is achieved by amplitude and phase equalization. Excellent results have been obtained in which 1000 feet of $\frac{7}{8}$-inch styroflex cable was nonreflectively equalized for transmission of 1-ns 10-to-90-percent rise-time pulses, 7-dB loss out to 100 MHz, 10 dB down at 350 MHz, time delay flat to 1 GHz [84]. Most modern samplers with built-in delay lines and trigger pick-off systems use some form of nonreflective delay-line equalization.

Superconductive lines, because of their very small metal losses, can be miniaturized. Also, operation at cryogenic temperatures reduces the dielectric losses. Consequently, such lines do provide large delay times with little pulse distortion. Rise times of 500 picoseconds have been preserved upon transmission through a 1360-foot length of 0.051-inch dielectric OD 50-ohm cable (Pb and Nb outer and inner conductors, respectively; polytetrafluoroethylene dielectric; 4.2°K operating temperature [89]). For a similar cable of 80 feet in length (130 nanosecond delay), a 45-picosecond rise time was observed for the line. There was an initial loss in step height of approximately 7 percent at 50 picoseconds which decreased to 1 percent in about 4 nanoseconds [94].

VII. Conclusion

In the final analysis, the utility of this sort of paper rests in the stimuli to others that it may provide. Here has been presented the author's view of the development and the present state-of-the-art of the measurement of baseband pulse rise times of less than 10^{-9} second.

As for the future, these suggestions may serve as guides for some or as points of departure for those of opposing views. In any event, it is the challenge evoked by suggestions which spur others on to prove, disprove, or suggest anew. . . .

Relevant to all of the areas discussed is the necessity for the development of improved models to serve as the basis for analysis and accurate design. The resolution of a system rise time into the contribution of the component parts of the system can only be done with adequate models. This is not a contrived generalization, but an urgent necessity. The formulation and verification of better models will be time-consuming and arduous tasks, but are none the less important.

At present it appears that for the required delay-line length, the rise time of the best (superconductive) delay line is of the same order as that of the best sampler. Improvements in superconductive lines or a reduction in time-delay requirements could increase the delay-line bandwidth. Quite probably, for recurrent waveforms recourse to random sampling will avoid the delay-line problem by simply eliminating the line.

For single-shot measurements on rise times of the order of 0.1 to 1 nanosecond, multiple sampling by pulse recirculation appears to be a practical method. Fiber optics and possibly image-storage systems will improve the performance and extend the usefulness of the TWCRT.

Pulse comparison systems will continue to be discarded and resurrected as the occasion demands. TEM-mode transmission lines will undergo more detailed analysis and efficient equalization techniques will be extended down into the fractional nanosecond domain.

References

[1] C. W. Womack, O. Gaddy, and N. S. Nahman, "Project JAYHAWK progress rept. for period 12 April 1959 to 11 July 1959," University of Kansas Research Foundation, Lawrence, under Contract DA 49-170-sc-2618.

[2] E. J. Martin, "A millimicrosecond cathode-ray synchroscope," in *Bull. Engrg. and Architecture*, no. 43. Lawrence, Kan.: University of Kansas Publications, 1959, pp. 19–31.

[3] S. Goldberg, Edgerton, Germeshausen & Grier, Inc., Boston, Mass., private communication.

[4] C. N. Winningstad, "A fractional millimicrosecond oscilloscope system utilizing commercially available components," *Rev. Sci. Instr.*, July 1958.

[5] W. Sheffer and N. S. Nahman, "Project JAYHAWK progress rept. for period 12 October 1962 to 11 January 1963," Electronics Research Lab., University of Kansas, Lawrence, under Contract DA 18-119-sc-2420.

[6] G. J. Frye, Tektronix, Inc., Beaverton, Ore. private communication.

[7] G. M. Lee, Central Research Labs. Inc., Red Wing, Minn., private communication.

[8] I. A. D. Lewis and F. H. Wells, *Millimicrosecond Pulse Techniques*, 2nd ed. New York: Pergamon, 1959, ch. 6.

[9] K. J. Germeshausen, S. Goldberg, and D. F. McDonald, "A high-sensibility cathode-ray tube for millimicrosecond transients," *IRE Trans. on Electron Devices*, vol. ED-4, pp. 152–158, April 1957.

[10] D. J. Brangaccio, A. F. Dietrich, and J. W. Sullivan, *Bell Sys. Tech. J.*, vol.37, pp. 447–460, March 1958.

[11] L. B. Wollaver, "CRT type no. KR-1," Edgerton, Germeshausen & Grier, Inc., Boston, Mass., Tech. Memo. B-69, February 1958.

[12] J. L. Turner, "CRT type no. KR-3B," Edgerton, Germeshausen & Grier, Inc., Boston, Mass., Tech. Memo. B-84, March 1958.

[13] Y. N. Prozorovski, "The SO-3 high speed oscillograph" (in Russian), *Pribory i Tekhnika Eksperimenta*, no. 5157, pp. 73–76.

[14] Lewis and Wells, *op. cit.*, section 3.5.1.

[15] G. J. Frye, J. D. Bruce, and N. S. Nahman, "A nanosecond time base for traveling-wave deflection cathode-ray tubes," *IRE Trans. on Instrumentation*, vol. I-10, pp. 85–89, September 1961.

[16] G. Longerbeam, Livermore Radiation Lab., Livermore, Calif., private communication.

[17] F. L. Katzmann, "Improving ultra-fast transient recording using fibre-optics CRT's," *Electronic Instr. Digest*, pp. 41–47, October 1966.

[18] N. S. Nahman, "A survey of millimicrosecond pulse instrumentation," in *Bull. Engrg. and Architecture*, no. 43. Lawrence, Kan.: University of Kansas Publications, 1959, pp. 5–9.

[19] Lewis and Wells, *op. cit.*, section 6.5.3.

[20] J. M. L. Jannsen, "A cathode-ray oscillograph for periodic phenomena of high frequencies," *Philips Res. Rept.*, pp. 205–240, June 1950.

[21] V. A. Vol, "To the theory of stroboscopic oscillography" (in Russian), *Radiotekhnika*, vol. 13, no. 8, pp. 63–70, 1958.

[22] ——, "The reproduction by a stroboscopic oscilloscope of periodic signals of arbitrary form" (in Russian), *Radiotekhnika*, vol. 14, no. 3, pp. 69–75, 1959.

[23] R. Carlson et al., "Sampling oscillography," *IRE WESCON Rec.*, pt. 8, pp. 45–51, 1959.

[24] D. Howard, A. Best, and J. Umphrey, "The wideband sampling gate—An analysis, characterization and application discussion," *WESCON Tech. Papers*, pt. 6, session 23, paper 1, 1966.

[25] J. M. L. Jannsen, "An experimental stroboscopic oscilloscope for frequencies up to about 50 Mc, Parts I and II," *Philips Tech. Rev.*, vol. 12, no. 2, pp. 52–59 and no. 3, pp. 73–82. August 1950.

[26] J. G. McQueen, "The monitoring of high-speed waveforms," *Electronic Engrg.*, vol. 24, pp. 436–441, October 1952.

[27] A. S. Farber, "Sampling oscilloscope for millimicrosecond pulses at a 30 Mc repetition rate," *Rev. Sci. Instr.*, vol. 31, pp. 15–17, January 1960.

[28] W. M. Goodall and A. F. Dietrich, "Fractional millimicrosecond electrical stroboscope," *Proc. IRE*, vol. 48, pp. 1591–1594, September 1960.

[29] R. Sugarman, "A fast oscilloscope for scintillation pulses," Brookhaven Nat'l Lab., Upton, N. Y., Rept. BNL 2646.

[30] ——, "Sampling oscilloscope for statistically varying pulses," *Rev. Sci. Instr.*, vol. 28, pp. 933–938, November 1957.

[31] G. B. B. Chaplin, "A method of designing transistor avalanche circuits with application to a sensitive transistor oscilloscope," *Proc. Transistor-Solid State Conf.*, session 3, February 1958.

[32] G. B. B. Chaplin, A. R. Owens, and A. J. Cole, "A sensitive transistor oscillograph with dc to 300 Mcs response," *Proc. IEE (London)*, vol. 106B, May 1959.

[33] P. R. Orman, "The pulse sampling oscilloscope," *Nuclear Electronics Conf.*, vol. 1, Paris, France, 1958.

[34] W. E. Bushar, "Sample method displays millimicrosecond pulses," *Electronics*, vol. 32, pp. 70–71, July 1959.

[35] R. Carlson, "A versatile new dc–500 Mcs oscilloscope with high sensitivity and dual channel display," *Hewlett-Packard J.*, vol. 11, pp. 1–8, January–March, 1960.

[36] J. J. Amodei, "Converting oscilloscopes for fast rise sampling," *Electronics*, vol. 33, pp. 96–99, June 24, 1960.

[37] C. N. Winningstad, "Nanosecond pulse measurements," *IRE WESCON Rec.*, January 1961.

[38] ——, "Storage to picoseconds—A survey of the art," *Electronic Ind.*, August 1963.

[39] G. Long, "Pulse reflections pin down discontinuities," *Electronic Design*, May 10, 1963.

[40] W. M. Grove, "A new dc–4000 Mc sampling scope plug-in with signal feed-through capability," *Hewlett-Packard J.*, vol. 15, April 1964.

[41] H. A. Zimmerman, "Nanosecond measurements with a sampling oscilloscope," *Electrotech.*, pp. 69–72, January 1965.

[42] G. J. Frye, "Oscilloscope sampling techniques," *Electronic Ind.*, pp. 108–112, June 1965.

[43] M. R. Kaufman, "Current measurements at nanosecond speeds," *Electronic Design News*, October 1965.

[44] H. A. Zimmerman, "Pseudo-Schmitt eliminates uncertainties in trigger logic," *Electronics Design*, pp. 30–35, November 22, 1965.

[45] W. Grove, "Sampling for oscilloscopes and other RF systems: dc through *X*-band," presented at the 1966 Internat'l Symp. on Microwave Theory and Techniques, Palo Alto, Calif.

[46] A. I. Best, D. L. Howard, and J. M. Umphrey, "An ultra-wideband oscilloscope based on an advanced sampling device," *Hewlett-Packard J.*, vol. 18, pp. 2–7, October 1966.

[47] W. Grove, "A dc to 12.4 GHz feedthrough sampler for oscilloscopes and other RF systems," *Hewlett-Packard J.*, vol. 18, pp. 12–15, October 1966.

[48] J. M. Umphrey, Hewlett-Packard Co., Colorado Springs Div., Colo., private communication.

[49] P. Emile, General Applied Science Lab., Inc., Westbury, N. Y., private communication.

[50] G. J. Frye and N. S. Nahman, "Random sampling oscillography," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-13, pp. 8–13, March 1964.

[51] J. T. Boatwright, "Random sampling—A statistical measurement approach," *WESCON Tech. Papers*, pt. 6, session 23, paper 4, 1966.

[52] F. W. Wenninger, Jr., "A sensitive new 1-GHz sampling voltmeter with unusual capabilities," *Hewlett-Packard J.*, vol. 17, July 1966.

[53] R. B. Patten, J. W. Sedlmeyer, and L. Fussell, Jr., "Apparatus for sampling electrical waves," U. S. Patent 3 278 846, October 11, 1966.

[54] R. Manhart, Elec. Engrg. Dept., University of Nevada, Reno, private communication.

[55] R. L. Peterson, "Switching properties of superconductive thin films," M.S. thesis, Elec. Engrg. Dept., University of Nevada, Reno, May 1966.

[56] A. J. Cummings and A. R. Wilson, "Cryogenic nanosecond pulse recirculator," *Proc. IEEE (Correspondence)*, vol. 52, p. 1749, December 1964.

[57] D. Rathbun, "Data transmission through long superconductive cables," Livermore Lab., Sandia Corporation, Livermore, Calif., Rept. SCL-TM-66-26.

[58] H. V. Shurmer, "Intrinsic frequency limitations for semiconductor microwave devices," *Radio Electronic Engrg.*, pp. 93–98, February 1966.

[59] R. L. Carbrey, "A strobed analog data digitizer with paper tape output," *Proc. FJCC*, pp. 707–715, 1964.

[60] O. L. Gaddy, "A comparison technique for measurement of millimicrosecond pulses," in *Bull. Engrg. and Architecture*, no. 43. Lawrence, Kan.: University of Kansas Publications, 1959, pp. 13–19.

[61] ——, "Theory and operation of a diode comparison technique for measuring sub-millimicrosecond pulse characteristics," M.S. thesis, Elec. Engrg. Dept, University of Kansas, Lawrence, 1959.

[62] ——, "A simple method of measuring fractional millimicrosecond

pulse characteristics," *IRE Trans. on Instrumentation*, vol. I-9, pp. 326–333, December 1960.
[63] D. C. Argouridis, "Self-sampling system for measurement of picosecond pulse characteristics," *Rev. Sci. Instr.*, vol. 33, December 1962.
[64] G. H. Honnold, "A method for determining the applicability of the Gaddy pulse comparison system to the measurement of unsymmetrical trapezoidal and nontrapezoidal pulse shapes," M.S. thesis, Elec. Engrg. Dept., University of Kansas, Lawrence, 1962.
[65] G. H. Honnold and N. S. Nahman, "On the applicability of the comparison method for picosecond pulse instrumentation," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-13, pp. 123–128, June–September 1964.
[66] R. Sugarman and F. C. Merritt, "Fast rise sampling oscilloscope and pulse generator," Brookhaven Nat'l Lab., Upton, N. Y., Rept. BNL 4146, 1959.
[67] J. C. Hubbs, "Comments on pulse measurement techniques," presented at the 1966 Colloquium on Pulse Voltage Parameters and Their Measurements (NBS unpublished document).
[68] R. L. Wigington and N. S. Nahman, "Transient analysis of coaxial cables considering skin effect," *Proc. IRE*, vol. 45, pp. 166–174, February 1957.
[69] N. S. Nahman, "A discussion on the transient analysis of coaxial cables considering high-frequency losses," *IRE Trans. on Circuit Theory*, vol. CT-9, pp. 144–152, June 1962.
[70] R. Pelissier, "The propagation of transient and steady state waves along electric lines, Parts I, II. and III" (in French), *Rev. Gén. Élec.*, vol. 59, pp. 379–399, September 1950, pp. 437–454, October 1950, and pp. 502–512, November 1950.
[71] R. Cazenave, "Distortion of a signal transmitted through a perfectly homogeneous coaxial line," *Cables et Transm. (Paris)*, vol. 5A, no. 4, pp. 279–314, 1951.
[72] P. Behrend, "Theory of pulse technique for coaxial cables," *Z. Angew. Physik (Germany)*, vol. 5, p. 61, February 1953.
[73] D. Seitzer, "The change in shape of steep fronted pulses by a uniform impedance discontinuity," *Arch. Elekt. Ubertragung*, vol. 15, pp. 303–307, 1961.
[74] R. L. Wigington, "The transient response of strip line," M.S. thesis, Elec. Engrg. Dept., University of Maryland, College Park, 1959.
[75] R. Stapelfeldt and F. J. Young, "The short pulse behavior of lossy tapered transmission lines," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-9, pp. 290–296, July 1961.
[76] D. Seitzer, "The change in shape of steep fronted pulses by multiple reflections along a transmission line," *Arch. Elekt. Ubertragung*, vol. 16, pp. 263–270, 1962.
[77] G. Brianti, "Distortion of fast pulses in coaxial cables numerical analysis and applications," MSC Division, CERN (European Organization for Nuclear Research), Geneva, Switzerland, May 3, 1965.
[78] C. E. Thompson, "Recovery of fast pulses transmitted through a coaxial cable," M.S. thesis, Elec. Engrg. Dept., University of Nevada, Reno, 1965.
[79] F. Jayne, "Just how good is braided coax," *Microwaves*, vol. 5, pp. 38–44, June 1966.
[80] C. W. Womack and N. S. Nahman, "Project JAYHAWK progress rept. for period 12 January 1957 to 11 April 1957," University of Kansas Research Foundation, Lawrence, under Contract DA 49-170-sc-1844.
[81] C. W. Womack, O. Gaddy, and N. S. Nahman, "Project JAYHAWK progress rept. for period 12 July 1957 to 11 October 1957," University of Kansas Research Foundation, Lawrence, under Contract DA 49-170-sc-1844.
[82] G. C. Temes and J. A. C. Bingham, "Iterative Chebyshev approximation technique for equalizer synthesis," Stanford Linear Accelerator Center, Stanford, Calif., Rept. SLAC-PUB-146, 1965.
[83] R. L. Peterson, "Frequency domain compensation of coaxial cables for best time domain response," Lawrence Radiation Lab., University of California, Livermore, Rept. LER 768-1, 1966.
[84] M. P. Ekstrom, "Combined amplitude and phase equalization of coaxial cables," Lawrence Radiation Lab., University of California, Livermore, Rept. LER 814-1, 1966.
[85] ——, "Combined amplitude and phase equalization of coaxial cables," M.S. thesis, Elec. Engrg. Dept., University of Nevada, Reno, in press.
[86] N. S. Nahman and G. M. Gooch, "Nanosecond response and attenuation characteristics of a superconductive coaxial line," *Proc. IRE*, vol. 48, pp. 1852–1856, November 1960.
[87] N. S. Nahman, "A miniature superconductive coaxial transmission line," Ph.D. dissertation, Elec. Engrg. Dept., University of Kansas, Lawrence, 1961.
[88] P. K. Shizume and E. O. Vaher, "Superconducting coaxial delay line," *IRE Internat'l Conv. Rec.*, pt. 3, pp. 95–99, March 1962.
[89] R. J. Allen and N. S. Nahman, "Analysis and performance of superconductive coaxial transmission lines," *Proc. IEEE*, vol. 52, pp. 1147–1154, October 1964.
[90] A. J. Cummings and A. R. Wilson, "High voltage properties of superconducting transmission lines," Edgerton, Germeshausen & Grier, Inc., Las Vegas, Nevada, Tech. Memo EGG 1183-1155, 1965.
[91] A. J. Cummings and H. Kuettner, "Superconducting high voltage delay line," presented at the 1965 14th Annual Symp. on Wire and Cable, Atlantic City, N. J.
[92] A. J. Cummings and A. R. Wilson, "High voltage pulse characteristics of superconducting transmission lines," *J. Appl. Phys.*, vol. 37, pp. 3297–3300, July 1966.
[93] ——, "New techniques for superconducting cables," *IEEE Trans. on Nuclear Science*, vol. NS-13, pp. 321–325, February 1966.
[94] B. J. Elliott, IBM Thomas J. Watson Research Center, Yorktown Heights, N. Y., private communication.
[95] Q. Kerns, F. Kirsten, and C. N. Winningstad, "Pulse response of coaxial cables," in *The Counting Handbook*. Berkeley, Calif.: Radiation Lab., University of California.
[96] "The Mark II—Single transient system," General Applied Science Labs., Inc., Westbury, N. Y., 1964.
[97] J. W. Sedlmeyer, R. B. Patten, and L. Fussell, Jr., "Analyzer for fast single events," in *Nuclear Electronics III*. Vienna, Austria: International Atomic Energy Agency, 1962.
[98] "A voltage sampling analog to digital converter (for recording nonrecurrent pulses)," Sandia Corporation, Albuquerque, N. Mex., Rept. SC-DR-66-99 under Contract AT (29-1)-789, p. 35.
[99] A. A. Fleischer and E. Johnson, "New digital conversion method provides nanosecond resolution," *Electronics*, vol. 36, April–June, 1963.
[100] H. H. Doemland and W. D. McCaa, "A cryogenic memory system," *Proc. IEEE (Correspondence)*, vol. 51, pp. 1767–1768, December 1963.
[101] W. D. McCaa, "An active cryogenic memory," M.S. thesis, Elec. Engrg. Dept., University of Kansas, Lawrence, 1964.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967
pp. 855-864
THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

Peak Pulse Voltage Measurement (Baseband Pulse)

A. R. ONDREJKA

Abstract—**Several methods are presently being used for the measurement of pulse voltage. Oscilloscopes are particularly useful because they provide information concerning the shape of the pulse, besides a measure of the peak voltage. In addition to the oscilloscope, several peak voltmeter circuits are mentioned. These include pulse stretching, sampling, and the slideback method. A standard pulse generator is described which provides a calibrated pulse voltage suitable for voltmeter calibration and other uses.**

I. Definitions of Terms and Measurement Considerations

THE voltage waveform of interest here is a train of baseband or video pulses. Ideally (in the mathematical sense), in pulses of this type the voltage excursion occurs in only one direction from zero or some fixed dc level. Usually the pulse duration is much less than the time interval between pulses, and the average voltage is small compared to the peak voltage.

The term "peak pulse amplitude" is defined as the maximum absolute peak value of the pulse excluding those portions considered to be unwanted, such as spikes [1]. Pulse rise time is defined as the time interval between the instants at which the instantaneous amplitude first reaches specified lower and upper limits (e.g., 10 percent and 90 percent) of the peak pulse amplitude.

Real pulses generated in the laboratory usually depart from the desired shape in one or more parameters. For example, in many cases the pulse may have such spurious effects as overshoot or ripple on its top. Ripple is a periodic (usually damped) variation of the pulse top which may endure for a considerable portion of the pulse duration. The

Manuscript received March 23, 1967.
The author is with the National Bureau of Standards, Boulder, Colo.

user is then faced with making a decision as to the peak amplitude. In the case of ripple, an acceptable compromise may be the determination of the average pulse amplitude which is defined as the average of the instantaneous amplitude taken over the appropriate part of the pulse duration. Overshoot generally requires a different treatment.

The accuracy of measurement of peak pulse voltage depends on the characteristics of the measuring system (which includes the measuring instrument and any associated sensor or transducer) relative to the nature of the pulse being measured. Of particular interest is the response time of the system. It is obvious that a measuring system having a rise time of 10 nanoseconds is incapable of accurately measuring the peak voltage of a pulse whose duration is 5 nanoseconds. On the other hand, measurement on a pulse with 20-nanosecond duration would be within the capability of the 10-nanosecond system, provided the pulse is approximately rectangular. In general, if the required minimum time resolution is t, the rise time of the measuring instrument should be no greater than $t/2$ for an uncertainty less than a few percent. In frequency-domain terminology, the upper frequency limit (3 dB) of the system should be at least $2/t$ and the phase characteristics should be such that the time of transmission of all significant frequency components in the pulse is constant (i.e., linear phase shift).

The precautions necessary in CW RF measurements are generally also required in pulse measurements. In addition to the aforementioned frequency and phase responses, these include such matters as 1) adequate shielding of measuring equipment, 2) loading effects of measuring equipment, 3) lead length which may introduce resonance effects, 4) proper terminations, and 5) nonlinear behavior of circuit elements [2].

II. Measurement Methods

Several methods for pulse voltage measurement have been developed in past years, some of which are in current use. The principles of operation of these are presented in the following.

A. Oscilloscopes

Oscilloscopes have found wide and significant application in the measurement of pulse voltage. They serve a dual function in pulse analysis in that they can be used to measure the pulse voltage and also provide data necessary for the delineation of the pulses. Delineation [2] is the process by which the relative shape of the pulse is described. Delineation of the pulse is important before the effects of spikes and other distortions can be considered. Even though measurements are to be performed with a voltmeter other than an oscilloscope, it is usually desirable to observe the pulse shape before measurement to avoid possible serious errors.

Oscilloscopes are generally classified into two broad categories, real-time and sampling [3]. In real-time oscilloscopes, the time necessary to display an event on the cathode ray tube is approximately equal to the actual time of occurrence of the event. For example, a pulse whose duration is 10 nanoseconds will require approximately 10 nanoseconds to be displayed. This can be contrasted with sampling oscilloscopes in which the time required to display the signal is a function of the number of samples required per display. For example, if one hundred samples are required per display, then the ratio of display time to actual time is at least 100 to 1.

The oscilloscope parameters of importance in pulse voltage measurement include rise time (or bandwidth and phase response), vertical sensitivity calibration, and vertical linearity. The rise time of the oscilloscope should be sufficiently short so that the display essentially represents the true shape of the input pulse. If the pulse rise time is faster than that of the oscilloscope, then the displayed pulse will be distorted and the displayed peak amplitude may only approximately represent the true peak amplitude. The determination of rise time (or bandwidth and phase response) of oscilloscopes is not an easy task. Methods which have been employed include the application of pulses with known rise time, and gain and phase response versus frequency measurements. Measurements on pulses having nanosecond and subnanosecond rise times are particularly difficult and it is in this range that the errors may become large.

Calibration of the vertical sensitivity and linearity is an obviously necessary procedure which should be carried out immediately prior to use of the oscilloscope.

Real-time oscilloscopes with 3.5-nanosecond rise-time capability are available commercially. Special types employing travelling-wave deflection systems have rise times as small as 0.2 nanosecond but have limited vertical sensitivity. Some have relatively high resolution because of their extremely narrow trace widths.

Sampling oscilloscopes with 0.028-nanosecond (28-ps) rise time have recently become available [4]. Certain problems exist which are peculiar to this type of instrument. One of these, commonly called "blow-by," is a distortion caused by capacitive feed-through of the input signal in the absence of a gating pulse to the sampling diode. Smoothing and shaping controls may also introduce distortions. These types of distortion can introduce errors of up to 10 percent.

Manufacturers' claims for the vertical sensitivity and linearity uncertainty are generally 2 to 3 percent. If care is exercised in the performance of the measurement, total uncertainties of 2 or 3 percent are possible. Experiments have been performed which indicate a possible minimum uncertainty of 0.3 percent in amplitude when the pulse is not severely distorted by the oscilloscope.

B. Pulse Stretchers

The method of pulse stretching was one of the earliest used in the art of peak voltage measurement. The basic principle involved is that of developing a steady direct voltage which is (in the ideal case) equal to the peak amplitude of the input signal. Measurement of peak pulse voltage is then reduced to the simple matter of measuring a direct voltage.

One of the simplest pulse stretchers is a capacitor which is charged by the pulse voltage through a diode as shown in Fig. 1. The relatively small forward resistance of the diode allows the capacitor to charge rapidly while the large diode back-resistance tends to prevent charge leakage from the capacitor. A high-input resistance voltmeter, such as an electrometer, is used to measure the voltage across the capacitor.

A major defect in the simple capacitor pulse stretcher is due to the fact that as the capacitor becomes charged to near the peak voltage of the input pulse, the forward resistance of the diode becomes large and approaches the back-value. Thus the capacitor is not able to charge to the full voltage of the input signal, and an offset exists between the capacitor voltage and the peak input voltage. If the characteristics of the diode are known, the magnitude of the offset voltage can be estimated and a correction made to the observed direct voltage. Typically, the offset voltage is of the order of 0.5 V.

Some commercially available, direct-reading peak voltmeters employ a unity gain amplifier to improve pulse stretching and to allow use of a low-resistance dc voltmeter. The input impedance of the amplifier is very high to minimize its loading effect on the capacitor.

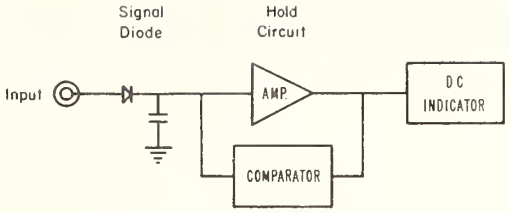


Fig. 1. Pulse-stretching voltmeter.

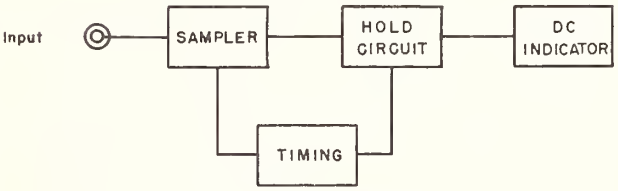


Fig. 2. Sampling voltmeter.

C. Sampling Voltmeters

The sampling method of pulse voltage measurement was designed to overcome some of the difficulties (e.g., offset voltage) described previously. In the sampling scheme shown in Fig. 2, the sampler senses the input pulse voltage for a time which is small compared to the duration of the pulse. The duration of the sample is short enough that variations in the input voltage during the sampling time are negligible. Ideally, the sampler does not discriminate between a sample taken within the constant-amplitude portion of a pulse and a sample taken from a dc signal. The output of the sampler is a pulse whose amplitude is proportional to the amplitude of the input signal at the sampled point. The proportionality factor can be determined, for example, by dc calibration. The voltage information from the sampler is then stored and stretched in such a way that it is available for measurement by conventional dc instrumentation.

The sampling scheme described is not necessarily peak reading. Two methods have been used to convert it to peak reading. In one method, a timing unit controls the position of the sample with respect to the input waveshape. The timing can be adjusted to locate the sample at the peak of the pulse. In another method [5], samples are taken at random locations and the sample with maximum amplitude is measured by a peak-reading voltmeter. This peak-reading voltmeter need not have the bandwidth of the sampler since the samples can be stored long enough for the voltmeter to respond.

A commercially available synchronous instrument has a claimed accuracy of ± 2 percent for 4-nanosecond duration pulses. An available random-sampling voltmeter offers ± 3 percent at approximately 4 nanoseconds and has a bandwidth greater than 1 GHz.

D. Slideback Voltmeter

The slideback voltmeter [6] compares an unknown voltage to an accurately known, adjustable, direct voltage. A peak pulse slideback voltmeter has been developed at the National Bureau of Standards, Boulder Laboratories, for use as a reference standard. The circuit, as shown in Fig. 3, employs a semiconductor diode which is back-biased by the dc slideback voltage V_B. In operation, a train of pulses from pulse generator PG (positive pulses for the circuit shown), applied to the input, causes the diode to conduct in the forward direction. If the amount of back-bias is increased from zero, a voltage is reached where the diode is just at cutoff and no forward pulse current flows through the diode and galvanometer G. This reverse direct voltage, which can be easily and accurately measured by the voltmeter V, is then equated to the peak pulse voltage. For greater sensitivity, a peak null indicator, such as an oscilloscope, replaces the galvanometer.

The accuracy of the system is primarily dependent on the validity of the assumption that, at a null condition, the dc bias voltage is equal to the peak pulse voltage. The validity of this assumption is, in turn, a function of the characteristics of the diode. If some voltage V_e across the diode is required to cause minimum detectable current to flow, then the dc slideback voltage would be less than the peak pulse voltage by an amount V_e, which is called the offset voltage. It is possible to measure V_e for a particular diode and apply a correction.

The choice of a diode must take into consideration a number of factors such as rise time, peak inverse rating, and forward V-I characteristics. With respect to the V-I characteristics, it is desirable that the diode offset voltage be as small as possible to allow the measurement of low peak voltages. The rise time of the diode limits the minimum

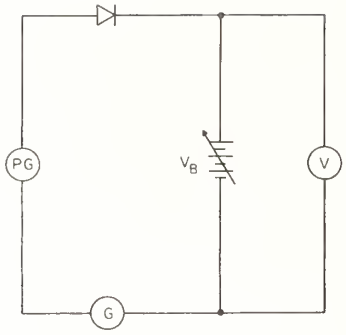


Fig. 3. Basic slideback voltmeter.

pulse width which can be measured while the peak inverse rating determines the maximum voltage range of an instrument employing a single diode. Diodes may, of course, be connected in series to increase the voltage range. Several heads, which contain the diode and load resistor, have been constructed for particular applications using various types of diodes. The 1N270 diode, for example, has a low offset voltage (40 mV) and a rise time of 10 nanoseconds. Hot carrier diodes are available with switching speeds on the order of 100 picoseconds and an offset voltage of approximately 300 mV.

In the latest version of the slideback voltmeter, the diode and bias batteries are mounted in a coaxial configuration to reduce distortions caused by undesirable reactances. This is shown schematically in Fig. 4. Using a hot carrier diode, the total rise time of the voltmeter is approximately 300 picoseconds. When proper corrections have been made to the measured value to compensate for the offset voltage V_e, the total uncertainty in the measurement can be as small as 0.3 percent. The range of the present instrument extends from 5 V to 1000 V.

E. A Standard Pulse Generating System [3]

For many applications, such as the calibration of peak voltmeters, it is often convenient to use a pulse generator whose output amplitude is known and stable. For this purpose, a zener limiter circuit has been developed which accepts pulses whose amplitudes are only approximately known and delivers flat-topped pulses whose amplitudes are constant and known to within a small uncertainty. The tandem circuit shown in Fig. 5 has proven superior to the more common single zener circuit.

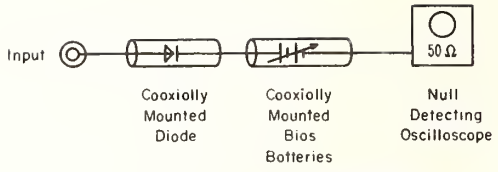


Fig. 4. Improved slideback voltmeter.

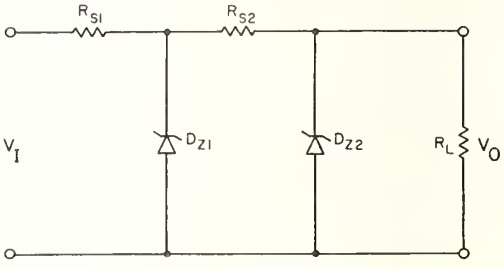


Fig. 5. Pulse limiter.

The zener limiter can be calibrated by applying direct voltage to the input and determining the voltage level at which limiting action occurs. The calibration is valid as long as the temperature of the junction is controlled and the input pulse is not of shorter duration than the rise time of the limiter circuit. The rise time of the limiter circuit can be made as short as 50 nanoseconds. Limiter circuits for several output voltages (up to 100 V) have been constructed and tested.

Analysis of errors in the standard pulse generating system indicates a maximum uncertainty of 0.25 percent. Maximum disagreement with the slideback voltmeter described previously is 0.3 percent, which is well within the maximum possible disagreement.

References

[1] "Standards on Pulses: Definitions of Terms—Part I, 1951," *Proc. IRE*, vol. 39, pp. 624–626, June 1951.

[2] "IRE Standards on Pulses: Methods of Measurement of Pulse Quantities, 1955," *Proc. IRE*, vol. 43, pp. 1610–1616, November 1955.

[3] C. N. Winningstad, "Nanosecond pulse measurements," presented at the 1961 WESCON Conv.

[4] W. M. Grove, "A dc to 12.4 GHz feedthrough sampler for oscilloscopes and other RF systems," *Hewlett-Packard J.*, vol. 17, October 1966.

[5] F. W. Wenninger, Jr., "A sensitive new 1 GHz sampling voltmeter with unusual capabilities," *Hewlett-Packard J.*, vol. 17, July 1966.

[6] A. R. Ondrejka and P. A. Hudson, "Measurement standards for low and medium peak pulse voltages," *J. Res. NBS* (*Engineering and Instrumentation*), vol. 70C, pp. 13–18, January–March 1966.

Measurement of RF Peak Pulse Power

PAUL A. HUDSON, MEMBER, IEEE

Abstract—This paper is a survey of the principal methods developed during the past twenty years for the measurement of RF peak pulse power. The basic principles involved for each method are described together with accuracies attainable under normal operating conditions. General techniques for pulse power measurement and precautions to be observed are also given.

I. Introduction

THE USE of pulse waveforms in the radio-electronics field has increased sharply in the past few years. Perhaps the best-known application of pulse modulated RF carriers is in radar systems. The number of radar and other pulse modulated carrier systems in use in the United States alone is estimated to be in the hundreds of thousands. More recent applications of pulse modulated carriers include IFF and Tacan and various forms of data-information transmission systems such as pulse-code modulation (PCM), pulse-duration modulation (PDM), and pulse-position modulation (PPM). The primary quantity of interest in pulse modulated RF transmitters is the power, and more particularly the power at the peak of the pulse.

This report includes a review of some of the salient techniques for measurement of RF pulse power.

II. Definitions of Terms

The term *peak pulse power* as used in this paper refers to *peak pulse power, carrier frequency*. The definition of the latter term is the power averaged over that carrier-frequency cycle which occurs at the maximum of the pulse of power [1]. A carrier-frequency pulse is also referred to as a burst of RF energy which endures for a finite time and is zero before and after the burst. The pulse repetition rate (PRR) is defined as the average number of pulses per second. The rise time t_r of a carrier-frequency pulse (powerwise) is defined as the interval between the instants when the power first reaches the 10-percent and 90-percent points of the power at the peak of the pulse. Also, pulse duration τ is the time interval between the first and last instants at which the power reaches a stated percentage (e.g., 50 percent) of the power at the peak of the pulse.

The duty factor for repetitive RF pulses is the product of the pulse repetition rate PRR and the pulse width τ. Typical duty factors for RF pulse systems range from 10^{-4} to 10^{-2}, and thus the instantaneous power is zero for all but a small fraction of the time. In general, peak display and peak detection techniques have been used in the design of RF pulse power measuring instruments. Descriptions of some of the more widely used measurement principles and instruments are given below.

Manuscript received March 7, 1967.
The author is with the National Bureau of Standards, Boulder, Colo.

III. Description of Salient Measurement Principles and Instruments

Various methods for peak pulse power measurement have been developed during the past 20 years. In general, the aim has been to develop instruments which indicate directly the measured quantity. Indirect methods such as calculation of the pulse power from a measurement of the average power are included in this report for the sake of completeness.

A. Calculation of Pulse Power from Average Power Measurements

The average power of a train of repetitive pulses can be measured with calorimetric or bolometric instruments, provided the time constants of the instruments are large compared to the pulse period. If the pulses are approximately rectangular in shape, the average power P_{av} and the pulse power P_p are related by the simple equation

$$P_p \approx \frac{P_{av}}{\tau PRR}. \tag{1}$$

The pulse repetition rate PRR can be measured easily and accurately with commercial electronic frequency counters. Measurements of pulse duration is usually accomplished by detecting the RF pulse envelope (as with a diode) and measuring the duration using a calibrated oscilloscope. In this measurement, it is assumed that the rise time of the diode and the oscilloscope are fast enough to yield a true representation of the actual pulse shape. If the pulse duration is large compared to its rise time, then this measurement can be quite accurate. When the pulse shape is irregular, a shape factor correction must be applied. Often the shape factor correction must be estimated and hence may be subject to a relatively large uncertainty.

B. Notch Power Meter

The notch power meter [2] is a terminating-type instrument and thus absorbs the power which it measures. It is basically a two-channel (Y) system wherein the amplitude of the detected or demodulated RF pulse envelope is compared to the detected amplitude of a reference CW level on a cathode ray tube (CRT). The CW signal is turned off (blanked) for a period of time which is larger than the duration of the RF pulse. The turn-off of the signal results in a notched (negative going pulse) display on the CRT. In commercial versions of the notch method, the RF pulse channel includes a tuned RF amplifier and a variable waveguide-below-cutoff attenuator. The variable attenuator is used to adjust the level of the input pulse so that its detected amplitude as indicated on the CRT is the same as that of the

reference CW level. A block diagram of a typical notch power meter is shown in Fig. 1.

Before performing a measurement, certain internal calibrations are necessary. First, the reference CW level is set to a known value (e.g., 1 mW) by means of a precision bridge, standard cell, and galvanometer. The gain of the two channels is equalized by alternately feeding the CW reference signal first through its own channel and then through the pulse channel. The variable attenuator in the pulse channel is adjusted to equalize the amplitude of the notch, and the attenuator dial is slipped to zero reading. After the RF pulse power is applied, the detected pulse envelope is positioned within the notch and the attenuator adjusted to equalize its amplitude to that of the notch. The input RF pulse power is then read on the attenuator scale in dBm. A typical CRT display is shown in Fig. 2.

Measurement uncertainties using commercial notch power meters range from 5 to 10 percent, and the precision is of the order of 1 to 2 percent, depending upon the input power level. One source of uncertainty is that the amplitudes of the reference CW and pulse signals displayed on the CRT are proportional to the respective peak RF voltages rather than power. Thus, even though the displayed voltages are equal, the power levels may be different if the impedances of the two channels are different. Other sources of uncertainty include the 1-mW reference CW signal and the variable attenuator.

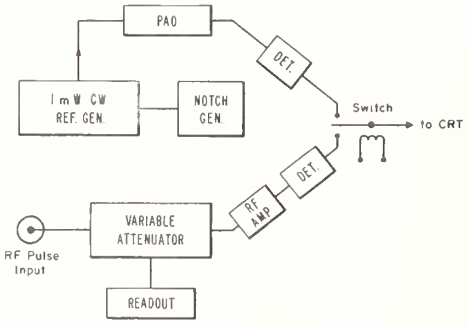


Fig. 1. Block diagram of notch power meter.

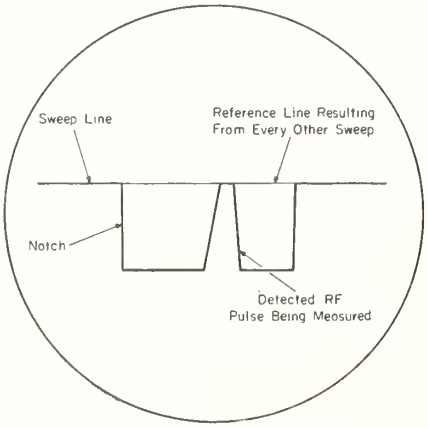


Fig. 2. Typical CRT display of notch power meter.

The power range of typical notch power meters extends from 0.1 mW to 2 kW (−10 to 63 dBm). Measurement accuracy is independent of pulse shape and duty factor over a wide range of pulse durations and PRR's.

C. Barretter Integration–Differentiation Pulse Power Meter

The basis of this technique is that any waveform may be integrated and the resulting shape differentiated to restore the original waveshape [3], [4]. A special barretter is used to integrate the input pulse of power, and the resulting waveshape is amplified and differentiated by an RC network. The resulting voltage pulse is thus proportional to the power envelope of the RF input pulse. The amplitude of the pulse is measured with a peak reading voltmeter calibrated in watts. A special calibration generator is included within commercial instruments as shown in Fig. 3. In calibrating these instruments, it is also necessary that the barretter sensitivity in terms of voltage output per mW of RF power input be known.

Sources of uncertainty in the barretter integration–differentiation technique include the calibration generator and the calibration factor or effective efficiency of the barretter mount. The effective efficiency of the mount can be determined using CW power. However, bridges employing audio substitution power should not be used because the time constant of the barretter is such that it will tend to follow the audio voltage. Manufacturers' claims for the uncertainty limit of peak pulse power measuring instruments employing the barretter integration–differentiation technique are 5 to 10 percent. Barretter mounts are available for frequencies from 500 MHz to 10 GHz for coaxial systems and to 18 GHz in waveguide. Typically, the power range extends from 5 mW to 300 mW and the measurements are not sensitive to duty factor over a wide range of PRR (50 to 10 000) and pulse duration (0.25 to 10 μs).

A new instrument based on this principle has recently been developed in the U.S.S.R. [5].

D. Methods Employing Peak Detection of RF Voltage

Semiconductor diodes have been used extensively as detectors of RF energy. As is well known, these devices tend to detect the peak voltage of the RF signal when they are provided with a filter circuit having the proper time constant. Instruments employing peak diode detection for CW power measurement have been on the market for a number of years.

The relationship between the peak RF voltage and the rms power P_{rms} is given by

$$P_{rms} = \frac{(V_p \times 0.707)^2}{R} = \frac{V_p^2}{2R} \qquad (2)$$

where R is the resistance of the load across which the diode is connected or the impedance of the coaxial line for the case of feed-through meters. The instrumentation for this type of meter involves simply a direct-reading meter calibrated in rms power.

In adapting the direct-reading peak detection method to

pulse power measurement, it has been the practice to supply additional circuitry to the instrument to furnish the current needed to drive the meter and thus give a continuous reading. The added circuit constantly compares the voltage across the meter with the instantaneous peak output of the diode (envelope peaks) and automatically adjusts the meter voltage so that it follows the peak diode output [6].

Instruments of this type, in both feed-through and termination types, are available commercially. The power range extends from 1 W to 25 kW at frequencies from 0.45 to 2300 MHz. Maximum uncertainty is claimed to be no greater than 8 percent of full scale.

In another embodiment of the peak detection technique [7], a variable dc reference voltage is set equal to the peak of the detected RF pulse envelope using a chopper and an oscilloscope as an indicator. A block diagram of the instrument is shown in Fig. 4. The input RF pulse is divided into two paths by the resistive power divider. The power in the through path is absorbed in the 10-dB pad and the 50-ohm termination. In the other path, the peak signal envelope is detected by the diode which is forward biased to an operating point of maximum stability and beyond the square-law region. Dividing the signal into two paths allows the calibration of the instrument with CW power, and the divider provides impedance isolation between the input and the diode. The chopper alternately connects the diode output and a reference dc voltage to the input of an auxiliary oscilloscope. The dc voltage is adjusted until its display on the CRT coincides with the peak of the detected pulse envelope. The reference voltage is monitored by a meter calibrated in watts.

The instrument may be calibrated using CW power by replacing the 50-ohm termination with a 10-mW power bridge to measure the power at the output of the 10-dB pad. Subsequently, the power loss between this point and the input connector must also be determined. The frequency range extends from 50 to 2000 MHz, and the full-scale power range is 200 mW. Maximum uncertainty of 0.6 dB is claimed.

Harmonic distortion in the signal being measured can cause large errors in peak-detection-type power meters. Well filtered (60 dB) sources should therefore be used with this type of instrument.

E. Sampling-Comparison Technique

In 1962, two methods for peak power measurements employing sampling-comparison techniques were developed independently at the Boulder Laboratories of the National Bureau of Standards and the Sandia Corporation, Albuquerque, N. Mex. [8], [9]. The basic principles of the methods are the same and only one of them (NBS) will be described.

In the sampling-comparison method, comparison is made between RF pulse power and CW power at the same frequency. Comparison is made by means of identical timewise samples extracted from both signals using a fast semiconductor diode switch. With reference to the block diagram of Fig. 5, the manually operated coaxial switch at point X allows the selection of either the CW or pulsed signals. The pulsed signal may come directly from the source or from the source via a directional coupler.

In making a measurement, the pulsed power is applied to the diode switch input from whence it goes to the normally closed output 1. A switching or gating pulser switches the input to output 2 for a period of time less than the RF pulse duration. The gating pulser is synchronized through a variable delay network with the pulser modulating the RF source. Thus, the switch is gated each time an RF pulse appears at the switch input. The timewise sample of the RF pulse at output 2 is detected with a thermocouple or diode detector, the dc output of which is measured and recorded. The pulsed power is then removed and CW power applied to the diode switch input via a calibrated directional coupler. Holding all parameters of the gating pulse fixed as before, the CW power is adjusted so that the detector output is the same as that previously recorded. The system thus samples the CW power in a timewise fashion as was the case with the pulse modulated power. There only remains the task of measuring the CW power incident upon the switch, and this power is equated to the peak pulse power. The diode which terminates output 1 on the switch may be

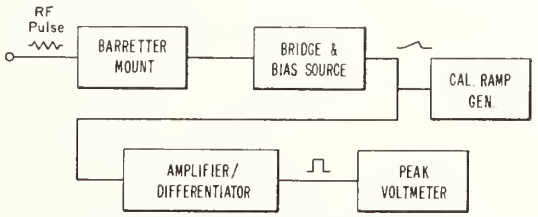


Fig. 3. Block diagram of barretter integration–differentiation method.

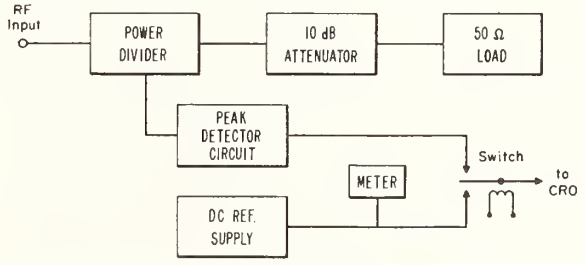


Fig. 4. Block diagram of pulse power meter employing peak diode detection—video chopper method.

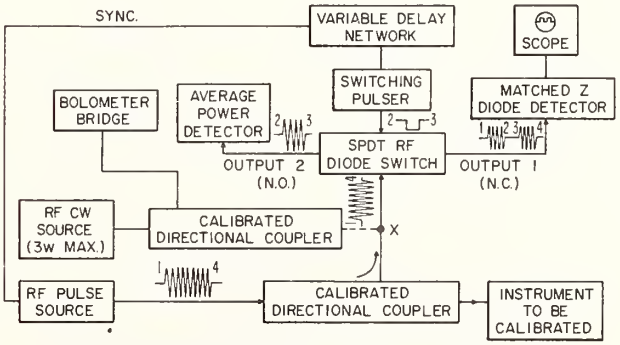


Fig. 5. Block diagram of sampling-comparison method.

used with an oscilloscope to view the pulse envelope shape and to aid in positioning the sampling pulse at the peak of the main RF pulse.

The range of the basic system extends from 0.02 W to 2 W, and the limit of uncertainty is 2 percent. When directional couplers are used to extend the range, as shown in the diagram, an additional 1 percent uncertainty is added.

Equipment used in construction of the sampling-comparison system is somewhat conventional except for the SPDT semiconductor diode switch. In order to minimize mismatch errors, the input VSWR of the switch must be as near unity as possible. An upper limit of 1.08 was set in the design of the system and commercially available units had considerably higher VSWR's. Thus it was necessary to construct a special switch [10].

Once set up, it is possible to make rapid, precise measurements with this system.

It should be noted that in the system described above it is necessary that the pulse repetition rate be very constant throughout the measurement process. Although commercially available pulse generators have adequately stable PRR's, in some cases this requirement may present a problem. Recently, a modification of the comparison technique has been reported [11] which does not require a constant PRR. The method employs sequential comparison of the pulsed RF power to a variable CW reference level at the same frequency as that of the pulsed carrier. With reference to Fig. 6, a known fraction of the RF pulse power is coupled to the secondary arm of a 4-arm calibrated coupler from whence the sample is fed to a receiver (detector) through a second coupler. The video output of the receiver goes to an oscilloscope and a trigger unit which gates a CW reference source *on* for a predetermined time in the interval between pulses. The output of the gated CW source is connected to a *standard* power meter through the secondary line of the first coupler and the primary arm of the second coupler. A portion of the CW power is coupled to the receiver by the second coupler, and its envelope is displayed on the oscilloscope along with the pulse envelope. The trigger unit causes the timebase to be initiated twice in each pulse-modulation period. Thus the CW and pulse envelopes are superimposed on the CRT, allowing precise adjustment of the reference level to that of the pulse. The peak pulse power is given by the power level as measured by the *standard* meter multiplied by the coupling coefficient of the calibrated coupler and the factor $n+1$, where n is the off/on ratio of the reference source.

Except for the trigger unit, the necessary components for constructing the system are readily available. An uncertainty limit of 3 percent is claimed by the developers of the system.

IV. Measurement Techniques and Sources of Error

In addition to the usual sources of error in RF power measurement, there are several additional ones which are peculiar to peak pulse power. Examples of the former include mismatch error, unaccounted losses between different measurement points, and errors in calibration of directional couplers or pads used as range extenders. A source of error in pulse power measurement is due to distortion in peak envelope detectors. Distortion in the detected envelope may be caused by harmonic distortion in the signal, or it may arise because the diode has a limited rise time response and therefore cannot follow the envelope. In many modern pulsed systems, including radar, the rise time of the modulating pulses is of the order of 1 ns with pulse durations of the order of 100 ns. Fast diodes are therefore required. As mentioned earlier, errors due to harmonic distortion in the RF signal can be reduced by adequate filtering of the generator.

Another source of error is due to bandwidth limitations which may exist in the pulse power measuring instrument. As is well known, a pulse modulated carrier has sidebands associated with it which contain an appreciable fraction of the power. The spectrum envelope of a rectangular pulse modulated carrier is shown in Fig. 7. The lobe width is $1/\tau$, where τ is the pulse duration. The spectrum is discrete with the spacing between lines being equal to the PRR. Approximately 99 percent of the power is contained in the frequency band $f_0 \pm 10/\tau$, and therefore the bandwidth of the measuring instrument should be at least this wide. Care must be exercised in using high-Q tuning devices on pulsed systems, such as in impedance matching or filtering. A system which is sharply tuned at f_0 may reflect an appreciable amount of sideband power. In many cases it is advisable to check the frequency distribution of the pulsed signal

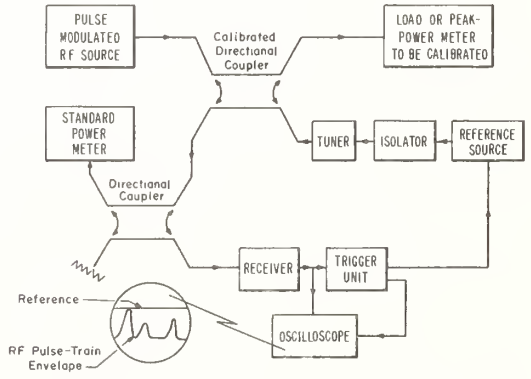


Fig. 6. Block diagram of method of pulse power measurement by sequential comparison with a reference.

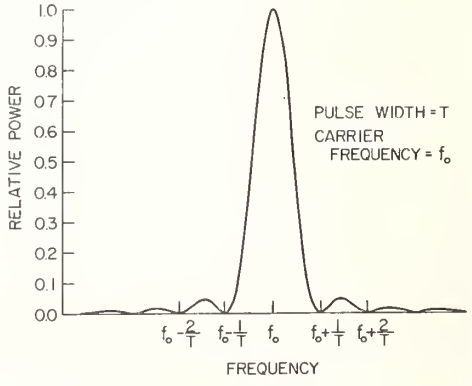


Fig. 7. Power spectrum envelope of rectangular pulse modulated RF carrier.

at the measurement point with a spectrum analyzer.

Leakage effects may also be a source of error, particularly when a low-level power meter is used to measure a relatively high power level with a directional coupler or attenuator pads. Interfering fields may come from either the RF generator or the modulator, and it may be necessary to provide shielding for the measuring instrument.

Tests and calibrations of most of the instrument types described previously indicate that, by and large, their error limits are within the manufacturers' specifications [12]. In critical applications, however, the user would do well to have his instrument calibrated or compared to a standard or other instrument of known error limit.

References

[1] "Standards on Pulses: Definitions of Terms, Part II, 1952," *Proc. IRE*, vol. 40, pp. 552–554, May 1952.

[2] D. F. Bowman, "The notch wattmeter," *Proc. Nat'l Electronics Conf.*, vol. II, pp. 361–371, October 1946.

[3] "Direct measurement of peak power," *Sperry Microline Monitor*, vol. 1, January 1961 (in-house publication of Sperry Microwave Electronic Co.)

[4] U. S. Patent 2 776 406.

[5] N. A. Pelykh, "Bolometric meter of pulsed UHF power," *Izmeritel'naya Tekhnika (USSR)*, pp. 57–59, February 1966. English transl.: *Measurement Techniques*, pp. 226–229.

[6] "RF directional wattmeter for peak or average power," *Watts New from Bird*, vol. 3, July–August 1966 (in-house publication of Bird Electronic Corp.).

[7] G. R. Polen, "Precision peak pulse measurements with the peak power calibrator," Boonton Radio Co., Notebook 35, 1964.

[8] P. A. Hudson, W. L. Ecklund, and A. R. Ondrejka, "Measurement of RF peak-pulse power by a sampling-comparison method," *IRE Trans. on Instrumentation*, vol. I-11, pp. 280–284, December 1962.

[9] C. A. Denney, C. L. Mavis, and C. J. Still, "Microwave pulse-power measurements," *IRE Trans. on Instrumentation*, vol. I-11, pp. 276–280, December 1962.

[10] W. L. Ecklund, "A low input VSWR coaxial diode switch for the UHF band," *IEEE Trans. on Microwave Theory and Techniques (Correspondence)*, vol. MTT-12, p. 359, May 1964.

[11] D. W. G. Bradstreet and E. J. Griffin, "Measurement of pulse-modulated RF power by sequential comparison with a reference," *IEE Electronics Lett.*, vol. 2, pp. 316–317, August 1966.

[12] S. Evendorff, "Microwave peak power measurement," *J. ISA*, vol. 12, pp. 71–75, May 1965.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967
pp. 851-855
The Institute of Electrical and Electronics Engineers, Inc.

Analysis and Performance of Superconductive Coaxial Transmission Lines

R. J. ALLEN, MEMBER, IEEE AND N. S. NAHMAN, SENIOR MEMBER, IEEE

Summary—This paper discusses the mechanisms affecting metallic conduction and dielectric properties in relation to a low temperature transmission line to provide a basis for analysis of the characteristic change, and the remaining losses at low temperature. Specific attention is given to lines employing Teflon or polyethylene as dielectrics, and operating up to 10 Gc. The construction and measured performance of a number of long superconductive coaxial lines is given in support of the analytical results.

Introduction

THE application of cryogenic techniques to the design of low loss coaxial transmission lines has provided lines with significantly improved characteristics. Attenuation and bandwidth of the small diameter cryogenic lines compare favorably with those of the largest of the room temperature cables.

The decrease in the series resistance loss of the metallic conductors at low temperature undergoes a further marked decrease when the metals enter the superconductive state. The shunt conductive loss due to the dielectric also decreases with decreasing temperature. Consequently, when operating at liquid helium temperature (4.2°K), the line exhibits characteristics approaching those of an ideal line.

Manuscript received June 11, 1964.

R. J. Allen is with Molecular Electronics Division, Westinghouse Electric Co., Baltimore, Md. He was formerly with The Bunker Ramo Corporation, Silver Spring, Md.

N. S. Nahman is with the University of Kansas, Lawrence, Kan.

Theory

Transmission Line Equations

The voltage on an infinitely long transmission line at a point x units from the line input terminals is given by

$$V(x, s) = V(0, s)e^{-x\Gamma(s)} \tag{1}$$

where $v(0, s)$ is the input voltage ($x=0$) and $\Gamma(s)$ the propagation function,

$$\Gamma(s) = \sqrt{Z(s)\,Y(s)} \tag{2}$$

in which $z(s)$ and $y(s)$ are the series impedance per unit length and the shunt admittance per unit length, respectively. $z(s)$ and $y(s)$ are analytic functions of the complex variable s which reduces to $j\omega$ on the real frequency axis, *i.e.*, for real frequencies, $s=j\omega$, and

$$Z(i\omega) = R(\omega) + jx(\omega), \qquad \text{ohms per unit length} \tag{3}$$

$$Y(i\omega) = G(\omega) + jB(\omega), \qquad \text{mhos per unit length.} \tag{4}$$

The voltage expression (1) is also valid for a line of finite length which is terminated in the line's characteristic impedance, $Z_0(s)$,

$$Z_0(s) = \sqrt{\frac{Z(s)}{Y(s)}}. \tag{5}$$

For low loss high frequency coaxial lines the series impedance per unit length consists of two terms, the inductive reactance per unit length sL and the skin effect impedance per unit length $z(s)$

$$Z(s) = sL + z(s) \tag{6}$$

in which the geometrical factors are the usual factors,

$$L = \frac{\mu}{2\pi} \ln\left(\frac{r_2}{r_1}\right), \qquad \text{henrys per meter} \tag{7}$$

$$z(s) = \frac{Z_{A1}}{2\pi r_1} + \frac{Z_{A2}}{2\pi r_2}, \qquad \text{ohms per meter.} \tag{8}$$

The subscripts 1 and 2 relate to the pertinent inner and outer coaxial conductors, respectively. r denotes a conductor radius, and Z_A the conductor surface impendance per unit square.

Also, for low loss high frequency coaxial lines the shunt admittance per unit length is given by

$$Y(s) = \frac{2\pi\epsilon_0 s}{\ln\left(\frac{r_2}{r_1}\right)} K^*(s) \tag{9}$$

where ϵ_0 is the mks free space dielectric constant, and $K^*(s)$ is the frequency dependent complex relative dielectric constant. r_1 and r_2 are the same as those denoted in (7) and (8).

The Metallic Normal Conduction Losses

In general, the high frequency surface impedances of the coaxial conductors contain the effects of the conductor losses; the real part of the complex surface impedance represents the metallic conduction loss while the imaginary part gives the related phase shift. For example, the lossless coaxial line has the equivalent circuit per unit length shown in Fig. 1(a) while the presence of a nonzero surface impedance $Z_A(j\omega)$ modifies the equivalent circuit as shown in Fig. 1(b). $z(j\omega)$ represents the series impedance per unit length contributed by the surface impedances of the two conductors, (8). This equivalent circuit [Fig. 1(a)] follows from a series expansion of (2) in which $Z(s)$ is given by (6), $Y(s)$ by sC and under the condition $z^{(s)}/sL \ll 1$ [1].

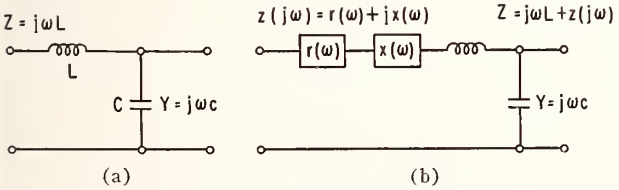


Fig. 1—Equivalent circuit per unit length for a lossless coaxial line.

In normal conductors at room temperature (as contrasted to superconductors) $z(j\omega)$ is proportional to $\omega^{1/2}145°$ hence, the real and imaginary components of $z(j\omega)$ are of equal magnitude. This follows from the mechanism of high frequency skin effect as found in a homogeneous plane conductor.[1]

At low temperatures of the order of 10°K normal conductors may exhibit the anomalous skin effect which occurs when the mean free path of the conduction electrons e exceeds the classical skin depth of penetration δ, *i.e.*, the skin depth as calculated from the dc conductivity is less than the electronic mean free path [2], [3]. Under such a condition, the electric field in the metal is not spatially constant over the distance defined by the mean free path. Furthermore, this situation is characterized by saying that the electric field at a point in the metal conductor is not a local field but rather a nonlocal field. In other words, the usual current density relation,

$$\bar{\jmath}(r) = \sigma\bar{E}(r); \qquad r \equiv x, y, Z \text{ spatial coordinates} \tag{10}$$

defines the current density $\bar{\jmath}$ at the spatial position r as being proportional to the electric field $\bar{E}(r)$ at the point r, the constant of the proportionality being the conductivity σ; for the anomalous case, the usual current density expression is replaced by

$$\bar{\jmath}(r) = \int_V f(\bar{E})dV \tag{11}$$

in which $\bar{\jmath}$ is proportional to a volume integral of a function of the electric field integrated over the entire volume of the metal. Hence, the current density at the point r is not simply proportional to the electric field at the point r; and the dc conductivity σ no longer provides the means of specifying $\bar{\jmath}$ or $\bar{E}$ from $\bar{E}$ or $\bar{\jmath}$, respectively.

When anomalous skin depth is *much less* than the electronic mean free path, the skin effect phenomenon is said to be in the *extreme anomalous limit*, and in this limit the anomalous surface impedance produces a $z(j\omega)$ that varies as $(j\omega)^{2/3}$, which obviously varies at faster rate vs ω than the classical $z(j\omega)$ variation of $(j\omega)^{1/2}$. Aside from the more rapid rate of increase vs ω, the magnitude of the anomalous skin effect is greater than that of the classical effect. Hence, the anomalous effect produces a skin loss (impedance) which is greater than the skin loss produced by the classical skin effect.

The Metallic Superconduction Losses

In superconductors local and nonlocal current densities are also encountered; however, in superconductors, the onset of local or nonlocal conditions depend upon the relative sizes of the coherence distance ξ_0 and the electric-field penetration depth λ rather than the normal state mean free path l and electric field penetration depth δ.

Essentially, the superconductive state can be characterized as a condensation of the entire electron gas of

[1] For coaxial lines, as long as the skin depth is much smaller than the respective conductor radius, the conductor surface can be considered to be a plane surface.

the normal state into a single state of bound electron pairs. The binding energy creates a gap which exists between the lower energy super state and the higher energy normal state. Thus, if a superconducting material is subjected to radiation of high enough energy, the bound electron pairs will be broken up and raised into the higher energy normal conducting state. According to the Bardeen, Cooper, and Schrieffer [4] theory at a temperature T such that $T/T_c \leq 0.5$, the energy necessary to break up a bound electron pair is approximately given by

$$\epsilon_0 = 3.5kT_c, \tag{12}$$

where T_c and k are the critical temperature below which the conductor remains superconductive in zero magnetic field and Boltzmann's constant, respectively. Consequently electromagnetic radiation at frequencies of the order of ϵ_0/h (h being Planck's constant) will excite transitions from the super to normal state.

The coherence distance ξ_0 can be thought of as the distance between the bound electron pairs of the superconductive state. Because the coherence distance is very large in terms of atomic interaction distances, the scattering processes which can scatter a normal state electron do not frequently act simultaneously on each electron of the bound electron pair. Consequently, the superconducting current can persist as the scattering process responsible for the normal state finite conductivity can do little to disrupt the pairing of the bound electron pairs in the superconductive state.

When the coherence distance ξ_0 is large compared to λ, which is the superconducting current's penetration depth into the superconductor, the superconducting current density is a nonlocal one. When the opposite situation obtains, the current density is a local one. A superconductor that principally has a local current density is called a "London" metal.

Bardeen, Cooper, and Schrieffer [4] formulated a tractable theory for superconductivity, and Mattis and Bardeen [5] extended the Bardeen, Cooper, and Schrieffer theory to include time dependent electric fields and nonideal superconductors (in the normal state $l \ll \infty$).[2] The significance of Mattis and Bardeen's result for the current density is that their current density in its most general form included local and nonlocal characteristics and applied to both the super and normal state. Furthermore, their expression for the surface impedance is an analytic function of the complex frequency variable S which satisfied the causal conditions [6], [7] required by physical systems.

The Bardeen, Cooper, and Schrieffer [4] theory depends upon a knowledge of three parameters, the critical temperature T_c, or the energy gap at $0°K$, the density of normal electronic states, $N(0)$ and the velocity at the Fermi surface v_0. Numerous studies [8] have shown that from consistent experimental data three such parameters can be deduced so as to account for the experimental data with the Bardeen, Cooper, and Schrieffer [4] theory. Consequently, if the three parameters are known for a given superconductor, the high frequency surface impedance can be closely predicted. However, it is necessary that the parameters must be known for the *specific* sample of the superconductor concerned; impurities, strains, etc. can create large differences between different samples of the same metal. A more detailed discussion of the effects of impurities and stains in relation to superconductive coaxial lines along with a detailed derivation of the Mattis and Bardeen surface impedance is given in [3].

Finally, it should be said that the theory of superconductivity is as yet far from completely resolved. For a lively discussion of the subject, the reader is referred to the recent writings of Anderson and Matthias [9].

In concluding this section, we can summarize our previous discussion by stating that the surface impedance of a superconductor is such that the high frequency losses increase sharply when the operating frequency approaches the energy gap frequency ϵ_0/h; above the energy gap frequency, the losses rapidly and asymptotically approach the normal state losses of the anomalous skin effect. For frequencies well below the energy gap frequency (microwave frequencies and downward) the losses per unit length are very small. Consequently, for short samples, such losses are negligible. On the other hand, for long samples, such as those used in microsecond delay lines, the losses are not negligible, and their presence imposes a bandwidth limitation.

Dielectric Properties

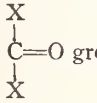
Dielectric Constant: Teflon and polyethylene are long chain polymers with as many as 100,000 X—C—X groups per molecule. Their dielectric constants are approximately 2, and frequency independent from 10^3 cps to 10^{10} cps. Theoretically, the dielectric constants should remain constant up to the optical region (5×10^{14} cps), except for absorption lines in the infrared. The losses are low and relatively insensitive to frequency. These dielectric properties serve to indicate that the largest contribution to dielectric constant comes from atomic and electronic polarizations. Examination of the symmetry of the pure molecule shows no net dipole moment, since the C—C bond has no dipole moment and the C—X bonds, which have dipole moments, are symmetrically arranged on the chain. In FEP Teflon, the X can be either an F or a CF_3. (The total amount of the fluorinated propylene is about 15 per cent by weight.) The position of the CF_3 on the chain is random. Because the CF_4 molecule has no dipole moment, the CF_3 group has the same dipole moment as a single C—F group, and the net dipole moment of the F—C—F

[2] The normal state mean free path l does not enter into the Bardeen, Cooper, and Schrieffer calculation of the superconductive current density if the superconductor is assumed to be ideal. Under such conditions, l is not limited to a relatively small value by scattering mechanisms, but rather it is large enough such that $\lambda/l \to 0$.

and CF_3—C—F groups is the same. It is assumed that the polarizations of the C—F bonds are additive. It follows that, even for the FEP Teflon, there will be no net dipole moment per molecule, without which the molecule has no tendency to align itself with the electric field. The motion of molecules in this fashion is usually a major cause of dielectric loss. While this mechanism is not present for the chemically pure material, impurities may cause dipole moments and thus cause losses.

The electric field also causes the charge in a molecule to redistribute itself which gives rise to an induced dipole. When molecules are brought together to form solids, the magnitude of the induced dipole moment per molecule may be altered a great deal by the forces of the atoms of neighboring molecules.

However, in Teflon and polyethylene, the covalent chemical bonds correspond to very short range interactions, such that only members of the particular chemical bond interact significantly with one another. For such materials, the polarization of the molecule in the solid is equal to the polarization of the isolated molecule. Furthermore, the polarization of the molecule is the sum of polarization of the different chemical bonds. Such bonds do not contribute losses as they will have absorption only in the infrared and optical regions. Since the polarization of the



group is known,

one can demonstrate that Teflon and polyethylene [10] fall into this class. The following calculation demonstrates this. The polymer molecules tend to be parallel to each other. Such regions are referred to as crystalline, while between the crystalline regions there are amorphous regions, so called because of their lack of regularity. The size of the crystalline region is quite a bit less than the mean length of a polymer chain, of 10^{-4} to 10^{-3} cm. For this reason, crystalline order does not extend over wavelengths greater than 10^{-4} to 10^{-3} cm. For 3-cm wavelengths, the material can be thought of as composed of randomly oriented molecules, and the Clausius-Mossotti formula is used to compute the dielectric constant. This formula is

$$\in = \frac{1 + \frac{8\pi}{3}\frac{\rho}{M}N_0\alpha}{1 - \frac{4\pi}{3}\frac{\rho}{M}N_0\alpha} \tag{13}$$

ρ is the density of the material, M is the molecular weight, N_0 is Avogadro's number, and α is the molecular polarization.

The molecular polarization depends on the number and the types of bonds in the molecule.

$$\bar{\alpha} = \bar{\alpha}_0 + \bar{\alpha}_1 n \tag{14}$$

where $\bar{\alpha}_0$ is the mean polarization of the end group and $\bar{\alpha}_1$ is the mean polarization of each X—C—X group. Using values of bond polarizabilities given in Landolt and Bornstein [11] for C—C and C—H gives $\bar{\alpha}_1 = 20.6$. This is about 10 per cent higher than the correct value for the H—C—H group, which is given as $\bar{\alpha}_1 = 18.3$. Using this value, $\epsilon = 2.30$ for polyethylene, in the limit as $n \to \infty$. The experimental value of ϵ corresponding to the same density used in the calculation was 2.3 [12]. The close agreement indicates 1 per cent or less of the dielectric constant is a result of these bonds. Because these bonds will not absorb or scatter, nonconstructively, energy from microwave electromagnetic field, the material is essentially lossless. Since the pure material, ignoring structural defects, is lossless, it follows that the impurities must give rise to the losses, assuming there is no interaction between molecules. Actually, there is a very weak interaction which can give rise to small losses. The picture of a lossless material with a small percentage of impurities causing small losses is consistent with the experimental facts.

It should be mentioned that the opaque character of the material is due to its macroscopic structure. The theory would predict that Teflon and polyethylene should be optically transparent with an index of refraction of 1.4. The opacity is due to the inhomogeneity of the material over wave lengths of the order of 10^{-2} cm. During crystallization of the polymer, many crystallites grow radially from a common nucleus, and the resulting "spherulitic" aggregate made up of many crystallites will be large enough to scatter visible light. Since the scattering is incoherent, the material is opaque [13]. Upon heating to the gel point, the spherulites dissolve and the material is transparent. The index of refraction squared differs from the dielectric constant by only a few percent; therefore, the major part of the polarization is electronic.

Since the absorption lines corresponding to the nuclear polarizaion are in the infrared and those due to the electronic polarization are in the ultraviolet, it is clear that these polarizations are lossless in the microwave region and below.

Dielectric Losses

Losses due to electrostriction: Losses in the pure material arise from electrostriction and electronic conduction. The former comes from the fact that under the influence of an electric field, the material contracts along the field and expands perpendicular to the field. The expansion and contraction are not quite reversible. Since the acoustical cutoff frequency of the material is above the microwave region, the alternating field will set the material into vibration. While Teflon does not have a lattice in the same sense as quartz, Teflon will

have a phonon spectrum corresponding to different modes of vibration. The interaction between modes will cause absorption of energy. All lattice modes except for that corresponding to the frequency of the electric field are in thermal equilibrium. The rate at which phonons produced by the electric field are absorbed and re-emitted at a different frequency will depend on the product of the total number of phonons present at both frequencies. In any such process, at least three types of phonons must be involved. The rate of energy absorption will be of the form

$$W = \sum_{w_2} \sum_{w_1} \bar{n}_1(\hbar w_1)\bar{n}_2(\hbar\omega_2)\frac{\epsilon_0^2}{8\pi} M(w_1, w_2, w) \qquad (15)$$

$\bar{n}_1(\hbar w_1)$, and $\bar{n}_2(\hbar w_2)$ are the average number of phonons present in their respective lattice modes which can be written as

$$\frac{1}{\epsilon^{(\hbar\omega_1/kt)} - 1} \quad \text{and} \quad \frac{1}{\epsilon^{(\hbar w_2/kt)} - 1} \cdot \frac{\epsilon_0^2}{8\pi}$$

is the energy density of the field and hence proportional to the number of phonons present. $M(w_1, w_2, w)$ is a measure of the coupling of the electric field to the lattice modes of frequency ω_1, ω_2 and restricts w_1, w_2 to the same magnitude of (w) to give M nonzero value.

Since w is $2\pi \times 10^{10}$ sec^{-1} and $T = 4.2°$K or greater, one can write $\bar{n}_1 n_2$ as

$$\bar{n}_1\bar{n}_2 = \frac{K^2T^2}{\hbar w_1 \hbar w_2} \qquad \text{for } T > 1°\text{K}. \qquad (16)$$

It follows that this process varies as T^2. Interactions involving more kinds of phonons will vary as higher powers of T. The form of M (w_1, w_2, w) depends on the actual structure of the material, including properties analogous to dislocations and impurities which are active in phonon scattering. Whether or not these contribute significantly to the losses can only be answered by computing M (w_1, w_2, w), which is not simple. Nonetheless, at worst, the temperature dependence indicated suggests the contribution from this process decreases by a factor of 5000 as the temperature decreases from 300° to 4°K.

Losses due to dipole moments: While the ideal polymer carries no net dipole moment, deviations of the molecule from its ideal folded structure could cause sections of the molecule to have a net dipole moment.

Noncancellation of these dipole moments results in a very weak interaction between dipoles and the lattice. A true orientational polarization will not be possible because the molecules' overall length causes them to be frozen in place. The interaction will be temperature dependent and show an exponential decrease analogous to that of a hindered rotation.

Loss due to conducting particles: Teflon and polyethylene have a finite conductivity, although extremely small, which obeys a law of the form

$$\sigma = \sigma_0 \exp(E/KT). \qquad (17)$$

E is the binding energy of the weakest bound electrons. For the pure material, $E \approx 10$ ev. A calculation by Pao and Bjocklund [14] indicates 10.07 ev between the $3s$ and $2p$ bands. Actually, the observed value is 1.1 ev. Impurities such as free radicals, although much less numerous, probably give rise to most of the free electrons because the binding energy of the electrons is less (about 1 ev). It follows that the loss mechanism depends on an exponential law and, hence, this contribution decreases exponentially. A Maxwell-Wagner type of polarization is held unlikely, since it would be inconsistent with the observed homogeneity of the material and the observed temperature dependence of the conductivity.

The Effect of Impurities: Impurities in the polymer can be

1) Chemical groups substituted for hydrogen or fluorine on the chain.
2) Free radicals at the chain end.
3) Trapped chemical impurities.

Presence of chemical groups in place of the hydrogen or fluorine on the chain: Chemical impurities can produce loss if they give the molecule a net dipole moment or if the impurity group has an absorption line in the microwave region. The former is considered unlikely since the loss tangent is flat in the microwave region, while the presence of an absorption line would be expected to produce a peak. On the other hand, dipole groups can give rise to the broad spectrum type of losses observed.

An example of such a mechanism would be the introduction of a carbonyl group, C=O, in place of the two hydrogens on the chain of polyethylene. Such an impurity can be introduced through oxidation. This has been proposed as a principal mechanism for losses, [15], [16]. The carbonyl group possesses a dipole moment. The potential is double valued, and there are two equilibrium positions. Classically, the carbonyl group cannot go from one position to another unless it has a sufficient amount of energy to overcome the potential barrier. This type of motion is a hindered rotation. In the crudest approximation, if ω is the frequency of vibration in the potential well, the frequency with which the particle has enough energy to surmount the barrier is

$$\omega_{12} = \omega e^{-V_0/KT}. \qquad (18)$$

Okamoto and Takeuchi [17] have studied the dielectric constant of high density Polyethylene from $-75°$ to 100°C with frequencies from 30 to 10^6 cps. They have found three loss peaks with activation energies of 25, 20, and 6 K cal per molecule (1.7×10^{-12}, 1.4×10^{-12} and 4.8×10^{-13} erg). The three loss peaks would imply that

the model of the potential well was oversimplified, as it should have more than one metastable position.

At low temperatures all of the positions are essentially frozen in.

Effect of free-radical and trapped chemical impurities: The principal effect of impurities will be to provide sources for electrons or traps. From the observed dc conduction of Teflon and polyethylene, it would appear that these impurities provide the majority of the electrons. For example, the temperature dependence of the dc conductivity of Teflon has been observed to be [18]

$$\sigma = \sigma_0 \exp(-1.1/KT).$$

Since the electrons of the C—F are held much tighter than 1.1 ev, impurities must contribute most of the electrons. Again the mechanism will contribute a very small loss at liquid helium temperatures.

In addition to providing conduction electrons, chemical impurities can also absorb energy if they possess a dipole moment. The best example of this would be water, which possesses a very large dipole moment, and will, if absorbed, cause the dielectric constant to rise. Teflon and polyethylene tend to absorb very little water. At low temperature the dipole moment of the water molecules will be frozen in place as the potential in which it aligns itself will have barriers much greater than KT at 4°K or 1/3000 ev.

Conclusions—Dielectric Properties: The mechanisms which are plausible enough to appear in the dielectric are all temperature dependent. The slowest temperature dependence corresponded to the loss mechanism which varied as T^2. If the dominant contribution of the losses at room temperature were due to these processes, the losses would decrease by a factor of 2×10^{-4}, giving a loss tangent of 4×10^{-8}. More rapid decay would be caused by terms having exponential behavior.

Migration of conducting particles at room temperature is insufficient by many orders of magnitude to account for existing losses. The effect of C=O groups does give the right order of magnitude for room temperature loss in highly oxidized polyethylene. The temperature dependence of relatively pure polyethylene does not decrease as fast as the simple model for C=O group would predict for $V_0=4.8\times10^{-13}$ erg.

It would be expected that the bending of the molecule would cause sections of it to have a net dipole moment. These sections will tend to align themselves with the field. The rotational barrier should be less than that for the carbonyl group, although it should be of the same order of magnitude.

It seems reasonable to conclude that the loss tangent of Teflon and polyethylene is less than 2×10^{-7} at $T=4.2$°K. Other dielectrics such as Cyanocel, polyvinyl chloride, and Tedlar which have much higher losses at room temperature, should have losses comparable to Teflon and polyethylene at 4.2°K, since most of their losses arise from dipole groups and decrease exponentially with $1/T$.

Delay Line Construction

Several types of coaxial lines were fabricated, using both polyethylene and Teflon dielectric.[3] The polyethylene line consisted of a 7 mil diameter niobium center conductor, 25 mil OD extruded polyethylene dielectric and an outer conductor of vacuum deposited lead. The line was 474.8 feet long and had a characteristic impedance of slightly under 50 ohms. Teflon dielectric lines had 15 mil center conductors and 51 mil OD extruded Teflon insulation to give 50 ohms characteristic impedance. Two types of lead outer conductor construction were used, vacuum deposition and pressure welding of sheet lead [3]. The OD of the heavy lead lines were approximately 90 mils. (See Fig. 2.) The Teflon insulated lines were approximately 1360 feet long, which provided 2 μsec delay.

Oxidation of the thin vacuum deposited lead outer conductor initially caused some deterioration of the line characteristics which was overcome by coating each layer on the spool with a film of vacuum deposited tin.

Fig. 2—Two μsec heavy lead FEP Teflon line.

[3] The seven mil niobium was obtained from the Fansteel Metallurgical Corporation, North Chicago, Ill. The polyethylene extrusion was done through the courtesy of the Polychemicals Division, Chestnut Run Plant, of the E. I. DuPont Corporation, Wilmington, Del. The 15 mil niobium was obtained from both Fansteel and the Kawecki Chemical Co., Boyertown, Pa. The FEP Teflon dielectric was extruded by American Super-Temp Wire Co., Winooski, Vt. The heavy lead jacket was applied by the National Lead Company, Perth Amboy, N. J.

The ends of the vacuum deposited lead lines were protected by encasing them in 18 to 20 inch lengths of lead tubing. Tubing for the 51 mil line was provided by lengths of pressure welded lead removed from other lines. The protective tubing for the 25 mil polyethylene line was made from resin core 60-40 electrical solder by removing the resin. The ID, fortunately, coincided exactly with the outer diameter of the cable. The solder also superconducted at above 4.20°K, although its critical current was not measured.

Connections to the 7 mil polyethylene line were made directly with Graymar "red-line" connectors which were mated to TNC adaptors and connectors and immersed in liquid helium. The coaxial leads extending from the TNC connectors were about 10 inches long and consisted of modified RG-141 cable. The outer conductor was removed from the conventional RG-141 cable and replaced with 0.003 inch wall stainless steel tubing. TNC connectors were used at room temperature.

The 51 mil dielectric Teflon lines used a tapered transition to the modified RG-141 cable instead of connectors immersed in the liquid helium. The transition had a length of $1\frac{7}{16}$ inches over which the center conductor outside diameter and the dielectric ID increased uniformly from 15 mils to 37 mils while the dielectric OD increased from 51 mils to 119 mils. GR-874 connectors were used at room temperature. An assembled vacuum deposited line providing 2 microseconds delay is shown with cryostat in Fig. 3.

Fig. 3—Assembled vacuum deposited outer conductor line with cryostat.

Performance Characteristics

The performance data germain to the analysis includes pulse response characteristics, total insertion loss, and delay time. The first two parameters can be related to losses while the third gives dielectric constant.

The pulse response for a two microsecond heavy lead jacketed FEP Teflon line is given in Fig. 4. Input and output pulses are shown superimposed on a 1 nsec per division scale. Rise time is seen to be under 0.5 nsec. It is seen that there is virtually no attenuation. Slightly poorer characteristics were obtained with the vacuum deposited outer conductor line (Fig. 5).

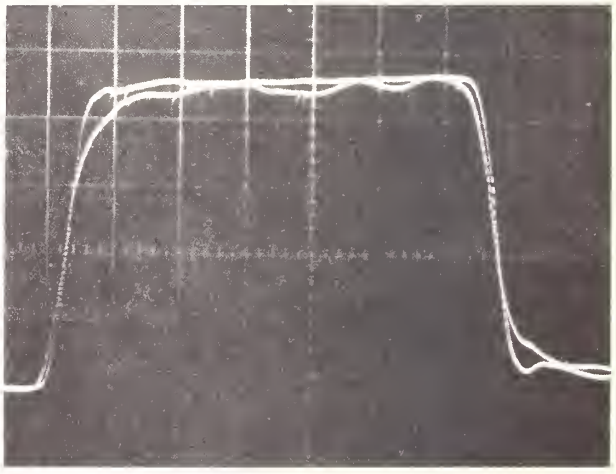


Fig. 4—Pulse response for heavy lead jacketed delay line. Scale 1 nsec division.

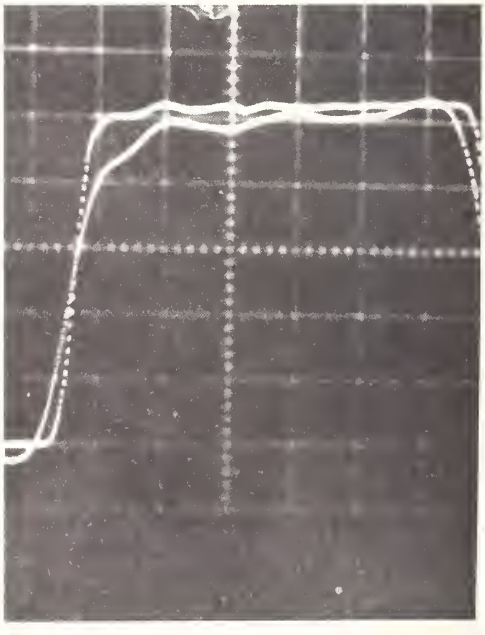


Fig. 5—Pulse response of vacuum deposited lead delay line. Scale 1 nsec/division.

Total insertion loss was measured up to 8 Gc using matched pairs of directional couplers and bolometers. Data for a heavy lead jacket line is given in Table I. The insertion loss of 8.2 db at 7 Gc can be studied to relate the measured loss to the analytical prediction.

An approximate separation of losses can be made by assigning conductor losses in accordance with the values of the surface resistivity of lead and niobium measured by Shizume and Vaher [19]. They obtained values at 9 Gc and 4.2°K of 7×10^{-4} ohms per square for lead and 4×10^{-5} ohms per square for niobium. Considering the geometry of the lines and assuming the w^2 dependency of resistance, the ac resistance becomes approximately 50 ohms. Along with 0.5 db loss due to input-output leads, this reduces the loss attributable to the dielectric to 1.7 db. The dielectric loss tangent corresponding to this loss is 7.3×10^{-6}. Since other loss mechanisms such as internal reflections have been neglected, it is reasonable to assume this to be an upper limit, with the actual value falling closer to the 4×10^{-8} or less suggested by theory.

The measurement of delay time, as shown in Fig. 6, showed essentially no dispersion, and permitted the dielectric constant to be found quite accurately. A value of $\epsilon=2.10$ was obtained for FEP Teflon and 2.25 was received for polyethylene (2.30 was found analytically for polyethylene).

TABLE I

INSERTION LOSS
2 μSEC (1360 FT) LINE WITH FEP TEFLON DIELECTRIC AND HEAVY LEAD OUTER CONDUCTOR

Frequency Gc	Insertion loss (peak) db
1	2
1.5	1.4
2	6
2.5	5
3	4
4	4
5	5.5
6	7
7	8.2
8	10.6

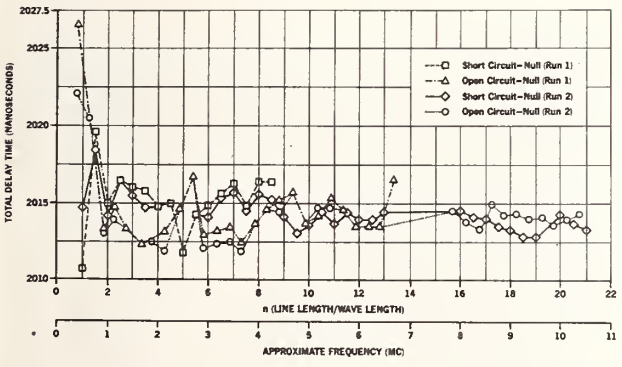


Fig. 6—Delay time by null resonance for a nominal 2.014 μsec line (two runs).

REFERENCES

[1] R. L. Wigington and N. S. Nahman, "The transient analysis of coaxial cables," PROC. IRE, vol. 45, p. 167–174; February, 1957.
[2] A. B. Pippard, "Metallic conduction at high frequencies and low temperatures," "Advances in Electronics and Electron Physics," Vol. VI, p. 1, Academic Press, Inc., New York, N. Y.; 1954.
[3] N. S. Nahman, "A Miniature Superconductive Coaxial Transmission Line," Ph.D. Dissertation, University of Kansas, Lawrence, 1961.
[4] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, "Theory of superconductivity," *Phys. Rev.*, vol. 108, pp. 1175–1204; December, 1957.
[5] D. C. Mattis and J. Bardeen, "The theory of the anomolous skin effect in normal and superconductive metals," vol. 3, p. 412; July, 1958.
[6] J. S. Toll, "Casuality and the dispersion relation; logical foundations," *Phys. Rev.*, vol. 104, pp. 1760–1770; December, 1956.
[7] J. R. MacDonald and M. K. Brachman, "Linear-system integral transform relations," *Rev. Mod. Phys.*, vol. 28, pp. 393–442; October, 1956.
[8] J. Bardeen, "Review of the present status of the theory of superconductivity," *IBM J. Res. and Dev.*, vol. 6, pp. 3–11; January, 1962.
[9] P. W. Anderson and B. T. Matthias, "Superconductivity," *Science*, vol. 144, pp. 373–381; April 24, 1964.
[10] J. Peoples, A. Guimento, and R. Allen, "Study of Dielectric Losses at Cryogenic Temperatures." U. S. Air Force, Rome, N. Y., TN-61-228 Rome Air Dev. Center. pp. 17–26; September, 1961.
[11] H. H. Landolt and K. Börnstein, "Zahlenwerte and Functioner," Springer-Verlag, Berlin, Germany, vol. 1, pt. 3, p. 510 et seq.; 1951.
[12] N. A. Lange, "Handbook of Chemistry," McGraw-Hill Book Co., Inc., New York, N. Y., p. 849; 1961.
[13] P. J. Flory, "Principles of Polymer Chemistry," Cornell University Press, Ithaca, N. Y., p. 49.
[14] Yoh Han Pao and R. F. Bjocklund, "Wave functions and energy eigenvalues for the polytetrafluoroethylene chain," *J. Appl. Phys.*, vol. 31, p. 1925–1934; November, 1960.
[15] S. P. Kabin and G. N. Mikhailov, *Invest. Akad. Nauk*, vol. 55R, p. 325, Fig. 22; 1958.
[16] A. R. Von Hippel, "Dielectric Materials and Applications," The Technology Press of MIT, Cambridge, Mass., p. 174; 1954.
[17] S. Okamoto and K. Takeuchi, *J. Phys. Soc.* (Japan), vol. 14, p. 378; 1959.
[18] J. F. Fowler and F. T. Farmer, "Conductivity induced in polytetrafluorotheylene by X-rays," *Nature*, vol. 174, pp. 136–137; July 17, 1964.
[19] P. K. Shizume and E. O. Vaher, "Superconducting Coaxial Delay Line," presented at IRE General Meeting, New York, N. Y.; March, 1962.

Reprinted from the PROCEEDINGS OF THE IEEE
Vol. 52, No. 10, October, 1964

PRINTED IN THE U.S.A.

Nanosecond Response and Attenuation Characteristics of a Superconductive Coaxial Line*

N. S. NAHMAN†, SENIOR MEMBER, IRE, AND G. M. GOOCH†

Summary—A miniature superconducting coaxial line is described which transmits nanosecond (nsec) pulses without any measurable change in rise time. Data in regard to the temperature dependence of the bandwidth over the range 293°K to 4.2°K are presented. The investigation is confined to ten and hundred foot lengths of cable.

Introduction

THE TRANSMISSION of a nanosecond (10^{-9} second) duration pulse over a relatively long transmission path is conveniently accomplished by a coaxial transmission line; the bandwidth of the line is primarily limited by its inherent skin effect attenuation.[1] Physically realizable coaxial lines are not uniform lines because connectors, dielectric supports, and random geometrical imperfections introduce nonuniformities which in turn produce reflective losses and may also produce higher order mode transmission of microwave frequencies. Skin effect losses for a given characteristic impedance may be reduced to any desired nonzero magnitude by increasing the respective dimensions of the coaxial line; however, mode conversion effects, if they are present, will occur at much lower frequencies.[1] Nanosecond systems by their very nature are inherently small physically and are not too compatible with large diameter coaxial lines. In nsec pulse oscillography it is usual practice to employ relatively large semirigid 50-ohm coaxial lines ($\frac{1}{2}$ inch to $\frac{7}{8}$ inch outer diameter) to provide signal delays of the order of 100 nsec in the vertical signal channel. Such lines possess an attenuation which is less than 10 db/100 feet at 4000 Mc.

This paper describes a miniature coaxial line (50 ohms, 0.091 inch diameter) that possesses a very low attenuation by virtue of the superconductivity phenomenon which occurs at low temperatures. The miniature line was constructed for use as a relatively lossless signal delay line for nsec pulse oscillography.

* Received by the IRE, May 18, 1960; revised manuscript received, August 29, 1960.

† Electronics Res. Lab., Center for Res. in Engrg. Science, Univ. of Kansas, Lawrence, Kans.

[1] S. Ramo and J. R. Whinnery, "Fields and Waves in Modern Radio," John Wiley and Sons, Inc., New York, N. Y.; May, 1956.

The Miniature Line

The miniature line was fabricated in a total length of 200 feet; its construction is as follows:

1) Inner conductor, niobium, 0.010 inch diameter;
2) Outer conductor, lead, 0.034 inch inner diameter, 0.091 inch outer diameter;
3) Dielectric, teflon.

Electrically, the line has a characteristic impedance of 50 ohms and a capacity of 29.1 picofarads per foot. Mode conversion occurs for frequencies on the order of 100 kMc. The line superconducts for temperatures less than 7.22°K, the critical temperature for Pb; the critical temperature for Nb is 8.7°K. The measured conductivities of the conductors at 293°K. (room temperature) are 6.14×10^6 mhos/meter for the Nb center conductor and 4.82×10^6 mhos/meter for the Pb outer conductor. Fig. 1 shows a photograph of the line and the associated transition to a type-*N* coaxial connector (note the boot match shown for size comparison).

The high-frequency attenuation of the line at 293°K was predicted by calculating the skin effect and dielectric losses. For liquid nitrogen temperatures (77.3°K) the dielectric losses were neglected, and the skin effect loss was based upon 77.3°K conductivities which were determined by direct measurement in a liquid N bath. No predictions were made as to the low temperature (4.2°K) attenuation of the line due to the fact that the superconductive losses would be very small and would not present any measurable value in the contemplated instrumentation system. This may be seen by referring to Fig. 2, an approximation to Gruneisen's curve.[2] The low temperature resistivity varies approximately as T^5, while for high temperatures it varies approximately as T. For the specific metal samples used in the line, Gruneisen's curve in conjunction with the two resistivity measurements (in the linear region of the curve) yields the resistivity constants[3–5]

$$\text{Nb: } \theta = 26.5; \qquad R_\theta = 1.3\times10^{-8}\ \text{ohm meters}$$

$$\text{Pb: } \theta = 70°\text{K}; \qquad R_\theta = 4.49\times10^{-8}\ \text{ohm meters.}$$

When the temperature is lowered, the normal resistivity (not the superconductive resistivity) approaches the residual resistivity (resistivity at 0°K). Fig. 2 shows that the normal conductivity at low temperatures be-

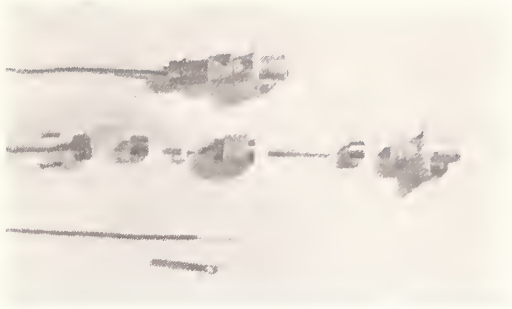


Fig. 1—The miniature line and its associated transition to a type-*N* connector.

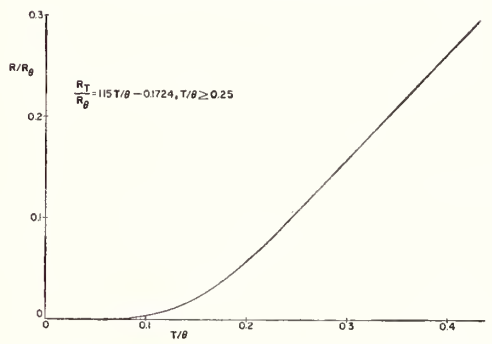


Fig. 2—Gruneisen's curve, $\theta \equiv$ Debye temperature, $R_\theta \equiv$ Debye temperature resistivity.

comes very large. If at low temperatures the conductivity is increased by a factor of 10^4, the line attenuation at any RF frequency would be reduced by a factor of 10^2; this in itself would be a remarkable reduction in attenuation which could not be fully detected by the contemplated instrumentation system. When superconductivity occurs, the conductivity (Pb) may increase by a factor of 10^{17} over the 293°K conductivity.[3]

The Experimental System

The experimental system is shown in block diagram form in Fig. 3(a) and 3(b). The insertion switch, which may be seen in Fig. 4, provides a convenient method for rapid insertion switching in the rigid coaxial system. Rigid coaxial lines are used to minimize the loss and instability in the over-all transmission system. The pulse transmission measurements were accomplished on a Project JAYHAWK traveling-wave synchroscope that employed a DuMont type KR 1524 P11M TWCRT.[6] The deflection response of the indicator is 6 db down at 2.1 kMc. The 100 feet of miniature line when wound on a 2 inch diameter form constitutes a cylindrical package of $3\frac{1}{4}$ inches in diameter and 5 inches long. The low-temperature system consists of two concentric dewars, the He flask having an internal radius of $2\frac{7}{8}$ inches and a depth of $15\frac{1}{4}$ inches; the N flask has an internal radius

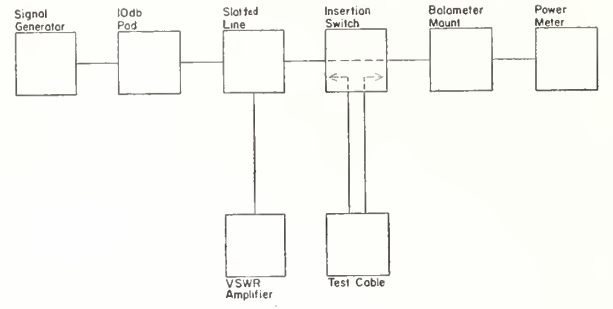
[2] C. Kittel, "Introduction to Solid State Physics," John Wiley and Sons, Inc., New York, N. Y.; August, 1955.

[3] The Pb values are of the same order as the usual values cited in the literature. On the other hand, the Nb Debye temperature is much smaller than the usual values (26.5°K compared to 252°K). This is not unusual; even the exact Gruneisen curve does not always agree with exeriment when the metal concerned is one of the transition metals. Fig. 2 is not intended to be quantitative, only qualitative; it is included at this point mainly to demonstrate qualitatively the variation of conductivity with temperature.

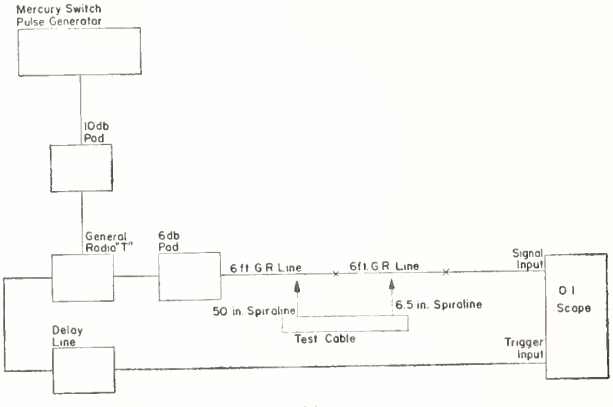
[4] "American Institute of Physics Handbook," McGraw-Hill Book Co., Inc., New York, N. Y.; 1957. Also, R. B. Scott, "Cryogenic Engineering," D. Van Nostrand Co., Inc., New York, N. Y.; 1959.

[5] N. F. Mott and H. Jones, "The Theory of the Properties of Metal and Alloys," Clarendon Press, Oxford, Eng.; 1936. (Reprinted by Dover Publications, Inc., New York, N. Y.; 1958.)

[6] Project JAYHAWK Progress Report, Contract No. DA 49-170-sc-2618, Univ. of Kansas Res. Foundation, Lawrence; April 12, 1959 to July 11, 1959.



(a)



(b)

Fig. 3—(a) Attenuation instrumentation. (b) Pulse response instrumentation.

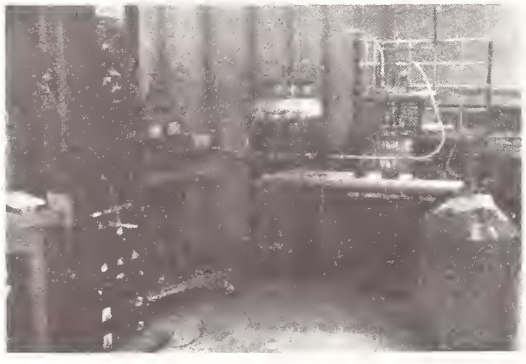


Fig. 4—Laboratory arrangement for attenuation measurements.

of $4\frac{1}{2}$ inches and a depth of $18\frac{1}{2}$ inches. The low-temperature operation of the system was at 4.2°K and He pumping was not required.

Experimental Results

The pulse response measurements employed an input pulse having the following characteristics:

1) 50 per cent pulse width 1.7 nsec.
2) 10 per cent–90 per cent rise time 0.4 nsec.
3) Amplitude 16.7 volts.

The pulse measurements were run on a 10 foot and a 100 foot length of line in order to demonstrate clearly the behavior of the line. The shorter length was necessary to obtain measurable responses when the line was operating at temperatures equal to or greater than 77.3°K [Fig. 5(a) and 5(b)]. Fig. 5(c) through 5(e) presents a graphic picture of the increase in bandwidth as the temperature moves from 293°K to 77.3°K to 4.2°K. Fig. 5(f) contains a superposition of the input pulse and the output pulse from the 100-foot line at 4.2°K. Notice that the rise time of the pulse is unaltered after passing through 100 feet of line, and that the upper portion of the output pulse has been reduced initially in amplitude by a reflection. The attenuation measurements demonstrate the presence of reflections.

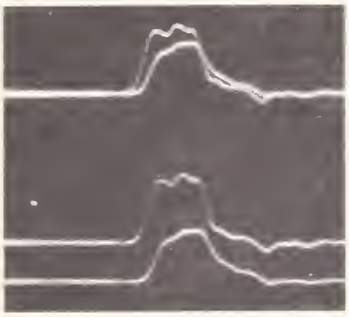
The attenuation measurements were also concerned with two lengths of line, 10 and 100 feet long. For temperatures equal to or greater than 77.3°K, the 10 foot length was employed; the results were multiplied by 10 and plotted in db/100 feet. In Fig. 6, *A* and *B* are the 293°K calculated and measured attenuation, respectively, and *C* is included the show the 293°K attenuation of an all-copper coaxial line having the identical dimensions of the miniature line. The attenuation of the copper line is less than the miniature line at 293°K, but is useless at 293°K for nsec pulse transmission. In Fig. 7, *A* and *B* show the measured and predicted 77.3°K attenuation.

In Fig. 8, *B* is the 4.2°K measured attenuation, while *A* is also included for comparison. This curve is the attenuation curve at room temperature of 100 feet of $\frac{1}{2}$-inch Spiroline rigid coaxial line. The 4.2°K attenuation curve has been adjusted in that the known rigid line losses and the input reflection loss have been subtracted from the over-all attenuation. Experimental error is in evidence which accounts for the negative attenuation. The variations in the 4.2°K attenuation are due to the multiple reflections that are present in the over-all system; however, even with the reflective losses the line outperforms the large 293°K Spiroline coaxial line. The reflection losses of the over-all system including all transmission paths are clearly placed in evidence when the miniature line is in the superconductive state (4.2°K).

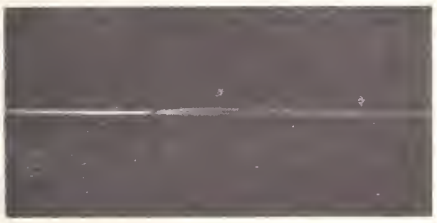
When the majority of the attenuation disappears in the transmission line system, all of the nonuniformities of the system become evident and are exhibited in the undamped reflections produced by the system. Some of the discontinuities of the various components in the given system can be measured separately, and their effect on the over-all transmission may be calculated.

Regarding the discontinuities of the superconductive line and its associated connectors, it is found that the very low operating temperature causes geometrical changes (contractions) in the line itself and significantly disturbs the relative positions of the line conductors in the connector constant impedance matching transitions. This, in turn, produces a relatively large frequency-dependent VSWR that is not experimentally repeatable due to the present design of the transitions.

(a)



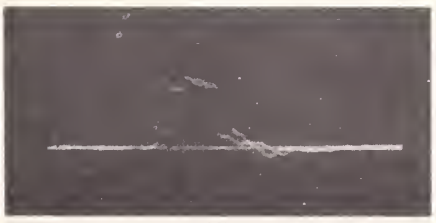
(b)



(c)

(d)

(e)



(f)

Fig. 5—(a) Upper trace: input pulse. Lower trace: output pulse, 10 feet, 293°K. (b) Upper trace: input pulse. Lower trace: output pulse, 10 feet, 77.3°K. (c) Output pulse, 100 feet, 293°K. (d) Output pulse, 100 feet, 77.3°K. (e) Output pulse, 100 feet, 4.2°K. (f) Superposition of input and output pulses, 100 feet, 4.2°K; input pulse shifted to left.

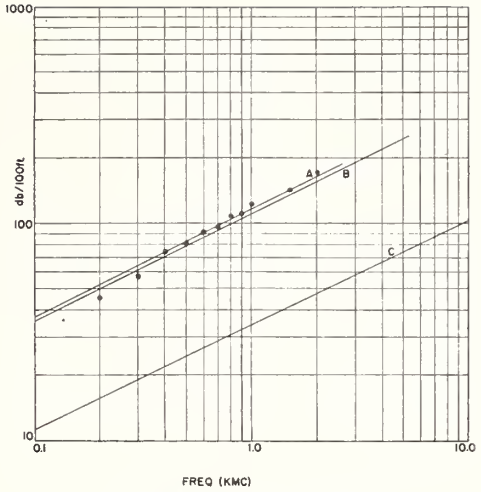


Fig. 6—Attenuation at 293°K.

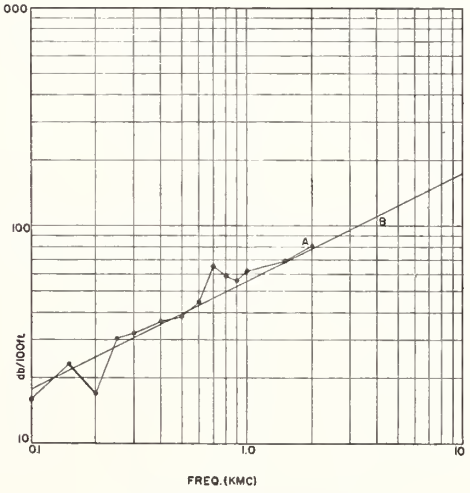


Fig. 7—Attenuation at 77.3°K.

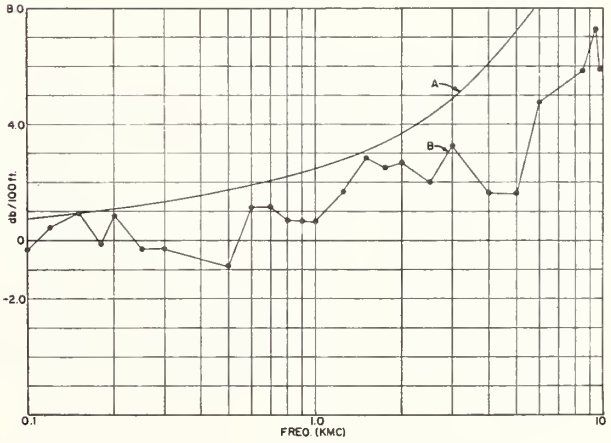


Fig. 8—$A = 100$ feet $\frac{1}{2}$-inch Spiroline cable (manufacturer's data) 293°K; $B = 100$ feet miniature line, 4.2°K.

Additional Remarks

The 4.2°K attenuation measurements in Fig. 8 are not to be interpreted as being indicative of the high-frequency attenuation to be found in superconductors. The majority of the resistive attenuation is due to the two short lengths of the miniature line that are not superconducting along their entire lengths. These two lengths are the input and output portions of the line that are not submerged in liquid He and are attached to the connectors shown in Fig. 1. The connectors are panel mounted in the base of a liquid N bucket; hence, the miniature line's extremities are held at 77.3°K.

In addition to nonuniform thermal contractions, the line attenuation is dependent upon the electrical characteristics of the particular Nb, Pb and teflon samples employed in the construction of the line. The fabrication techniques of the individual materials and the line itself introduce into the respective materials various stresses and strains which, in turn, alter their electrical characteristics. The Pb outer conductor was fabricated about the teflon coated Nb center conductor by a pressure-welding technique. The Pb was chemically pure, and subsequent measurements on the completed line showed that the conductivity was 4.82×10^6 mhos/meter at 293°K and 20.17×10^6 mhos/meter at 77.3°K. Using these values and tabulated values of Gruneisen's function[5] the Debye temperature was found to be 106.7°K as compared to the published value 96.3°K, normalization being referred to 293°K. The Nb Debye temperature was found to be 56.5°K using the conductivities 6.14×10^6 mhos/meter (293°K) and 24.1×10^6 mhos/meter (77.3°K), normalized at 293°K.

The teflon dielectric was extruded upon the Nb center conductor. The Nb wire is not pure and contains 0.25 per cent impurities.[7] The 10-mil Nb wire was formed by sintering a metallic powder into an ingot and drawing to size. Slade and Smallman[8] have found that 3-mil samples of the same wire even after vacuum annealing are not exceptional superconductors. The maximum supercurrent that could be carried by the 3-mil wire before the Silsbee destruction of the supercurrent occurred was approximately 50 per cent of that which the Silsbee Hypothesis predicts.[9]

Future studies on the miniature line will be concerned with the Silsbee effect and the measurement of the conductivities from 293°K down to the superconduction critical temperatures. Also, the *X*-band dielectric properties of teflon at 4.2°K are to be investigated. The results of these studies in conjunction with a study of the high-frequency behavior of superconductors is expected to provide a basis for calculating the bandwidth of the line.

Conclusion

This investigation has demonstrated that a miniature superconducting line may be employed as a relatively lossless wideband coaxial line that is able to transmit nanosecond pulses with no apparent distortion. Furthermore, the experiments performed within this investigation in conjunction with the present-day knowledge of low temperature normal conductivity indicates the feasibility for designing a miniature low-temperature normal conducting line which would possess an attenuation several orders less than 293°K attenuation. One point that should be kept in mind is that if the losses vanish in a transmission line, the line geometry must be perfectly uniform if reflections within the line itself are to be avoided. Under lossless conditions, such internal reflections would not be damped by resistive loss, and the resulting multiple reflections could affect the signal-to-noise ratio of the transmission line; hence, under some circumstances in a given line, a combination of normal and superconductors would be appropriate, *e.g.*, inner conductor—super and outer conductor—normal.

Acknowledgment

The authors would like to acknowledge their discussions with R. L. Wigington and the ardent help of their graduate research assistant, A. J. Cummings. Also, the authors are indebted to their sponsor who supported their work within DA 49-170-sc-2618 and DA 18-119-sc-899.

[7] Per cents: Fe—0.02, Ni—0.01, W—0.01, Ta—0.05, Ti—0.02, Zr—0.02, O—0.05, N—0.05, C—0.02. Manufactured by Fanstell Metallurgical Corp., Chicago, Ill.

[8] A. E. Slade and C. R. Smallman, Arthur D. Little Co., Cambridge, Mass.; private communication with C. R. Smallman.

[9] D. Shoenberg, "Superconductivity," Cambridge University Press, Cambridge, Eng., 2nd ed.; 1952.

Reprinted from the PROCEEDINGS OF THE IRE

Vol. 48, No. 11, November 1960

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 70C, No. 1, January–March 1966

Measurement Standards for Low and Medium Peak Pulse Voltages

A. R. Ondrejka* and P. A. Hudson*

National Bureau of Standards, Boulder, Colo.

(November 10, 1965)

Two systems for the measurement and standardization of peak pulse voltage have recently been developed in the Radio Standards Engineering Division of the National Bureau of Standards. Designed for use with pulses having durations as short as 10 nanoseconds, the systems represent the initial effort by NBS for establishment of standards in this field and it is expected that they will serve as the basis for a calibration service.

One of the systems is a refined, slideback voltmeter employing a solid state diode. The rise time is less than 10 nanoseconds and the instrument is useful for measurements in the range 5V to 100V as presently constructed.

The second system standardizes the output peak voltage level of pulse generators by means of cascaded zener diodes. The clipping level of the diodes is determined first by d-c voltage measurements and the system is subsequently used to generate pulses of known peak amplitudes. The rise time of the diodes is approximately 30 nanoseconds.

The accuracy of both systems is within 0.25 percent, and comparison measurements have been made between the two, and with other NBS a-c and rf (CW) voltage standards.

Key Words: Measurement, peak, pulse, pulse-limiting, slideback, standard, voltage.

1. Introduction

The use of pulse waveforms has increased significantly during recent years, resulting in a wider interest in measurement of pulse parameters. Pulses are used in a variety of applications such as high speed electronic computers, medical applications, pulse amplitude modulation of carrier waves, and for testing of broadband circuits and components. The time domain reflectometer is a good example of the latter. In addition, progress has been made in the state of the art of generation of fast-rise, short-duration pulses. Subnanosecond rise times are now common and work is progressing on picosecond rise-time pulsers.

The parameters of interest in measurements on pulses include the rise time and the peak voltage amplitude. Various types of instruments for performing these measurements are commercially available, the most common of which is the cathode ray oscilloscope. Other types include direct reading peak voltmeters which cover a wide dynamic range from a few millivolts to several kilovolts at pulse widths down to a fraction of a microsecond. The specified uncertainties of the instruments used for peak voltage measurement are generally 1 to 2 percent.

As a result of the wider application of pulse waveforms and the overall higher accuracies required in present-day measurements, the National Bureau of Standards has received requests for calibration services in the pulse parameters. It is the purpose of this paper to describe two systems for measurement and standardization of peak pulse voltage in the range 5V to 100V for pulses having durations of at least 10 nsec. These systems represent the initial efforts at NBS for establishing standards in this field and it is expected that they will serve as the basis for a calibration service.

Two independent systems were developed to allow intercomparison and thus better evaluate the absolute accuracies. One system is designed to measure the pulse voltage across its input terminals while the other generates a predetermined, accurately known, pulse voltage level. Intercomparison can then be carried out without the need for a transfer instrument.

The development of standards for pulse rise time measurement is planned for the future.

2. Principles of Operation and Design

The voltage waveform of interest here is a train of so-called "d-c" or unidirectional pulses [1].[1] That is, ideally (in the mathematical sense), the voltage excursion occurs in only one direction from the zero position. Usually the pulse width is much less than the time interval between pulses and the average voltage is small compared to the peak voltage. One of the problems is to obtain an accurate, continuous reading of a waveform whose instantaneous value is zero the greater part of the time.

*Radio Standards Engineering Division, NBS Boulder Laboratories, Boulder, Colo.

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

2.1. The Standard Measurement System

In the design of a system, for peak pulse voltage measurement, there are certain features which are necessary or desirable. Of primary importance is the system accuracy which requires that the design eliminate as many sources of error as possible even though some sacrifice in ease of operation may have to be made. Other desirable features include wide dynamic range, insensitivity to pulse duty cycle, and a continuous output reading. Since it is desired to use the system for measurements on fast-rise, short-duration pulses, it must also have an extremely broad bandwidth.

After theoretical and experimental investigation of various known measurement techniques, it was decided that a system based on the slideback principle would adequately fulfill the requirements mentioned above. Slideback voltmeters have been in use for a number of years and have been described in the literature [2, 3]. The principle involves comparing the unknown voltage to an accurately known, adjustable voltage. The circuit is arranged so that the two voltages oppose each other and null detection methods are used to indicate when the voltages are equal. For the measurement of d-c voltage, high accuracy and precision are attainable. In applying the principle to pulse voltage measurement a diode is employed to limit current drain from the d-c slideback source during the time interval between pulses. This time interval may be several orders of magnitude (e.g., 10^4) greater than the pulse width. The version used here can be described by referring to the circuit in figure 1, which is included for the purpose of illustrating the principle. In the absence of an input signal and with the d-c slideback voltage set to zero, no current flows in the circuit as evidenced by a null indication on the galvanometer. When a train of pulses (positive pulses for the circuit shown) is applied to the input, the diode becomes conducting and may be returned to its original nonconducting state by adjusting the reverse d-c (slideback) voltage until the diode is cut off and the galvanometer is again nulled. The reverse d-c bias voltage, which can be easily and accurately measured, is then equated to the peak pulse voltage.

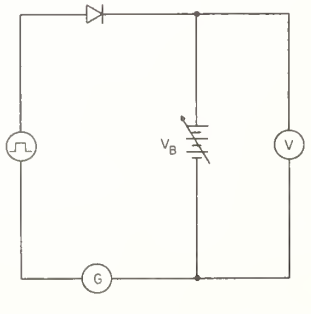


FIGURE 1. *Basic circuit of slideback principle.*

The accuracy of the system is primarily dependent on the validity of the assumption that, at a null condition, the d-c bias voltage is equal to the peak pulse voltage. The validity of this assumption is, in turn, a function of the characteristics of the diode. In an ideal diode both the back resistance and the forward conductance would be infinite while the rise time and shunt capacitance would be zero. The problem then was to choose an available diode having as nearly as possible these ideal characteristics. One of the critical characteristics is the forward conductance at low values of applied voltage. The forward conductance is important because of the finite sensitivity of the null detector. If some voltage, V_e, across the diode is required to cause a detectable current to flow, then the d-c slideback voltage, V_B, would be less than the peak pulse voltage, V_p, by an amount V_e which we have called the offset voltage. It is possible to measure the V_e of a particular diode and apply a correction. Some uncertainty would, however, be involved in this correction and, in general, it is desired to keep the correction as small as possible.

The forward diode current, I, is given approximately by the equation [4]

$$I = -I_s(\epsilon^{qv/kT} - 1) \tag{1}$$

where

I_s = reverse saturation current
q = electronic charge
v = applied voltage
k = Boltzmann's constant
T = Temperature, °K.

Thus the forward current is a function of the reverse current, I_s. This suggests the choice of certain types of point-contact diodes which have relatively large reverse currents. In the point-contact diodes tested, the reverse current did not reach a saturation value but, rather, tended to increase with increasing reverse voltage. The resultant heating effect caused changes in the forward characteristics. Thus the point-contact types were not suitable for this application. Junction diodes, on the other hand, are usually characterized by extremely low reverse currents and would not be suited because of the resulting large forward offset voltage. A diode type which is intermediate between these two is the metal-semiconductor type. In one version of this type, a gold wire is welded to n-type germanium to form a rectifying contact as in the 1N270 diode. The reverse resistance of the 1N270 is approximately 10 MΩ as compared to 1000 MΩ for typical junction types. A comparison of V–I curves for the 1N270 and a 1N4445 junction diode is shown in figure 2. The conductance of the 1N270 at the voltage levels shown is approximately 600 times greater than that of the junction type. Since conduction in metal-semiconductor diodes is by majority carriers, they have a better high frequency response (fast rise time) and in addition the shunt capacity is comparatively small. The rise time of the 1N270 is approximately 10 nsec and at present it appears to have the best overall

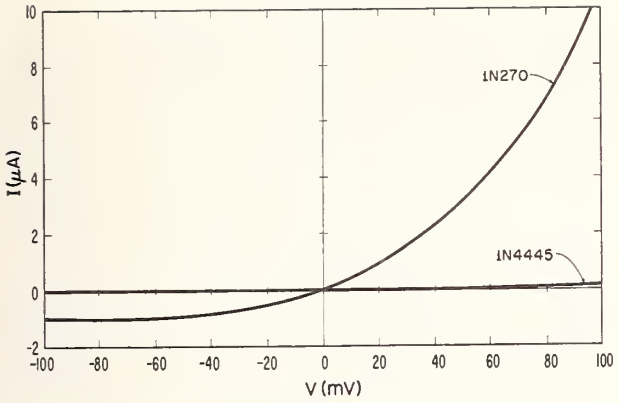


FIGURE 2. *Comparison of V–I characteristics of 1N270 gold bonded diode and 1N4445 junction diode.*

characteristics of the several diodes tested for this application.

Another important factor in the accuracy of the system is the means used for detecting when the diode is just at cutoff. The galvanometer detector mentioned earlier is not suitable because it responds to average current and thus suffers a considerable loss in sensitivity over that obtainable with a peak-responding detector. In addition, nulling the galvanometer would not guarantee equality between the pulse and d-c slideback voltages because of the reverse leakage current through the diode during the time interval between pulses. This current tends to reach a saturation value, I_s, of the order of 1μA (see fig. 2). The average reverse current, $\overline{I_R}$, then is given by

$$\overline{I_R} = I_s(1 - ft_p) \tag{2}$$

where f is the pulse repetition rate and t_p is the pulse duration. Typical pulse duty cycles, ft_p, range from 10^{-4} to 10^{-1} and thus $\overline{I_R}$ is equal to at least 0.9 I_s. The average forward current through the diode is given by

$$\overline{I_F} = I_p f t_p \tag{3}$$

where I_p is the peak current and a flat-top pulse is assumed. At null, $\overline{I_F} = \overline{I_R}$, and

$$I_p = \frac{I_s(1 - ft_p)}{ft_p} . \tag{4}$$

When the pulse duty cycle is 10^{-1}, for example, $I_p = 9\ I_s$, or approximately 9 μA. Referring again to figure 2, this current corresponds to 80 mV forward voltage and the peak pulse voltage must be larger than the d-c bias by this amount in order to null the galvanometer. At lower duty cycles, the pulse voltage would need to exceed the d-c bias by even greater amounts.

Due to the fact that the pulse duty cycle for narrow pulses is difficult to measure accurately and the reverse saturation current, I_s, varies with temperature, it is not practical to apply a correction for this effect.

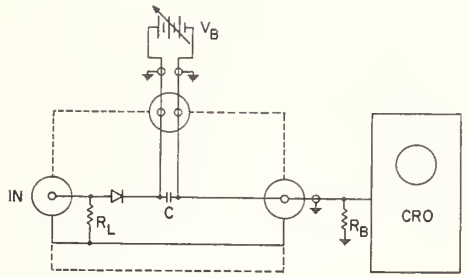


FIGURE 3. *Circuit diagram of slideback voltmeter system.*

A better detector was devised by substituting a coaxial resistor and a high gain cathode ray oscilloscope for the galvanometer. The circuit diagram of the complete measuring system is shown in figure 3. This is a peak detecting system and does not depend on average values of d-c current as previously. The CRO is set for capacitive coupling across the resistor, R_B. Pulses of current flowing through R_B are detected by the CRO and the d-c bias voltage is adjusted so that the top of the pulse just recedes into the baseline. The baseline is not affected appreciably by the reverse current[2] except for a very small differentiated pulse at the leading and trailing edges of the input pulses which is caused by the shunt capacitance of the diode.

Since the CRO detector has limited sensitivity, there will be some small forward pulse current, I_e, through the diode which is not detectable. As mentioned earlier, the voltage, V_e, which causes this current to flow is the difference (offset) between the peak pulse voltage and the d-c bias voltage. That is, $V_e = V_{\text{pulse}} - V_{\text{d-c}}$. Thus, when the d-c bias voltage has been adjusted to cause the peak of the pulse to just recede into the baseline, the diode will appear to be cutoff when in reality a small current continues to flow. It is possible to estimate the magnitude of the offset voltage by use of the *V-I* curve of figure 2 and a knowledge of the detector sensitivity. The detector sensitivity is 5 mV/cm and resolution is approximately 0.2 mm on the oscilloscope CRT which corresponds to a minimum detectable voltage of 0.1 mV. This voltage is the IR drop across the 50 Ω coaxial resistor, R_B, which terminates the coaxial line to the oscilloscope input and thus the minimum current is 2 μA. Referring to figure 2, it can be seen that at this current the offset voltage is 40 mV. This voltage is added to the measured d-c bias voltage as a correction. Due to thermal effects the offset voltage may vary $\pm$ 10 mV and this figure is taken as an uncertainty and it is independent of the input voltage level.

[2] It should be noted that when the bias voltage is equal to the input pulse voltage the reverse current goes to zero during the time the pulse is impressed across the diode. This interruption of the reverse current results in the generation of a small positive pulse. Since the reverse current is approximately 1 μA and the resistor is 50 Ω, the pulse voltage is 50 μV. The top of this pulse is the true baseline; however, due to the limited sensitivity of the CRO it appears only as a widening of the trace.

The detector sensitivity limits the ability of the operator to determine when the diode is just at cutoff and this is another factor which affects the accuracy of the measuring system. It was found experimentally that the random error caused by this factor alone resulted in variations in the reading of the peak pulse voltage as large as ± 20 mV from the mean. In order to make a quantitative determination of this error, the standard deviation of the mean was calculated for a sample of 30 measurements made with the input voltage held constant. This value was calculated to be ± 10 mV and was taken as an additional error which, again, is independent of the input pulse voltage level.

The other source of error is in the measurement of the d-c bias voltage. This measurement is made with a digital voltmeter which is accurate to within 0.05 percent of the indicated voltage. The overall system accuracy is estimated to be within 0.05 percent ± 20 mV.

To complete the description of the measuring system, reference is made to figure 3. That portion of the circuit enclosed by the dashed lines is mounted in a small, round, metal container 1½ in. in diameter and 2½ in. long. The load resistor, R_L, consists of three film resistors connected in parallel. These resistors are easily removed so that R_L may be adjusted to match any generator impedance from 50 to 10,000 Ω. The parallel arrangement reduces the inductive reactance as compared to a single resistor and also increases the power dissipation capability. A second mount employing a 50 Ω coaxial resistor is used for measurements on pulses having the fastest rise times.

The slideback voltage is obtained from a battery pack containing ten 9 V batteries and one 12 V battery. Any or all of the batteries may be switched into the circuit and a 10-turn potentiometer across the 12 V battery serves as a vernier. The capacitor, C, connected across the battery serves as a path for the pulse current. Batteries cause much less distortion of the pulse than the electronically regulated supplies which were tried. The reason for the distortion by the regulated supplies has not been investigated extensively, but it is believed to be due to the filter capacitors and the frequency dependence of the internal impedance. Connection between the diode mount and the bias supply is made by means of a shielded cable. A 50 Ω cable is used to connect to the oscilloscope detector and this cable is terminated by the 50 Ω coaxial load, R_B, which is connected in parallel with the CRO input. The bandwidth of the CRO vertical amplifier is 50 MHz.

Since R_L may be as large as 10,000 Ω, the slide back system may also be used as a high impedance voltmeter. That is, it can be connected in parallel with the output of a pulse generator to measure and monitor the voltage level. The loading effect due to R_L is, in most cases, negligible.

2.2. The Standard Generating System

In order to provide a cross-check on the measuring system described above, a second independent system was developed. It was desired that the second system standardize the output level of a pulse generator to an accurately predictable value. In addition it was desirable that the output level be constant over relatively long periods of time. Intercomparison of this system and the first could then be carried out without the need for a transfer standard.

The most promising method for achieving these objectives appeared to be a circuit employing zener (reference) diodes to clip or limit the pulse generator output. The voltage level at which limiting occurred could be determined using accurately measured d-c voltage. It would then be assumed that pulse voltage limiting would occur at the same level. In order to establish the validity of this assumption, an investigation of the characteristics of zener diodes and their associated circuits was required.

The regulating action of zener diodes is now fairly well understood. This action occurs in reverse biased junction diodes when the reverse voltage, V, exceeds a certain critical value, V_c. The diode current increases extremely rapidly with $V > V_c$ causing the voltage drop across the junction to maintain a nearly constant value as shown in figure 4 [5]. The rapid increase in current is caused by an avalanche effect in the semiconductor material and is due mainly to majority (i.e., electron) carriers. Thus, the action can be fast and is highly desirable for pulse voltage limiting to avoid undue increase in the rise time of the output pulse over that of the input pulse.

A basic shunt zener regulating circuit is shown in figure 5a, with the a-c equivalent circuit of the zener junction shown in figure 5b. The junction capacitance, C_z, is of the order of a few hundred picofarads which results in the storage of charge in the junction.

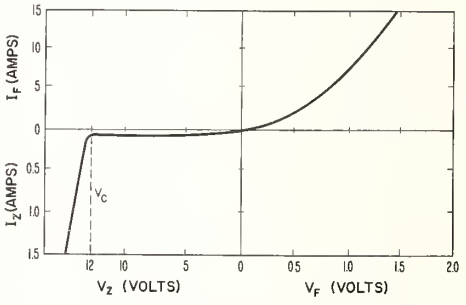


FIGURE 4. V–I *Characteristics of zener diode.*

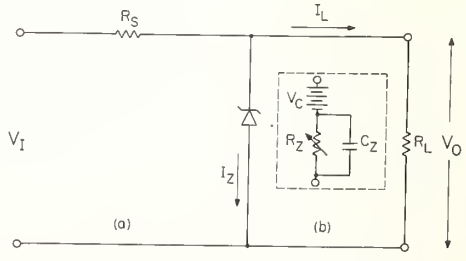


FIGURE 5 (a) *Basic zener regulating circuit;* (b) *A–C equivalent circuit of zener junction.*

This effect tends to limit the high frequency (and rise time) response. In designing the circuit for use as a d-c calibrated pulse voltage limiter, standard equations from handbooks were employed. The basic equation relating changes in output voltage, V_0, to changes in input voltage, V_i, load resistance, R_L, and temperature, T, is given by

$$dV_0 = \left(\frac{\delta V_0}{\delta V_i}\right) dV_i + \left(\frac{\delta V_0}{\delta R_L}\right) dR_L + \left(\frac{\delta V_0}{\delta T}\right) dT. \qquad (5)$$

In our application the second term in the above equation is not important because R_L is constant and $dR_L = 0$. The temperature effect on V_0 is, however, very important and required the use of diodes having low temperature coefficients. The importance of temperature coefficients is due to the fact that when calibrating V_0 using d-c voltage, the power dissipated in the junction may be several orders of magnitude greater than when the circuit is subsequently used with pulses. Temperature coefficients for the 10 and 50 W diodes used here are less than 0.1 percent/°C. It was desired to limit $\left(\frac{\delta V_0}{\delta T}\right) dT$ to 0.1 percent and hence the maximum temperature rise for a 50 W diode is 1 °C which corresponds to 1 W power dissipation in the junction. (The thermal resistance of these diodes is 1 °C/W.)

During the d-c calibration of V_0, the magnitude of V_i was adjusted so that the zener current I_z was large enough to insure good regulation (low zener impedance) and at the same time not exceed 1 W of power in the junction. In subsequently using the calibrated circuit on a pulse generator, it was assumed that $V_{i(\text{pulse})}$ could be set no closer than ± 10 percent of $V_{i(d\text{-}c)}$. It was required that the change in V_0 due to the uncertainty in $V_{i(\text{pulse})}$ be limited to 0.1 percent. Thus the value of $\frac{\delta V_0}{\delta V_i}$ in eq (5) could be no greater than 0.01 Using this information it was possible to calculate the value of the series resistor, R_s, for each diode by use of the equation

$$\frac{\delta V_0}{\delta V_i} = \frac{1}{1 + R_s\left(\frac{1}{R_Z} + \frac{1}{R_L}\right)}. \qquad (6)$$

In a typical example for a 12 V, 10 W zener and a 50 Ω load, the value of R_s is 400 Ω. The voltage drop across R_s is approximately 120 V and it must absorb 25 W of power. The same degree of regulation can be achieved with less power dissipation and voltage drop in R_s by connecting a second higher voltage zener circuit ahead of the first as shown in figure 6. The regulation required for each circuit is now only 0.1 (10 percent), and the values of R_{s_1} and R_{s_2} can be reduced to approximately 40 Ω each. The rise time of the double diode circuit is 50 nsec as compared to 30 nsec for the single diode.

Zener diodes are manufactured with breakdown voltages ranging from a few volts to well over 100 V.

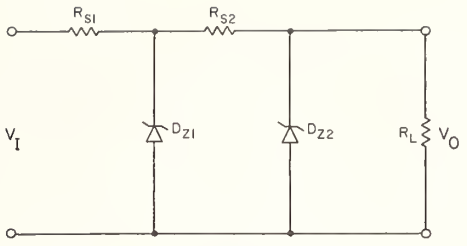


FIGURE 6. *Double zener circuit.*

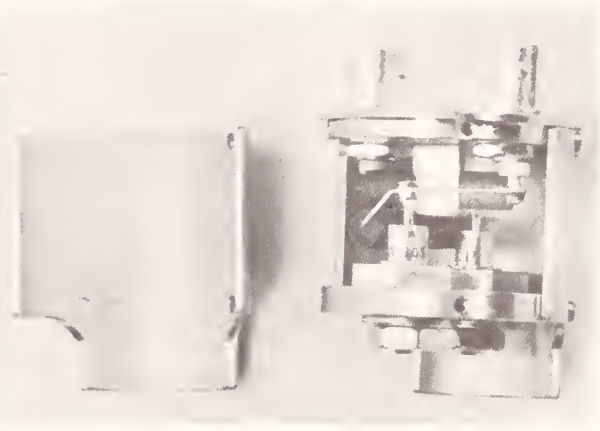


FIGURE 7. *Mount for zener diodes and resistors used in the standard generating system.*

Several circuits for various voltages in this range have been built and evaluated. In general, at the higher voltage levels it was found that the rise time could be improved by stacking several low voltage diodes in series rather than using a single high voltage unit.

Each zener circuit is mounted in a 2-in. diam by 2-in. high brass case as shown in figure 7. The performance of the circuits was improved by immersing the circuit and case in liquid nitrogen. Nitrogen is allowed to fill the case through holes in the top and bottom. This served to insure that the 0.1 percent temperature effect was not exceeded and allowed the series resistor, R_s, to be smaller in physical size.

3. Performance Characteristics and Results of Intercomparison

As mentioned earlier the estimated uncertainty of the standard measuring system using the 1N270 diode is no greater than 0.05 percent ±20 mV. Due to the fixed 20 mV uncertainty and a maximum allowable error of 0.5 percent, the lower limit on the useful voltage range is 5 V. The upper limit of 100 V is fixed by the maximum allowed reverse voltage of the diode. The rise time of the circuit is 10 nsec and the minimum width of the peak of the pulse is 10 nsec.

Tests were made on the standard measurement system to determine the effect of pulse duty cycle on the accuracy. Within the limits of precision, the measuring system was found to be insensitive to duty cycle in the range 10^{-5} to 5×10^{-1}. These tests were made by varying separately both the pulse duration and the repetition rate.

In the zener circuits, as in the measuring system, the effects of varying the duty cycle were negligible. The errors in the system include the 0.1 percent due to temperature effects, the 0.1 percent due to the uncertainty in setting $V_{i(\text{pulse})}$ equal to $V_{i(\text{d-c})}$ and the 0.05 percent error in the d-c voltmeter used to measure the limiting voltage, V_0. Thus, the overall estimated accuracy is 0.25 percent.

The zener circuits may also be used to simply regulate the output of a pulse generator. In this application variations in the output amplitude of the pulse generator are reduced by a factor of 100 to 1 at the output of the zener circuit. Higher regulation can be obtained by increasing the value of the source resistor, R_s.

Comparison measurements have been made between the two systems and with the NBS a-c and CW rf voltage standards. At 100 Hz the disagreement was less than 0.1 percent, at 10 MHz and 30 MHz less than 0.3 percent, and at 100 MHz less than 2.5 percent. The uncertainty of the 100 MHz rf CW thermal voltmeter used in the comparison was approximately 1 percent.

Agreement between the standard measurement system and the standard generating system was within 0.3 percent at all voltage levels where comparison was made. These voltage levels were 6 V, 12 V, 24 V, 47 V, and 100 V.

A photograph of the console containing both standard systems is shown in figure 8.

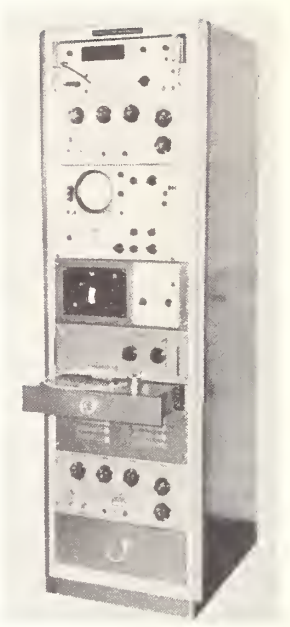


FIGURE 8. *Accessory equipment for both pulse voltage standards.* Included are the digital voltmeter, pulse generator, oscilloscope, and adjustable d-c bias supply.

(Paper 70C1-216)

4. Conclusion

The accuracies of the two systems for measurement and standardization of peak pulse voltage have been reasonably well established to be within 0.25 percent. These systems are considered an adequate basis for establishment of a calibration service for presently available commercial instruments.

Further work is planned to improve the performance characteristics of the systems and to develop new superior systems. Extension of the voltage range above the present 100 V will be an important part of this work.

5. References

[1] Standards on pulses: definition of terms, Proc. IRE, **39**, No. 6, (June 1951).
[2] F. E. Terman and J. M. Pettit, Electronic Measurements, pp. 34–36, (McGraw-Hill Book Company, 2d ed., 1952).
[3] M. C. Selby, High frequency voltage measurement, NBS Circular 481, (Sept. 1, 1949).
[4] Charles L. Alley and Kenneth W. Atwood, Electronic Engineering, p. 64, (John Wiley & Sons, Inc., New York, N.Y., 1962).
[5] Aldert Van Der Ziel, Solid State Physical Electronics, p. 287, (Prentice-Hall, Inc., 1957).

4.a. **Random sampling oscillography,** G. J. Frye and N. S. Nahman, IEEE Trans. Instr. & Meas. IM-13, No. 1, 8-13 (March 1964).

A description is given of a sampling oscilloscope technique in which the sampling pulse recurrence rate is not in any way synchronized with the signal waveform. The technique eliminates the need for a signal channel delay line. An experimental system which employs only semiconductor devices is described, and experimental data demonstrating the practicality of the system is presented.

4.b. **On the applicability of the comparison method for picosecond pulse instrumentation,** G. H. Honnold and N. S. Nahman, IEEE Trans. Instr. & Meas. IM-13, No. 2 & 3, 123-128 (June-Sept. 1964).

A graphical technique has been developed to allow the determination of the applicability of the Gaddy pulse-comparison system to the measurement of picosecond pulse characteristics. This paper presents the development of this graphical technique and its application to several pulse shapes. Experimental results are given which illustrate the usefulness of the graphical technique for the determination of applicability. The limitations of the system are discussed and its major area of usefulness is mentioned.

4.c. **Measurement of RF peak-pulse power by a sampling-comparison method,** P. A. Hudson, W. L. Ecklund, and A. R. Ondrejka, IRE Trans. Instr. I-11, Nos. 3 & 4, 280-284 (December 1962).

A method is described whereby RF peak-pulse power may be measured by comparison with CW power at the same frequency. Comparison is made by first sampling the power in a small portion of the RF pulse width and subsequently sampling the same portion of the CW signal whose power level is accurately known. Sampling is accomplished with a fast SPDT coaxial solid-state switch synchronized with the RF pulse. A time delay network allows the sample to be taken anywhere along the RF pulse width.
Estimated maximum error for peak power levels to 10 kW is 3 percent. High CW levels are not necessary since the switch is used in conjunction with directional couplers to cover the power ranges of interest.

4.d. **A discussion on the transient analysis of coaxial cables considering high-frequency losses,** N. S. Nahman, IRE Trans. Prof. Grp. Circuit Theory CT-9, No. 2, 145-152 (June 1962).

The physical processes of HF loss are discussed. Two cases in transient analysis for coaxial cables are presented; the first considers skin effect in plated conductors while the second is an analysis based upon an attenuation approximation of the form $(\text{frequency})^m$, $0 < m < 1$. A graphical transient analysis technique is described which allows one to easily analyze cables whose losses are due to a combination of physical processes. Generalized curves based upon the $(\text{frequency})^m$ approximation are presented by which the transient response of a cable can be rapidly evaluated for purposes of engineering design.

4.e. **Transient analysis of coaxial cables considering skin effect,** R. L. Wigington and N. S. Nahman, Proc. IRE 45, No. 2, 166-174 (February 1957).

A transient analysis of coaxial cables is made by considering the skin effect of the center conductor as the distorting element. Generalized curves are presented by which the response of any length of coaxial cable can be predicted if one point on the attenuation vs. frequency curve is known. An experimental check on the analysis is made by comparing measurements and prediction of the responses of several different coaxial cables.

Also see 10-a.

5. Noise

Papers

Abstracts

See also 7.5.

Noise Standards, Measurements, and Receiver Noise Definitions

C. K. S. MILLER, W. C. DAYWITT, AND M. G. ARTHUR

Abstract—**This paper consists of four sections covering 1) basic principles of noise measurement, 2) the switching radiometer, 3) a survey of noise sources, and 4) concepts of noise factor and noise temperature. The first section presents basic formulas used in analyzing radiometers. The second discusses the switching radiometer, briefly tracing its development and usage in the standards field. The third section surveys the development of hot and cold thermal noise sources, noise diodes, and gas-discharge noise generators. The last section presents and discusses the basic definitions of receiver noise performance.**

I. Introduction

NOISE generally implies any random disturbance which corrupts a desired signal and reduces the certainty with which an observation or measurement may be made. In observing a star, the observer becomes aware of the sensation of the flashing or twinkling. The twinkling is apparently random and is caused by fluctuations in the index of refraction of the atmosphere. In this case the medium fluctuates randomly, perturbing any signal traveling through it, and we have the signal partially obscured by noise. A discriminating look at a continuous sinusoidal signal will reveal a randomness in the amplitude and phase which is in itself a corruption to the signal.

The random quality exhibited by fluctuating or noise-like phenomena requires statistical mathematical methods for its analysis. Generally, one must idealize noise-like characteristics to formulate a mathematical model whose properties approximate those of the natural phenomenon. There are numerous references to which the newcomer can turn for aid [1]–[5].

The intent of this paper is to fill in some of the voids that exist in the literature on the subject of noise measurements. The following sections cover four distinct areas. Section II gives the basic principles of noise measurement, presenting four areas of uncertainty in measuring an effective noise temperature by comparison with a standard. Section III discusses the switching radiometer, briefly tracing its development and usage in the standards field. Section IV surveys the field of reference noise sources. The survey covers the development of heated and cooled resistors and tries to draw from the literature the salient points of consideration in constructing either kind. Noise sources requiring calibration are also considered, and noise diodes are reviewed in the light of attempting to show the need for calibration. Gas-discharge noise sources are covered from their inception to the types of devices currently available commercially. Section V is a discussion of the concepts of noise factor and noise temperature as related to their definitions.

Manuscript received March 7, 1967.

The authors are with the National Bureau of Standards, Boulder, Colo.

II. Basic Principles of Noise Measurement

Some techniques and formulas for calculating the effects of thermal radiation in communication systems have proved so useful and convenient and of such widespread application that they are worth gathering in one place. These expressions are most easily derived by assuming thermodynamic equilibrium and applying the second law of thermodynamics. However, they have proved to be applicable to many nonequilibrium situations. Their range of applicability includes systems whose components can be specified by some temperature, either effective or physical, and an impedance. The bandwidth of importance is assumed narrow enough to consider the power density constant in frequency. It is to this type of problem that the following noise formulas apply.

Before proceeding to the formulas per se, it is first necessary to review a number of expressions from microwave network analysis [6]. These expressions are, of course, applicable to lower frequency circuits and reflect only a preference for the language of microwave theory.

For immediate convenience, we introduce the circuit in Fig. 1 and restrict ourselves to noiseless components and to a CW signal source.

In Fig. 1, Γ_g, Γ_1, Γ_2, and Γ_L are reflection coefficients, S is the scattering matrix that characterizes the two-port, and P_{1d} and P_{2d} are the net power delivered to the right of reference planes 1 and 2, respectively. Z_M and Z_N are characteristic line impedances which are not necessarily identical. The reflection coefficients and the scattering matrix S are measured at their respective reference planes 1 and 2. Γ_1 is related to the matrix S and Γ_L by the following formula:

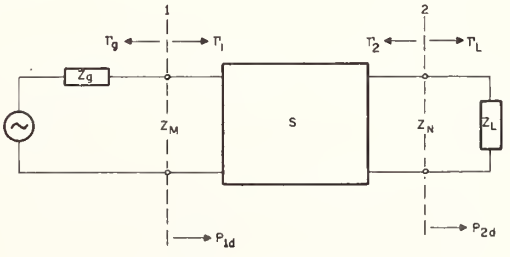


Fig. 1. Circuit schematic used to define M, N, η, and α.

$$\Gamma_1 = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}. \tag{1}$$

A similar expression holds for Γ_2.

There are four derived expressions of interest. They are the mismatch factors M and N associated with the respective reference planes 1 and 2

$$M \equiv \frac{(1 - |\Gamma_g|^2)(1 - |\Gamma_1|^2)}{|1 - \Gamma_g\Gamma_1|^2}, \tag{2}$$

$$N \equiv \frac{(1 - |\Gamma_L|^2)(1 - |\Gamma_2|^2)}{|1 - \Gamma_L\Gamma_2|^2}; \tag{3}$$

the efficiency

$$\eta = \frac{P_{2d}}{P_{1d}} = \frac{(Z_M/Z_N)|S_{21}|^2(1 - |\Gamma_L|^2)}{(1 - |\Gamma_1|^2)|1 - S_{22}\Gamma_L|^2}; \tag{4}$$

and a quantity "α" which is the ratio of the available power at the output to the available power at the input,

$$\alpha \equiv \frac{P_{2a}}{P_{1a}} = \frac{M\eta}{N} = \frac{(Z_M/Z_N)(1 - |\Gamma_g|^2)|S_{21}|^2}{(1 - |\Gamma_2|^2)|1 - S_{11}\Gamma_g|^2}. \tag{5}$$

Obviously,

$$P_{1d} = MP_{1a} \tag{6}$$

and

$$P_{2d} = NP_{2a}.$$

Both α and η are measures of the lossiness of the two-port S in that, *if S is lossless*,

$$\eta = 1 = \alpha. \tag{7}$$

As such, α will turn out to be a direct measure of how much noise the two-port will contribute to the output. Furthermore, α is independent of Γ_L and η is independent of Γ_g.

With the preceding definitions and expressions in mind, it is possible to write very convenient expressions for the relevant noise powers used in radiometer analysis. We assume a constant spectral density across the bandwidth of interest and describe the source in terms of an effective temperature. For example, the available power from an impedance heated to a uniform temperature T is characterized by the physical temperature T itself. That is

$$T \equiv \frac{P_a}{kB} \tag{8}$$

where k is Boltzmann's constant, B is the bandwidth, P_a the available power, and T will be referred to as the available temperature. It should be kept in mind, however, that T need not be a physical temperature. For example, T may correspond to the available power of an impulse noise source whose physical temperature may be quite different from T. Furthermore, we will adhere to the common but sloppy practice of using the terms "temperature" and "noise power" synonymously.

Figure 2 is typical of many radiometers or portions thereof. In Fig. 2 we have a noise source, either under measurement or being compared to another source of noise power; the source is followed by some two-port, for example an attenuator at ambient temperature T_A or perhaps a section of lossy transmission line. We are interested in either the available power, T, at reference plane 2 or the power delivered, NT, across plane 2 to the receiver. NT consists of two parts: a part that is generated by the noise source and delivered to the receiver, $T_sM\eta = T_sN\alpha$; and a part generated in the two-port itself, $NT_A(1-\alpha)$. That is

$$\begin{aligned} NT &= M\eta T_S + NT_A(1 - \alpha) \\ &= N[T_S\alpha + T_A(1 - \alpha)]. \end{aligned} \tag{9}$$

Thus, for example, with proper tuning ($N=1$), the maximum available power to the receiver is

$$T = T_S\alpha + T_A(1 - \alpha), \tag{10}$$

and should the component be lossless,

$$T = T_S.$$

It is this last formula for NT that finds its way into so many radiometer applications. To illustrate its use, it will be applied to the comparison of an unknown noise source to a noise standard. In this case the two-port becomes a precision attenuator and, for convenience, we will assume that its input and output are sufficiently isolated from its interior so that S_{11}, S_{22}, and Γ_2 can be considered to be constant as the attenuator is changed from one setting to another. The comparison is described schematically in Fig. 3 [7]. The quantities in Fig. 3 have meanings analogous to those of Figs. 1 and 2. The comparison is made by adjusting α with the standard first connected to the M reference plane (α_S) and then adjusted again when the unknown is connected to this same plane (α_x) so that the receiver samples the same

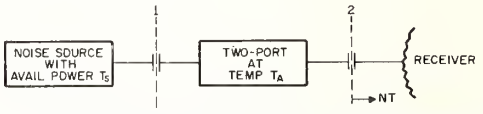


Fig. 2. Block diagram used to depict the net average noise power, exclusive of receiver noise, delivered to the receiver.

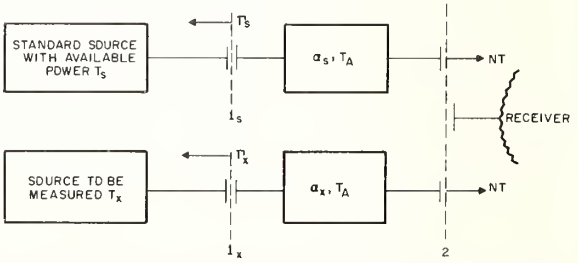


Fig. 3. Block diagram used to determine an unknown noise temperature T_x by comparison against a known noise temperature T_s.

NT both times. Our assumption of isolation insures that N is the same in both cases and that noise generated in the receiver and reflected back from N remains the same in both cases. Therefore, the effect of receiver noise cancels out and need not appear in the equations. For simplicity, we will assume infinite receiver resolution to changes in NT. The balance equation then reads

$$NT_S\alpha_S + NT_A(1-\alpha_S) = NT = NT_x\alpha_x + NT_A(1-\alpha_x),$$

or

$$T_x = (T_S - T_A)\frac{\alpha_S}{\alpha_x} + T_A. \tag{11}$$

Assuming equal characteristic impedances, $Z_{M_S} = Z_{M_x}$ and remembering that Γ_2, S_{11}, and N are constants for the assumed isolation,

$$\frac{\alpha_S}{\alpha_x} = \left|\frac{1-S_{11}\Gamma_x}{1-S_{11}\Gamma_S}\right|^2\left(\frac{1-|\Gamma_S|^2}{1-|\Gamma_x|^2}\right)\frac{|S_{21}|_S^2}{|S_{21}|_x^2}. \tag{12}$$

The *phases* of the quantities S_{11}, Γ_x, and Γ_S are seldom if ever measured. Therefore, the first factor in α_S/α_x is uncertain by approximately $2|S_{11}|(|\Gamma_x|+|\Gamma_S|)$ and indicates why the input to the attenuator section is usually tuned ($S_{11}\approx 0$). This first factor causes what is commonly called "mismatch error." Both $|\Gamma_x|^2$ and $|\Gamma_S|^2$ are usually very small and slight inaccuracies in their measurement have little effect on the accuracy of T_x so that the second factor can be considered to be exact. The third ratio can be measured to a high degree of accuracy, but inaccuracy in its measurement will still noticeably affect the accuracy of T_x. The uncertainty in T_x, then, mainly stems from four causes: uncertainty due to mismatch:

$$(\delta T_x)_M \le \left[1-\left(\frac{1+|S_{11}\Gamma_x|}{1-|S_{11}\Gamma_S|}\right)^2\right]\left(\frac{1-|\Gamma_S|^2}{1-|\Gamma_x|^2}\right)\frac{|S_{21}|_S^2}{|S_{21}|_x^2} \cdot (T_S - T_A) \tag{13a}$$

$$\approx 2|S_{11}|(|\Gamma_x| + |\Gamma_S|)T_x; \tag{13b}$$

uncertainty in the calibration of the S_{21} ratio:

$$(\delta T_x)_{S_{21}} \le \left|\frac{1-S_{11}\Gamma_x}{1-S_{11}\Gamma_S}\right|^2\left(\frac{1-|\Gamma_S|^2}{1-|\Gamma_x|^2}\right)(T_S - T_A)\delta\left(\frac{|S_{21}|_S^2}{|S_{21}|_x^2}\right) \tag{14a}$$

$$\approx T_x\frac{\delta(|S_{21}|_S^2/|S_{21}|_x^2)}{(|S_{21}|_S^2/|S_{21}|_x^2)}; \tag{14b}$$

uncertainty in the standard temperature T_S:

$$(\delta T_x)_{T_S} \le \frac{\alpha_S}{\alpha_x}\delta T_S \approx T_x\frac{\delta T_S}{T_S}; \tag{15}$$

and finally, the uncertainty in T_A:

$$(\delta T_x)_{T_A} \le \left|1-\frac{\alpha_S}{\alpha_x}\right|\delta T_A \approx \left|1-\frac{|S_{21}|_S^2}{|S_{21}|_x^2}\right|\delta T_A. \tag{16}$$

III. Discussion of Radiometers for Noise Temperature Comparisons

There are many types of radiometers [8], [9], and they have varied applications. For example, radiometers are used in radio telescopes to chart maps of radio noise sources in outer space, to measure temperatures and temperature variations of stars and sky, to search for icebergs in the oceans, to guide all-weather missiles, to plot temperature profiles of the earth, to study plasmas, and to seek out cancerous areas in human patients. Basically, the radiometer is an instrument to measure noncoherent radio waves.

The radiometer is used as a comparator of noise sources, usually two noise sources, a standard and an unknown. For this purpose the switching radiometer is most frequently used and will be surveyed here. Colvin [8] and Knight [9] give a more general survey including those radiometers excluded in the present discussion.

In 1946, Dicke [10] suggested using a switch in the front end of a radiometer (Fig. 4). Dicke constructed a wheel made of an absorbing material and shaped in such a way that, when rotated by a motor in a slotted section of the waveguide, it produced a nearly square wave modulation with close to equal times in and out of the waveguide. If we assume the wheel and antenna were both nonreflecting, then the effect of the wheel is one of disconnecting the antenna and connecting an equivalent resistance (the absorbing material of the wheel) to the receiver. Note that if the radiation from the wheel and the antenna is the same no change will be noticed on the output meter, while if there is a difference a modulated signal will result and the deflection on the output meter will change.

This scheme was adapted for use in the measurement of standards [11]. The first modification was quite apparent (Fig. 5). The front end of the radiometer was modified so that the switch alternately connected one or the other of the two noise sources, each of which was actually on a separate arm. As depicted, a matching section could be included if so

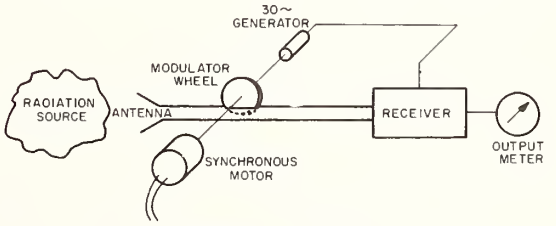


Fig. 4. Schematic diagram of the radiometer used by Dicke.

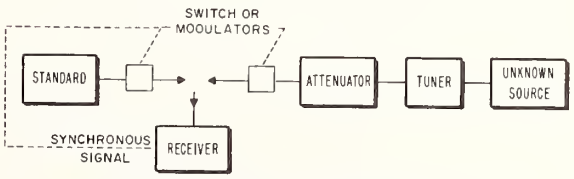


Fig. 5. Block diagram of a switching radiometer used to compare noise sources.

desired. An attenuator was included to lower the noise power level of the unknown, which was usually of considerably higher output than the thermal standard. With subtle differences, and sometimes novel switch designs, this scheme has been employed with a modulator and a synchronous detection signal [12], [13] and with a switch without a synchronous detection signal [14], [15].

The advantage in alternately sampling the two noise sources is that the receiver need only have a stability of the order of time required to complete a full switching cycle. This is extremely valuable since the gain variations and temperature changes encountered are seldom critical for time periods of fractions of a second. The problems caused by receiver instability increase if the switch is of a manual variety and increase even further if the standard and unknown are manually interchanged and connected directly to the receiver [16], [17]. The disadvantages are: 1) the insertion loss of the two paths must be the same or accurately known; 2) asymmetries in the junction of the arms can exist; 3) the switch must be repeatable to the desired degree of accuracy; and 4) the attenuator must have an absolute calibration or the calibration uncertainty in this insertion loss will provide an appreciable error.

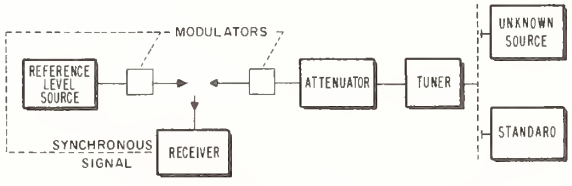


Fig. 6. Block diagram of a modified switching radiometer used for calibration of noise sources.

To improve this type of radiometer, further modifications were made [18]. The modifications (Fig. 6) consist of: 1) using one arm of the previous arrangement as a reference noise source, the only requirement being that it have a stable output over the period of time required to make a measurement, and 2) converting the other arm to a comparison arm, where the comparison of standard and unknown is made. The requirement for the attenuator in the reference arm is that it be stable, and for the attenuator in the comparison arm that it be calibrated for difference attenuation measurements. It should be noted that a difference attenuation calibration is inherently more accurate than an absolute attenuation calibration. In addition, a manually operated switch may be used for ease in connecting standard and unknown to the comparison arm of the radiometer, but the switch must have identical paths (i.e., insertion loss the same) and be repeatable. This modified radiometer has been used by a number of noise metrologists [18], [7], [9], [20], [21]. A further modification of switching the IF rather than the RF signal has also been used successfully [22].

The modified switching radiometer has been analyzed quite extensively and the sources of error are quite well defined [7], [19], [22]. A block diagram (Fig. 7) of the WR90 modified switched radiometer at the National Bureau of Standards is shown. It has a reflectometer incorporated into the comparison arm to simplify tuning the input and measuring the reflection coefficient of the source. To circumvent the problem of matching the attenuator to the transmission line, the attenuator is calibrated with 60 dB of isolation permanently affixed to either side so that the attenuator mismatch effects are included in the calibration.

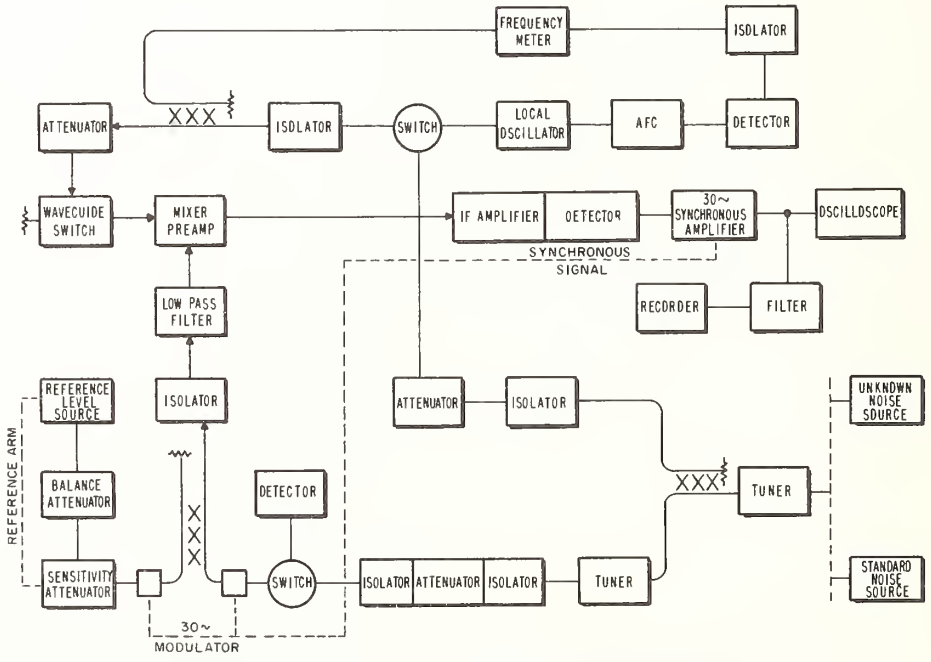


Fig. 7. Block diagram of the NBS modified switching radiometer used to provide microwave noise calibration services.

IV. Survey of Reference Noise Sources

A. The Basic Standard of Noise

Physical measurements of natural phenomena are limited by thermal noise. The noise can be minimized or reduced to the degree that the physical temperature can be reduced. Since thermal noise is directly related to temperature we have the basis of a basic standard. The determination as to whether a thermal noise source can be used as a basic standard lies in whether or not its output is calculable from its dependent parameters.

1) Heated Resistor for Rectangular Waveguides: A heated resistor has been recognized as a thermal noise standard for some time. The difficulty comes in making a practical noise source for use at radio frequencies. A further difficulty arises in applying an analysis to the source which can satisfactorily calculate the noise output.

In 1928, Johnson [23] showed experimentally that a resistor with no externally applied voltage across it has a measurable electrical noise across its terminals. In the same year, Nyquist [24] reported a theoretical calculation of thermal-noise voltage in a resistor. This work has been followed by those who have tried to fabricate practical thermal-noise standards. By present standards some of these attempts may be judged to have been a little crude and somewhat optimistic. Many of these attempts have made contributions to the state of present thermal-noise standard capabilities by the thoroughness of their reporting.

In 1956, Hughes [13] constructed a "hot load" as he called the resistive wedge that he heated to temperatures between 160°C and 230°C. Hughes used three copper-constantan thermocouples (calibrated against a batch of thermometers which were themselves consistent to within 0.5°C) to measure his wedge. The thermocouples were imbedded into the wedge. He reported his error from his heated resistive noise source to be less than 0.5°K. However, Hughes reported no allowances made for the effects of the gradient in the waveguide connecting his noise source to the radiometer.

In the same year, Sutcliffe [14] constructed a resistive wedge which he placed in a steel waveguide to generate thermal noise. He used chromel-alumel thermocouples (calibrated against thermometers to 360°C) and operated the wedge at a nominal 600°C. Sutcliffe did not construct his thermocouples into his wedge so his temperature measurements were taken during simulated rather than operational conditions. He concerned himself with attenuation effects of the gradient in the connecting waveguide but since his measured values in the hot and cold states were so similar he concluded this effect was negligible. Sutcliffe reported that although he could maintain the source temperature to within 1°C, the thermal noise power source was known to only 10°C.

In 1957, Knol [16] constructed a heated resistive noise source using a platinum waveguide and a zinc-titanate based wedge. Knol heated his wedge to 1063°C and reported being able to keep it at the melting point of gold to within ±1°C. He did not consider the effects of the connecting waveguide; however, he chose platinum to avoid oxidation of the waveguide.

In the same year, Reynard [25] constructed a quartz wedge that was coated with platinum and silicon monoxide. The wedge was placed in a nickel WR 12 waveguide and the operating temperature ranged between 600°C and 700°C. Reynard tried to avoid a waveguide that could become lossy through oxidation. He tried to measure the gradients of the wedge and connecting waveguide as best he could for the size waveguide he was using. He reported that he could maintain the gradient of the load to less than 1°C. There have been other workers who have constructed heated resistors for millimeter wavelengths [26].

In 1958, Sees [27] discussed some considerations and sources of errors to be considered in constructing and using hot-body noise sources. Sees recognized that the waveguide would not only attenuate the noise power generated by the heated resistive wedge but that the waveguide would also generate noise power because of its resistive qualities. Both the attenuation and generation would be functions of the temperature gradient of the waveguide. He correspondingly formulated a method of calculating the resulting macroscopic attenuation and generation effect.

With this background, Estin et al. [18] constructed a hot-body noise source using a silicon-carbide wedge and a gold waveguide. Their oven was a vertical oven and their effort was not completely successful, but it did provide a good beginning for further work at the National Bureau of Standards to develop a hot-body reference noise standard. Later, Wells et al. [7] reported their progress in the development of a hot-body noise source and the National Bureau of Standards announced the availability of a calibration service [28].

Wells et al. used two hot-body noise sources; one was a silicon-carbide wedge in a gold waveguide, the other a zinc-titanate wedge in a platinum-13 percent rhodium waveguide. The shape of the two wedges were different so in effect two entirely different designs were used. They operated the two noise sources over wide temperature ranges. These features showed their standard to be calculable since the differences in configuration did not affect the results, and variations in material and operating temperature could be accounted for as corrections to the noise output to the hot-body noise source. Wells et al. reported having implemented Sees' [27] method of calculating a correction to the noise power output of the resistive wedge due to the resistivity and temperature gradient of the connecting waveguides. They reported the gradient of the wedge to be less than 1°K, but due to the calibration of the thermocouples to ±2.5°K and the temperature effects in the connecting waveguide, they claimed the effective temperature of the load known to an accuracy of ±4.1°K at 1000°C.

It would be remiss not to recognize the work of Birger and Sokov [29], who were apparently unaware of the efforts of all who preceded them with the exception of Sutcliffe. They constructed a wedge of green carborundum on a

ceramic base and a connecting waveguide of nickel with internal gold-plating as their thermal noise generator. The operating temperature was a nominal 600°C. They considered the attenuating effects of the connecting waveguide. They constructed a total of four noise generators, one in each waveguide size WR90, WR137, WR187, and WR284, and reported the effective temperature errors not exceeding ±8°C.

Recently, Liedquist [19] reported that the Research Institute of Swedish National Defence (FOA) had purchased a hot-body noise standard from an industrial organization in Japan. The silver-palladium alloy waveguide is gold plated on the inner surface. The reported operating temperature is 1000°K using a single thermocouple for measurement and control. The waveguide necessarily uses a transition from circular to rectangular waveguide. Liedquist reported knowing the wedge temperature to 3°K and the effective temperature to 8°K.

Also, recently, Halford [22] reported the construction of a thermal standard composed of a zinc-titanate wedge in a waveguide composed in part of duralumin and in part of silver plated steel and operated at 400°C. During noise comparisons, temperature measurements are performed by a platinum resistance thermometer located in a channel of the bar which contains the waveguide port. He correlates his thermometer readings with the temperature of the wedge by an auxiliary experiment where he actually measures the wedge temperature with a thermocouple probe. Halford reported knowing the wedge temperature to ±0.5°K and the effective temperature of the source to ±1.5°K.

2) Heated Resistor for Coaxial Waveguide: Halford [22] reported a second effort to build a thermal standard in a coaxial configuration for operation in the 2–4 GHz frequency range. This arrangement uses a carbonyl iron powder bounded by a suitable adhesive cement as the material of the wedge. The wedge is cone shaped with the inner conductor passing through its center and tapered towards the outer conductor. The wedge is operated at 400°C and the temperature measured by a thermometer in an auxiliary channel neighboring the outer conductor. The noise power is coupled to the waveguide by a tunable coaxial to waveguide adapter arrangement. He reported some room for improvement since the coaxial wedge had a local gradient of 3°K.

Zucker et al. [30] built a coaxial heated resistor noise source in 1958. It was designed for use in the 0–1000 MHz frequency range with an operating temperature of 1300°C. The generator was designed with the outer conductor tapered to the inner conductor. The 50-Ω resistive character of the generator was obtained by the deposition of a pyrolytic carbon film on a ceramic base. Temperature measurements of the generator were made by a pyrometer through special ports in the side of the oven. No consideration is recorded by Zucker et al. of the attenuation and generation effects due to the temperature gradient in the connecting coaxial line. They reported that an experienced operator should be able to measure the source temperature to ±3°K.

Gordon-Smith and Lane [20] briefly describe a coaxial heated resistor noise source in a paper published in 1964. They used a 50-Ω metal-oxide resistor having a low temperature coefficient of resistance and an outside diameter of 0.3 cm and length of 1.0 cm. They report that the resistor was heated by immersion into a stirred oil bath maintained at 280°C and that the temperature gradient of the resistor was less than 1°C. Gordon-Smith and Lane considered the attenuation of the coaxial line but not the generation of noise in the line. They reported knowing the effective temperature to ±3.75°K.

Also in 1964 Rusinov and Sorochenko [31] reported the construction of a heated cylindrical carbon resistor for use as a thermal noise standard. The cylindrical carbon resistor is heated from within and is the center conductor as shown in Fig. 8. The outer conductor is composed of two adjoining conical transition sections which are finned for cooling. The resistor is operated at 200–250°C and an accuracy of 3 to 4 percent is claimed.

From the efforts of these authors we recognize that the two major contributors to errors to a heated-resistor noise source are: 1) determining the temperature of the resistor, and 2) evaluating the effects of the connecting transmission line on the output of the resistor.

For determining the effective temperature of the resistor, ideally we would want the resistor to be homogeneous, have a conjugate impedance match, and an infinite thermal conductivity. Similarly, in determining the effects of the transmission line, ideally we would want a stable material with infinite electrical conductivity.

Having less than the ideal resistor, we must measure 1) the gradient in the resistor material to determine an average temperature, and 2) the gradient in the material (from imbedded temperature sensors to the radiating surface) to determine the average temperature of the radiating mass. Unfortunately in physical models the rate of radiation in a transmission line enclosure is not uniform across the radiating surface of the resistor. For one shape of wedge used at the National Bureau of Standards the majority of the power is radiated from a cross section approximately one-third the length of the radiating surface back from the tip of the

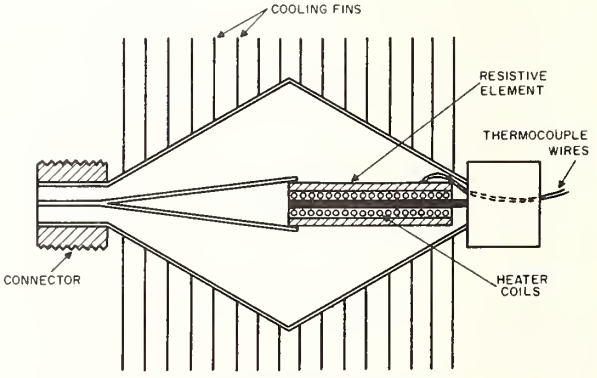


Fig. 8. Schematic diagram of the coaxial thermal noise source used by Rusinov and Sorochenko [31].

wedge. Generally speaking, the lower the temperature gradient, the greater the certainty of the temperature.

Evaluating the effects the transmission line has on the output of the resistor requires 1) knowing the temperature gradient of the transmission line, and 2) knowing the resistive properties of the transmission line as a function of temperature. The attenuation effect of the transmission line can be measured, but for small attenuation values the measurement may be less accurate than calculating the attenuation. This calculation has been programmed for computer solution. To calculate these effects, a material must be selected for the transmission line that will not oxidize at the highest temperatures expected or have uncertain Curie point effects.

The attenuation effect will lower the effective noise temperature while the generation effect will raise the effective noise temperature. These effects can be added because uncorrelated random noise signals are additive. With a heated resistor the attenuation effect is larger than the generation effect; hence, the total effect of the transmission line will be to lower the effective temperature of the resistors. However, considering only the attenuating effect of the connecting transmission line results in a greater error than totally neglecting both attenuation and generation.

3) Cooled Resistor: The noise characteristics of a low-noise system can be more accurately measured by using a cooled resistor as opposed to a heated resistor. Depending on the degree of cooling required, the cooling mechanism will vary. When cryogenic fluids are used, the complexity of the apparatus can compare to that of apparatus used with high temperature resistors.

In 1961, Stelzried [32] constructed a cold resistor in a connecting coaxial line. The coaxial transmission line had both stainless steel conductors coated first with 50 microinches of copper followed by 40 microinches of gold. The resistor was a nominal 50 Ω constructed to give a VSWR less than 1.05 at 960 MHz when immersed in L-He (liquid helium, 4.2 K). He measured the gradient of the coaxial transmission line for differing L-He levels and gives a graph of his calculated equivalent noise temperature for these levels.

In 1962, Mezger and Rother [33] constructed, for the 1–5 GHz frequency range, three configurations where the resistive material was the same and changes were made in the connecting coaxial line. In each case the inner conductor had a 6-mm OD and the outer conductor had a 16-mm ID and was 2 mm thick. One coaxial line was formed from silver-plated brass. A second coaxial line was formed with a 100-mm intermediate section formed from German silver. This section had the wall thickness reduced to 0.5 mm for both conductors (this means the inner conductor was hollow for this section) with the intent of minimizing the heat transfer. The third coaxial line had, in addition to the insulating section, a vented hollow inner conductor to permit the cryogenic fluid to rise to the same level in the inner conductor to that surrounding the outer conductor. Using the first two configurations, the effectiveness of the insulating section using L-N_2 (liquid nitrogen) lowers the noise by 2.5 K. They used the second two configurations to evaluate the effectiveness of cooling as a function of contact surface by changing the depth of L-N_2.

In 1964, Eisele [34] constructed a wedge in a section of WR137 waveguide which was cooled to L-He temperatures. The wedge was cut from a single crystal ruby. It was sliced to 0.050 inch, tapered and coated with nichrome and inserted in the connecting waveguide in a manner assuring good thermal contact. For the insulating section, he used a copper-plated stainless steel waveguide. Eisele selected a configuration where the waveguide connection emanated from the bottom of the cryostat. He reported that the noise temperature of the refrigerated termination, excluding transmission line contributions, approached the temperature of the bath to less than 1 °K.

In 1965, Stelzried [35] reported having constructed a WR284 waveguide L-He cooled termination. He used a copper-plated stainless steel section as an insulating section and reported an equivalent noise temprature referenced to the connecting flange as 5.0°K.

Also in 1965, Penzias [36] reported having constructed a WR229 waveguide L-He cooled termination as a reference noise source. The waveguide was 90 percent copper brass. The wedge had a pyramid shape and was bathed in L-He. A Mylar septum prevented the L-He from filling the waveguide. Penzias used the same design with L-N_2. With the L-N_2 he measured a VSWR of 1.05 of the termination compared to 1.04 for L-He. Penzias reported that, for the case of L-N_2, the effective equivalent noise output of the device was calculated to be 78.1 K. Penzias measured the temperature gradient of the waveguide by diode thermometers and it appears that he calculated the effects of the connecting waveguide based on Stelzried's [32] earlier work.

In the same year, Jurkus [37] reported having constructed two WR137 waveguide cooled noise sources. The load element, as Jurkus called the resistive material, is made from a resin loaded with iron powder which is shaped in four wedges which fit into the corners of the waveguide. A copper-plated thin-walled 6-inch section of stainless steel waveguide is used as a thermal insulator. The temperature gradient was used to calculate the effect of the connecting waveguide by the method suggested by Wells et al. [7]. Jurkus calculated the effects of the gradient on each of the two noise sources for immersion in L-N_2 and L-O_2 (liquid oxygen). He also reports on the effect of the depth of immersion in L-N_2 on the gradient.

In 1966, Menon and Albaugh [38] reported having constructed three cooled noise sources for use at 2 cm, 9.5 mm, and 3.5 mm. For the thermally isolating section, thin-walled brass and electroformed copper waveguide were used. The resistive element was horizontally positioned through the side of a flask close to the bottom. This configuration was used to reduce the dependence of the effective temperature on the liquid level of the cryogenic fluid (L-N_2 in this case).

From the efforts of these authors, we can surmise major difficulties that will need consideration in construction and in determining the errors encountered in using a cooled-resistor noise source. The same two major contributions to

errors that were mentioned with heated-resistor noise sources will apply for cooled-resistor noise sources, namely: 1) determining the temperature of the resistor, and 2) evaluating the effects of the connecting transmission line on the output of the resistor.

The same ideal conditions apply as previously stated. With limitations in materials that we can choose for the resistor, we can plan for the most effective thermal contact with the coolant as possible. This means keeping the resistor immersed deep enough or bathed in the coolant. The temperature of the coolant must be known. The boiling point of cryogenic coolants are well defined and they are dependent on pressure. Altitude changes can effect the boiling points of cryogenic fluids and, to a lesser degree, so can local barometric pressure changes. Designs for resistors that provide a reflection coefficient magnitude Γ or VSWR that is a minimum and is independent of temperature are preferred.

The effect the transmission line has on the effective noise temperature of the resistor requires 1) knowing the temperature gradient of the length of transmission line, 2) knowing the resistive properties of the transmission line as a function of temperature, and 3) knowing that the transmission line is free of any condensates. The same comments made previously in regard to heated resistors still apply. The attenuation and generation effects can probably be calculated more accurately than measured for known resistive properties and a measured temperature gradient of the transmission line. The same computer program solves the gradient effects of cooled loads. With cooled resistors, the connecting transmission line has the effect of raising the effective noise temperature output of the resistor, since the generation of noise in the transmission line is greater than the attenuation. Here, too, considering only one effect results in a greater error than in neglecting the total effect.

Keeping the transmission line free of condensates is an additional significant problem. Evacuating, purging, or pressurizing, or some combination of these three, is necessary to keep the transmission line free. For waveguide transmission lines, a window of some kind is usually necessary; polystyrene [35], Teflon [37], or Mylar [34], [36], [38], have been used. A configuration where changing coolant liquid level does not change the effective temperature is preferred. The transmission line connector should be considered as a source of error and has not been identified as such by any of these authors. The new 14-mm and 7-mm precision connectors for coaxial line are improved over the Type *N*. Still these precision connectors and good flanges (in waveguide) can be serious sources of error to cooled resistors.

B. Noise Sources Needing Calibration or Comparison to a Standard

Noise sources that do not have completely calculable outputs cannot qualify as primary standards. If, however, they are stable, repeatable from one usage to another, and have bandwidths broad enough that they can be used to make noise measurements, then they are valuable in that they may be compared to a standard and so function as an interlaboratory standard. In the case where no absolute standard exists per se, then obviously no comparison can be made. This is a condition that had existed in noise for quite a long time but is now being remedied. The noise sources being used for measuring noise figure, noise factor, and effective input noise temperature need a value and, in the absence of a standard, one of two procedures has been followed. Either a single noise source was isolated and assigned a value and in turn used to assign similar values to other comparable noise sources, or each noise source used had an output that was calculated by an agreed method based on operating conditions and known parameters. The agreed method of calculation needs to be uniformly adapted by all users so that comparable devices have comparable outputs. Both of these methods are still in use in the noise field where standards are lacking. Unfortunately the adoption of these methods can encourage apathy toward the development of standards, and sometimes measurements are attempted with inadequate standards. In this light it may be said that the noise source field is undergoing growing pains.

The noise sources generally used in the field are of two major varieties: a) the noise diode, and b) the gas-discharge noise source. The noise diode is used to around 500 MHz and can be extended in frequency range to higher frequencies with some modifications. The gas-discharge noise source is a comparatively new addition to the noise field and it has been used successfully from 300 MHz and into the millimeter-wave regions. A pleasant discussion of these two noise sources is to be found in Hart [17]. Maxwell and Leon [39] recognized the need for calibrating noise diodes and used a hot-cold resistive standard in 1956. Prinzler [40] used a heated resistor as a standard for his calibration of noise diodes in 1958.

1) The Noise Diode: A noise diode has an output that is generally referred to as shot noise. Shot noise normally implies a fluctuating current, although it could be applied to the fluctuating voltage developed across a conductor through which shot noise current flows. Schottky [41] observed that a flow of electrons from a thermionic cathode was irregular because each emission was a random, independent event. The fluctuating current caused by this irregular flow of electrons he called the "Schroteffekt," from which we get the term "shot noise."

To cover the literature written about the noise diode would be an extensive work in itself. There has been a bibliography [42] made of many of the published papers in noise prior to 1954, so only a few pertinent papers will be referenced.

Tien and Moshman [43], through use of a mathematical model based on the generation of random numbers by the Monte Carlo method, simulated the electron emission at the cathode of a noise diode. Using this, they tried to predict results at low frequencies and high frequencies. They felt the results agreed well with existing theory for low frequencies, but found no correlation at high frequencies. Kosmahl [44] tried to measure the correlation between the induced grid noise current and the fluctuations in the output, but found none.

Most of the noise diodes presently on the market have been described in the literature [45] at one time or another. These devices have their own characteristics as a function of their tube geometry. They are often frequency sensitive and have differing emitter characteristics depending on the material and construction of the emitter. They have noise output characteristics that have some predictability, and it depends on how well the output must be known as to how much faith you should put in the calculated output. In England, a noise generator designed by Harris [46] is employed as an interim noise standard.

The Soviet Union established a thermal noise standard in 1964 in the decimeter bands for the calibration of noise diodes and gas-discharge noise sources [31]. The National Bureau of Standards expects to offer a calibration service at 3 MHz based on a thermal noise standard before the end of 1967.

2) The Gas-Discharge Source: The first suggestion to use a gas-discharge tube as a noise source was made by Mumford [47]. Mumford first used a fluorescent lamp mounted in the *H*-plane of a waveguide. The fluorescent lamp, he reported, gains most of its noise radiation from the mercury gas-discharge in the tube. Mumford used lamps with 10 different types of fluorescent coatings and one germicidal lamp (i.e., no fluorescent coating) to find that the fluorescent coating was not the source of the noise since all of them had comparable noise outputs. He concluded that it was the gas-discharge alone that produced the noise.

The fluorescent lamp noise source appears to have been quite temperature sensitive and some effort was made trying to predict the noise output as a function of mount temperature [48], [49]. The statement has been made that "the noise power, in general, is derived both from the thermal velocities, which are characterized by the electron temperature, and from the dc power, which is characterized by the average current" [50]. The facility of being able to calculate the electron temperature and identifying it with the noise temperature when applied to pure gas discharges provided the basis for suggesting that gas-discharge noise sources be used as microwave noise standards [15]. The trouble is that for pure gas discharges the measured values obtained [51]–[53] did not seem to always support these claims; in fact, the noise temperature is usually less than the electron temperature. These discrepancies could mean that there exist unaccounted for effects [11], [54], or even that the noise temperature as measured is not simply related to the electron temperature. Consequently, gas-discharge noise sources need calibration against a standard to ascertain a numeric value that will classify the noise output.

Johnson and Deremer [53] had trouble extending Mumford's design [47] to 10 GHz, since this design gave a drastically reduced bandwidth of the noise source. They found that with a tube mounted in the *E*-plane of the waveguide at 10° to the axis of the waveguide that the VSWR was much lower than a comparable design using the *H*-plane. Decreasing the angle lowered the VSWR, but lengthened the elliptic waveguide cut to accommodate the tube and also meant the tube itself had to be lengthened. No optimum design was resolved with these limitations. Johnson and Deremer suggested a 10° *E*-plane mount for general use and suggested that a smaller angle could be selected for more exact measurements. In their experiments they varied the gas fill presure from 3 to 30 mm of Hg and tried different gases for gas fill in the tubes. They found gas mixtures minimized problems with fluctuations in the plasma, but also appeared to slightly increase the VSWR. Tubes were designed to have a discharge diameter of 1/4 to 1/3 of the inside width of the waveguide and to operate at discharge currents of 200–250 mA which meant gas fill pressure was to be 20–30 mm of Hg. To eliminate the dependence of noise output on the operating temperature, only inert gases were used.

Originally, Johnson and Deremer designed a series of mounts for use in the 3 to 30 GHz range. This design has been available commercially and has only been modified by manufacturers to satisfy their own specifications. We note that Knol [51] independently built a 10° *E*-plane mount for his own use at about the same time. Others in the field have extended this design for general usage and to increase the upper frequency limits [55], [25], [56], [57].

More recently, it was found that the tube in the typical Johnson and Deremer mount design, if once removed, could not easily be replaced and still repeat the effective noise temperature originally measured [58]. Other weaknesses in current commercial design were noted and an improved mount design was suggested with a 9° insertion angle [58] and a 7° insertion angle [22]. Further, an improvement in tube design is also suggested [22].

To extend the useful range of the gas-discharge noise source to lower frequencies, Johnson and Deremer [53] tried to construct tube shapes that would fill the entire cross section of the coaxial transmission line with a gaseous discharge. They did not regard this work as the most successful but could not pursue the matter due to the termination of the contract. They were aware of a contractural program designed to construct a coaxial helical mount and did not pursue this avenue since it would have been a duplication of effort.

To build a good helical-type coaxial mount we need sufficient coupling between the gas-discharge and the transmission line, and a good impedance match. The helical center conductor is a slow-wave structure as is a lumped-constant low-pass filter. Spencer and Strum [59] believed it was possible to make a coaxial mount on the helix idea that would be usable from 30–5000 MHz. Such a design hinged upon whether the gas tube could be constructed long enough for the necessary length of helix. Spencer and Strum give a lot of design criteria in their paper and state they built a mount having a flat noise output from 200 MHz to 3 GHz. Using the low-pass filter concept they built a coaxial source usable from 50 to 300 MHz. Others have reported having successfully built coaxial helical mounts for gas-discharge tubes [39], [60]–[63].

Alma'ssy and Frigyes [64] present a novel suggestion on building a mount for a gas-discharge noise source. They constructed a coaxial transmission line mount where the inner conductor is replaced by a gas tube having comparable dimensions to the inner conductor. "The filament voltage

reaches the tube through a broadband metal extension while the anode is directly connected to the inner conductor." One end may be terminated, thereby simplifying the construction. The device was built, and with a discharge current of 120 mA no difference in output was detected between six tubes. The device was operated between 1800 and 2200 MHz. Another different idea was to construct a gas-discharge tube in two-conductor lines [65].

At present the National Bureau of Standards has the capability to perform calibrations in the frequency range 8.2 to 18.0 GHz, which is covered by two waveguide sizes (WR90 and WR62). The National Bureau of Standards expects to perform calibrations in the 2.60 to 3.95 GHz frequency range (waveguide size WR284) in the immediate future. The remaining frequencies have no standards available to them for calibration. To take up this void Lee and Olsen [66] devised a technique where, by fabricating special tubes, they made transitions between bands.

V. Concepts of Noise Factor and Noise Temperature

A. Introduction

In 1942, North wrote a paper on "The Absolute Sensitivity of Radio Receivers" [67]. This is perhaps the earliest paper that proposed a standardized measurand for the noise performance of a radio receiver. The author suggested a number, N, that he termed "noise factor," as being a fundamental description of the internal noise of a receiver. In 1944, Friis [68] wrote a paper in which he discussed "noise figure" at greater length, and this paper is considered by many to lay the foundation for this quantity. Since 1944, many papers on the subject of the noise performance of amplifiers and receivers have appeared. Such papers have been stimulated by the need to describe this aspect of devices, and by the need to clarify the meaning and application of the noise factor concept.

In recent years, the noise performance of receivers and amplifiers has been described in terms of an effective input noise temperature [69], T_e. For many persons, T_e is a more meaningful measurand than is noise factor, F. However, for a device for which both quantities are defined, a simple relationship exists between the two (23).

The concept of noise temperature comes quite naturally from the relationship between the noise power P_a available from a resistor at a uniform temperature T, which is

$$P_a = kT \text{ watts per hertz.} \tag{17}$$

For a device having a noise power P_a per unit bandwidth available from one of its ports, a noise temperature of T degrees Kelvin can be associated with that port. Thus the internal noise of an amplifier or receiver, or in fact of any network containing a source of noise, can be expressed in terms of a noise temperature.

B. Noise Factor (Noise Figure)

The noise factor (noise figure), $F(f)$, of a linear two-port, at a specified input frequency f is defined as the ratio of 1) the total noise power per unit bandwidth N_t at a corresponding output frequency available at the output port when the noise temperature of the input termination is standard (290°K), to 2) that portion of 1), N_s, that is engendered at the input frequency of the input termination at the standard noise temperature (290°K) [70]. Thus

$$F(f) = \frac{N_t}{N_s}. \tag{18}$$

Although this definition is concise, complete, and useful, it must not be misapplied or confused with the more than fifteen additional and different types of noise factor which may be found in the literature.

An alternate definition of noise factor is based upon the concept of signal-to-noise ratio [68]. If S_i/N_i and S_0/N_0 are the signal power-to-noise power ratios at the input port and output port of the linear two-port, respectively, when the input input termination is at 290°K, F is defined by the relationship

$$F = \frac{S_i/N_i}{S_0/N_0}. \tag{19}$$

Thus, noise factor serves most directly as a measure of the extent to which the noisy two-port degrades the signal-to-noise ratio of the input power.

Equation (19) also shows that, in order to produce a unity output signal-to-noise ratio when the source temperature is 290°K, a signal power S_i' is required where

$$S_i' = FkT_0 \text{ watts per hertz} \tag{20}$$

and

$$T_0 = 290°\text{K}. \tag{21}$$

Noise factor, $F(f)$, as defined by the IEEE is a rather idealized concept. It describes the noise performance of a *linear two-port* at a *single* operating frequency. To describe the noise performance over a *band* of frequencies, an average noise factor, $\bar{F}$, is defined [70]. To distinguish the two, the term spot noise factor [70] is sometimes used to identify the single frequency factor. Noise factor is not defined for a nonlinear two-port, nor for a multiport transducer [see Section V-C)].

Noise factor is a hyperbolic function of source impedance. It reaches a minimum value at some value of source impedance called the "optimum source impedance." This minimum noise factor is called the optimum noise factor [71], and is independent of source impedance.

Noise factor is based on the temperature of the source impedance being standard, a condition that does not always prevail in practice. Thus, the noise performance of a transducer will differ from that predicted by its noise factor when it operates in a system where the temperature of its input termination is not standard. Furthermore, since the experimental procedures for measuring F can be carried out using source impedance temperatures that are different from 290°K, the quantity that is often measured is not F but, rather, an analogous quantity F'. F' is defined the same way as F except that the noise temperature of the in-

put termination, T_s, is taken to be that which prevails during the measurement. It can then be shown that

$$(F' - 1)T_s = (F - 1)T_0 \tag{22}$$

so that

$$F = 1 + (F' - 1)\frac{T_s}{T_0}. \tag{22a}$$

In practice, effort is made to reduce F to its minimum value of unity, thus making the two-port as near as possible to an ideal, noise-free transducer.

C. Effective Input Noise Temperature, T_e

The effective input noise temperature T_e (of a multiport transducer with one port designated as the output port) is defined as the noise temperature in degrees Kelvin which, assigned simultaneously to the specified impedance terminations at the set of frequencies contributing to the output, at all accessible ports except the designated output port of a noise-free equivalent of the transducer, would yield the same available power per unit bandwidth at a specified output frequency at the output port as that of the actual transducer connected to noise-free equivalents of the terminations at all ports except the output port [69].

To help understand this definition, two block diagrams are shown in Fig. 9. In the upper diagram, the input ports of the *noise-free equivalent* of the multiport transducer are terminated in source impedances, each at temperature T_e. The power per unit bandwidth available from the transducer is P_L. In the lower diagram, the input ports of the *noisy* multiport transducer are terminated in source impedances of the same impedance values as before, but each such input termination is noise-free (noise temperature of 0°K). The power per unit bandwidth available from this transducer is also P_L. The value of T_e necessary to produce the same value of P_L in both diagrams is equal to the value of the effective input noise temperature of the (actual) noisy multiport transducer.

Note that each of the diagrams of Fig. 9 requires the use

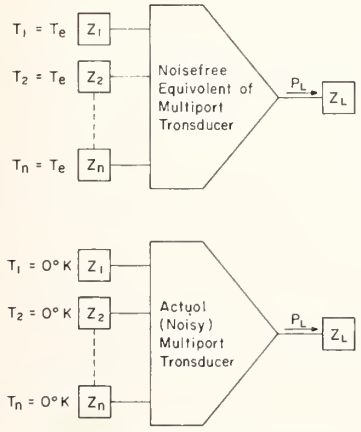


Fig. 9. Diagrams illustrating the definition of T_e.

of fiction (e.g., noise-free equivalent components). However, the real quantity, T_e, of the real, noisy multiport transducer is established through the use of this fiction. Furthermore, no difficulty is presented to the measurement of T_e through this use.

The definition of T_e pertains to output power at a single frequency. As with noise factor, an average effective input noise temperature, $\bar{T}_e$, is defined to describe the noise performance of a multiport transducer over a band of output frequencies [70].

For a linear two-port transducer with a single input and a single output frequency, T_e is related to F [70] by

$$T_e = 290\,(F - 1). \tag{23}$$

Thus the dependence of T_e upon source impedance is also a hyperbolic function. An optimum value of T_e occurs when the input ports are terminated with source impedances that minimize T_e; this optimum T_e is independent of source impedance, and is characteristic of the transducer alone.

T_e is a function of frequency, but is not a function of the temperatures of the input terminations, except that in a nonlinear transducer it may be a function of signal level.

D. Operating Noise Temperature, T_{op}

In a system of transducers, such as may be comprised of one or more sources (e.g., an antenna) and one or more amplifiers, filters, and frequency translators connected in cascade, a useful figure of merit is the operating noise temperature, T_{op}, of a *system* under operating conditions. T_{op} is defined as the temperature in degrees Kelvin given by

$$T_{op} = \frac{N_0}{kG_s} \tag{24}$$

where N_0 is the output noise power per unit bandwidth at a specified output frequency flowing into the output circuit (under operating conditions), k is Boltzmann's constant, and G_s is the ratio of 1) the signal power delivered at the specified output frequency into the output circuit (under operating conditions), to 2) the signal power available at the corresponding input frequency or frequencies to the system (under operating conditions) at its accessible input terminations [67].

In a linear two-port transducer with a single input and a single output frequency, if the noise power originating in the output termination and reflected at the output port can be neglected, T_{op} is related to the noise temperature of the input termination (source impedance), T_i, and the effective input noise temperature, T_e, by the equation

$$T_{op} = T_i + T_e. \tag{25}$$

An average operating noise temperature, $\bar{T}_{op}$, is defined [69] when performance over a band of output frequencies is required.

In practice, a system with the smallest T_{op} will provide the best performance from the standpoint of noise degradation of information.

E. Measurement Techniques for $\bar{F}$ and $\bar{T}_e$

Various methods have been devised for measuring the noise factor and/or effective input noise temperature of transducers. A fairly complete description of these methods has been covered previously [70], [72].

References

[1] W. R. Bennett, *Electrical Noise*. New York: McGraw-Hill, 1960.

[2] D. K. C. MacDonald, *Noise and Fluctuations: An Introduction*. New York: Wiley, 1962.

[3] A. Van der Ziel, *Noise*. New York: Prentice-Hall, 1954.

[4] J. J. Freeman, "Electrical noise," *Electro-Technology*, pp. 126–144, November 1960.

[5] W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*. New York: McGraw-Hill, 1958.

[6] D. M. Kerns and R. W. Beatty, *Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis*. New York: Pergamon, 1967.

[7] J. S. Wells, W. C. Daywitt, and C. K. S. Miller, "Measurement of effective temperatures of microwave noise sources," *IRE Internat'l Conv. Rec.*, pt. 3, pp. 220–238, March 1962; also in *IEEE Trans. on Instrumentation and Measurement*, vol. IM-13, pp. 17–28, March 1964.

[8] R. S. Colvin, "A study of radio-astronomy receivers," Stanford Radio-Astronomy Inst., Publ. 18A, October 31, 1961.

[9] J. Knight, "Evaluation and analysis of radiometers," thesis, University of Toronto, Canada, September 1961.

[10] R. H. Dicke, "The measurement of thermal radiation at microwave frequencies," *Rev. Sci. Instr.*, vol. 17, pp. 268–275, July 1946.

[11] L. W. Davies and E. Cowcher, "Microwave and metre wave radiation from the positive column of a gas-discharge," *Aust. J. Phys.*, vol. 8, no. 1, pp. 108–128, 1955.

[12] J. J. Freeman, "Noise comparator for microwave," *Television News* (*Radio-Electronic Engineering Section*), vol. 49, p. 11, March 1953.

[13] V. A. Hughes, "Absolute calibration of a standard temperature noise source for use with *S*-band radiometers," *Proc. IEE* (*London*), vol. 103, pt. B, pp. 669–672, September 1956.

[14] H. Sutcliffe, "Noise measurements in the 3-cm waveband using a hot source," *Proc. IEE* (*London*), vol. 103, pt. B, pp. 673–677, September 1956.

[15] K. W. Olsen, "Reproducible gas-discharge noise sources as possible microwave noise standards," *IRE Trans. on Instrumentation*, vol. I-7, pp. 315–318, December 1958.

[16] K. S. Knol, "A thermal noise standard for microwaves," *Philips Research Repts.*, vol. 12, pp. 123–126, April 1957.

[17] P. A. H. Hart, "Standard noise sources," *Philips Tech. Rev.*, vol. 23, pp. 293–324, July 1962.

[18] A. J. Estin, C. L. Trembath, J. S. Wells, and W. C. Daywitt, "Absolute measurement of temperatures of microwave noise sources," *IRE Trans. on Instrumentation*, vol. I-9, pp. 209–213, September 1960.

[19] L. Liedquist, "Absolutkalibrering av bruskällor inom Mikrovågsområdet" (Absolute calibration of noise sources in the microwave field), *Elteknik*, vol. 9, pt. 3, pp. 43–47, March 1966.

[20] A. C. Gordon-Smith and J. A. Lane, "Standardization of coaxial noise sources," *Electronic Letts.*, vol. 1, pp. 7–8, March 1965.

[21] R. A. Andrews, Jr., "Calibration of gas-discharge noise sources in an industrial standards laboratory," presented at the 20th Ann. ISA Conf., Los Angeles, Calif., October 1965, Preprint 42.1-1-1965.

[22] G. J. Halford, "Noise comparators and standards for *S* and *X* bands," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-15, pp. 310–317, December 1966.

[23] J. B. Johnson, "Thermal agitation of electricity in conductors," *Phys. Rev.*, vol. 32, p. 97, July 1928.

[24] H. Nyquist, "Thermal agitation of electric charge in conductors," *Phys. Rev.*, vol. 32, p. 110, July 1928.

[25] A. I. Reynard, "Precision instruments for calibrating radiometers at 4.3 millimeters wavelength," U. S. Naval Research Lab., Rept. 4927, May 1957.

[26a] W. Jasinski and G. Hiller, "Determination of noise temperature of a gas-discharge noise source for four-millimeter waves," *Proc. IRE* (*Correspondence*), vol. 49, pp. 807–808, April 1961.

[26b] Q. V. Davis, "A high temperature termination for use at short millimeter wavelengths," *J. Sci. Instr.*, vol. 40, pp. 524–525, November 1963.

[27] J. E. Sees, "Fundamentals in noise source calibrations at microwave frequencies," U. S. Naval Research Lab., Rept. 5051, January 1958.

[28a] "Calibration of microwave noise sources," *NBS Tech. News Bull.*, vol. 47, pp. 31–34, February 1963.

[28b] "Calibration of waveguide noise sources," *Federal Register*, vol. 28, p. 7639, July 26, 1963.

[29] L. A. Birger and I. A. Sokov, "Reference thermal noise generators: high and ultra-high frequency measurements," *Izmeritel. Tekhn.*, pp. 47–50, January 1962. Translation in *Meas. Tech* (*USSR*), 1962.

[30] H. Zucker, G. Baskin, S. I. Cohn, J. Lerner, and A. Rosenblum, "Design and development of a standard white noise generator and noise indicating instrument," *IRE Trans. on Instrumentation*, vol. I-7, pp. 279–291, December 1958.

[31] Yu. S. Rusinov and R. L. Sorochenko, "Primary standard of noise radiation in the decimeter band," *Pribory Tekh. Eksper.* (*USSR*), pp. 121–122, May–June 1964. Translation in *Instrum. Exper. Tech.* (*USA*), pp. 612–613, May–June 1964.

[32] C. T. Stelzried, "A liquid-helium-cooled coaxial termination," *Proc. IRE* (*Correspondence*), vol. 49, p. 1224, July 1961.

[33] P. G. Mezger and H. Rother, "Kühlbare Eichwiderstände als Ranschtemparaturnormale" (Coolable calibrating resistance as noise temperature standards), *Frequenz*, vol. 16, no. 10, pp. 386–391, 1962.

[34] K. M. Eisele, "Refrigerated microwave noise sources," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-13, pp. 336–342, December 1964.

[35] C. T. Stelzried, "Temperature calibration of microwave thermal noise sources," *IEEE Trans. on Microwave Theory and Techniques* (*Correspondence*), vol. MTT-13, pp. 128–130, January 1965.

[36] A. A. Penzias, "Helium-cooled reference noise source in a 4-kMc waveguide," *Rev. Sci. Instr.*, vol. 36, pp. 68–70, January 1965.

[37] A. Jurkus, "Cold loads as standard noise sources," *Proc. IEEE* (*Correspondence*), vol. 53, pp. 176–177, February 1965.

[38] R. C. Menon and N. P. Albaugh, "Cooled loads as calibration noise standards for the mm-wavelength range," *Proc. IEEE* (*Correspondence*), vol. 54, pp. 1501–1503, October 1966.

[39] E. Maxwell and B. J. Leon, "Absolute measurement of receiver noise figures at UHF," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-4, pp. 81–85, April 1956.

[40] H. Prinzler, "A saturated diode noise source for 20 cm wavelengths and its absolute calibration with a heated resistance," *Nachrichtentechnik*, vol. 8, pp. 495–500, November 1958.

[41] W. Schottky, "Uber Spontane Stromschwankungen in Verschiedenen Elektrizitatsleitern," *Ann. Phys.* (*Leipzig*), vol. 57, pp. 541–567, December 1918.

[42] P. L. Chessin, "A bibliography of noise," *IRE Trans. on Information Theory*, vol. IT-1, pp. 15–31, September 1955.

[43] P. K. Tien and J. Moshman, "Monte Carlo calculations of the noise near the potential minimum of a HF diode," *J. Appl. Phys.*, vol. 27, pp. 1067–1078, September 1956.

[44] H. Kosmahl, "Correlation factors for noise fluctuations at the potential minimum of a diode (triode)," *Arch. Elekt. Übertragung*, vol. 10, pp. 353–357, August 1956.

[45a] R. W. Slinkman, "Temperature-limited noise diode design," *Sylvania Technol.*, vol. 2, pp. 6–8, October 1949.

[45b] H. Johnson, "A coaxial-line diode noise source for UHF," *RCA Rev.*, vol. 8, pp. 169–185, March 1947.

[45c] H. Groendijk, "A noise diode for ultra-high frequencies," *Philips Tech. Rev.*, vol. 20, pp. 108–110, 1958–1959.

[45d] R. Kompfner et al., "The transmission-line diode as a noise source at centimeter wavelengths," *J. IEE* (*London*), vol. 93, pt. IIIA, p. 1436, 1946.

[46] I. A. Harris, "The design of a noise generator for measurements in the frequency range 30–1250 Mc/s," *Proc. IEE* (*London*), vol. 108B, November 1961.

[47] W. W. Mumford, "A broad-band microwave noise source," *Bell Sys. Tech. J.*, vol. 28, pp. 608–618, October 1949.

[48] E. L. Chinnock, "A portable, direct-reading microwave noise generator," *Proc. IRE*, vol. 40, pp. 160–164, February 1952.

[49] W. W. Mumford and R. L. Schafersman, "Data on the temperature dependence of *X*-band fluorescent lamp noise sources," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-3, pp. 12–17, December 1955.

[50] P. Parzen and L. Goldstein, "Current fluctuations in the direct-

current gas-discharge plasma," *Phys. Rev.*, vol. 82, pp. 724–726, June 1951.
[51] K. S. Knol, "Determination of the electron temperature in gas discharges by noise measurements," *Philips Research Repts.*, vol. 6, pp. 288–302, August 1951.
[52] T. J. Bridges, "A gas-discharge noise source for eight-millimeter waves," *Proc. IRE*, vol. 42, pp. 818–819, May 1954.
[53] H. Johnson and K. R. Deremer, "Gaseous discharge super-high-frequency noise sources," *Proc. IRE*, vol. 39, pp. 908–914, August 1951; also, "Super-high-frequency noise source," US Signal Corps, Project 322C-1, Final Rept., June 1, 1947–March 31, 1951.
[54] H. Prinzler, "Investigations of noise generators in the microwave region," *Acta Tech. Hungar.*, vol. 42, nos. 1–3, pp. 283–292, 1963.
[55] N. Houlding and L. C. Miller, "Discharge tube noise sources," TRE Memo 593, October 1953.
[56] R. Saier, "Rauschgeneratoren für Zentimeter-Wellen," *Frequenz*, vol. 14, pp. 68–70, February 1960.
[57] P. A. H. Hart and G. H. Plantinga, "An experimental noise generator for millimetre waves," *Philips Tech. Rev.*, vol. 22, no. 12, pp. 391–392, 1960–1961.
[58] C. K. S. Miller, W. C. Daywitt, and E. Campbell, "A waveguide noise-tube mount for use as an interlaboratory standard," *Acta IMEKO III*, pp. 371–381, 1964.
[59] W. H. Spencer and P. D. Strum, "Broadband UHF and VHF noise generators," *IRE Trans. on Instrumentation*, vol. PGI-4, pp. 47–50, October 1955.
[60] M. Kollanyi, "Application of gas-discharge tubes as noise sources in the 1700–2300 Mc/s band," *J. Brit. IRE*, vol. 18, pp. 541–548, September 1958.
[61] H. Schittger and D. Weber, "Über einen Gasentladungs-Rauschgenerator mit Verzögerungsleitung," *Nachr.-Techn. Fachber.*, vol. 2, pp. 118–120, 1955.
[62] A. D. Kuz'min and A. N. Khvoshchev, "A broadband noise generator for the decimeter region," *Radiotekhnika*, vol. 13, no. 7, pp. 36–42, 1958.
[63] H. Montague, "Coaxial UHF noise source," U. S. Naval Research Lab., Rept. 4560, August 1955.
[64] G. Alma'ssy and I. Frigyes, "New microwave noise generator for the 2000 Mc/s band," *Periodica Polytech. Elect. Engrg. (Hungary)*, vol. 4, no. 4, pp. 293–303, 1960.
[65] R. I. Skinner, "Wide-band noise sources using cylindrical gas-discharge tubes in two-conductor lines," *Proc. IEE (London)*, vol. 103B, pp. 491–496, July 1956.
[66] R. A. Lee and K. W. Olsen, "Absolute values of excess noise ratio traceable to the Bureau of Standards," Bendix, Red Bank, N. J., Engrg. Data Release, Issue 40, File G-19, p. 12, May 1962.
[67] D. O. North, "The absolute sensitivity of radio receivers," *RCA Rev.*, vol. 6, pp. 332–343, January 1942.
[68] H. T. Friis, "Noise figures of radio receivers," *Proc. IRE*, vol. 32, pp. 419–422, July 1944.
[69] "IRE Standards on Electron Tubes: Definition of Terms 1962 (62 IRE 7.S2)," *Proc. IEEE*, vol. 51, pp. 434–435, March 1963.
[70] "Description of the noise performance of amplifiers and receiving systems," *Proc. IEEE*, vol. 51, pp. 436–442, March 1963.
[71] "Representation of noise in linear two-ports," *Proc. IRE*, vol. 48, pp. 69–74, January 1960.
[72] "IRE Standards on Methods of Measuring Noise in Linear Two-ports, 1959," *Proc. IRE*, vol. 48, pp. 60–68, January 1960.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967
pp. 865-877
THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 66C, No. 4, October–December 1962

A Precision Noise Spectral Density Comparator

C. M. Allred

(May 18, 1962)

The theory is given for a precision comparator that measures the ratio of two noise spectral densities. The relative error of a single measurement is also derived. The comparator described removes or alleviates many of the problems in high-speed switching, and since the instrument operates under null conditions, the null position is essentially independent of amplifier noise and gain instabilities.

1. Introduction

The relative strength of two power sources or signal levels is often desired. As the strength of the two power sources decreases the determination of this ratio becomes increasingly difficult. When the desired signals have levels that are approximately equal or smaller than the internal noise level of a very low noise amplifier, the problems become severe. This is due not only to the masking effect of the internal amplifier noise but also to the instability of the amplifier gain. This in turn arises from the enormous amounts of gain needed.

The usual practice is to sample the outputs of the two levels at a switching rate that is high compared to the instability drift rate of the amplifier. An attenuator is usually used to reduce the level of the stronger signal to that of the other signal. The early pioneer of this method was the Dicke Radiometer [Dicke, 1946].

The switch with its noise, losses, and instabilities is often a limiting factor in such methods. The attenuator and the necessity of impedance matching is another problem source. These problems are eliminated or greatly alleviated by the method described herein.

2. Principle of Operation

The basic problem in such measurements is to make the amplifier gain drift essentially unimportant. In principle, the usual radiometer samples the two sources so that the variation of amplifier gain affects the amplified source levels of both sources equally. A preferable method would be to have both sources simultaneously present and no switching. The problem then, is to separate the two amplified sources at the output. If the two sources are statistically independent, they could be separated by correlation techniques. In principle this is what is done in the system described below.

Correlation techniques are an old tool and there are systems described in the literature that in various respects resemble the one discussed here [Fink, 1959; Freeman, 1958, p. 274].

The block diagram of the system is shown in figure 1. The individual components and also the configuration can take on different forms without changing the basic principles. The system described here is similar to one presently under construction.

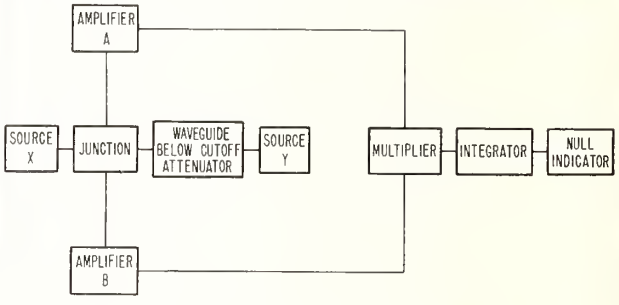


FIGURE 1. *Block diagram of comparator.*

2.1. Sources

The two sources may be very general. They may consist of random noise, pulse singals, CW, etc., or combinations of various types. It is assumed their output levels are constant during the measuring time.

2.2. Attenuator

For precison results the attenuator may be an accurate wave guide-below-cutoff (piston) attenuator. Small changes in attenuation may also be determined by measuring changes in the input level to the attenuator.

2.3. Junction

This simple but important component is shown in figure 2. It is a low-loss symmetrical coaxial T. Opposite to where the center leg joins the T, very small slits, perpendicular to the axis of the two side arms, are cut into the outer conductor. These are symmetrically placed on the T. The junction then is made an integral part of the attenuator so that the fields (TE_{11} mode) within the attenuator guide impinge upon the slits of the T. In effect, the T acts as a center-tapped secondary of a transformer. The source X is applied to the center tap.

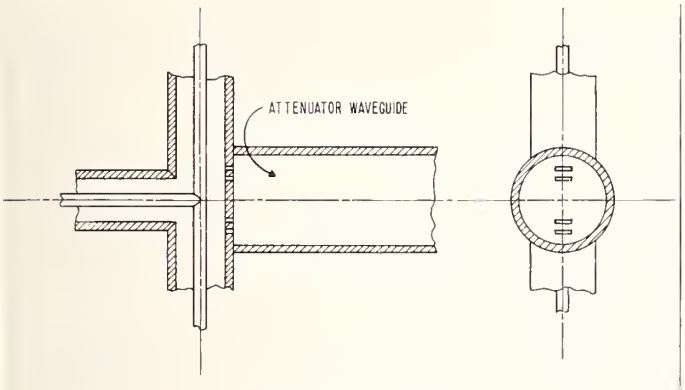


FIGURE 2. *Junction.*

2.4. Amplifiers

There is some value in having the two amplifiers similar, though this is not necessary. The amplifiers should be linear; at least the frequency components produced by the nonlinearity should be prevented from reaching the input to the multiplier. For some modes of system operation it is essential for the pass band characteristics of the amplifiers to remain constant during a measurement while the gain can vary. This is what primarily happens in an amplifier as the mutual conductance of a vacuum tube varies. The stability of the pass band can be enhanced by keeping the amplifier broadbanded except for stable band limiting devices.

2.5. Multiplier, Filter, and Indicator

The multiplier has the function of forming the product of the input voltages. The output of the multiplier, which is this product, is applied to the filter. The filter permits only the d-c and very low frequency components of this product to reach the indicator. Of course the filter and multiplier could be a single integral unit. The indicator responds to both positive and negative d-c voltages and is used only to obtain a null.

3. General Theory

Only the basic general theory will be presented. This describes, under certain assumptions, what the system measures and the expected error (due to the inevitable random fluctuations in the output) in this measurement. Possible errors due to various components deviating from the assumptions will be reserved for a forthcoming report when a specific system is being described.

The assumptions used in the following theory are:

1. The system, excluding the multiplier, is considered linear.

2. The amplifiers and filter, while linear, may be considered varying in time. The unit impulse response and Fourier system response functions of these units are then time variable. The analysis could have also been carried through with the assumption that the unit impulse functions were sample functions from a linearly independent ergodic process with the essential results being the same.

3. The signals present are all assumed to be sample functions from random processes that are bounded, and are singly and jointly ergodic as well as singly and jointly stationary. The periodic signals are considered to occur with a uniform random phase distribution over the ensemble.

4. The components, other than what is stated or implied above, are considered to be ideal. For example, the multiplier performs the proper product. Also, the junction is symmetrical, etc.

Sources X and Y respectively give rise to random signals $x(t)$ and $y(t)$ that are real time varying functions and contain, in general, both periodic and nonperiodic parts. The input to amplifier A is

$$x(t)+y(t). \tag{1}$$

To amplifier B, it is

$$x(t)-y(t). \tag{2}$$

The output of amplifier A, $e_a(t)$, is

$$e_a(t)=\int_{-\infty}^{\infty} h_a(\alpha, t)[x(t-\alpha)+y(t-\alpha)+z_a(t-\alpha)]d\alpha \tag{3}$$

where $h_a(\alpha, t)$ is the unit impulse response of amplifier A, $z_a(t)$ is the internal noise of amplifier A referred to its input, and α is the variable of integration in the convolution integral. Likewise the output of amplifier B, $e_b(t)$, is

$$e_b(t)=\int_{-\infty}^{\infty} h_b(\beta, t)[x(t-\beta)-y(t-\beta)+z_b(t-\beta)]d\beta \tag{4}$$

where the b subscripts refer to amplifier B. The quantities $e_a(t)$ and $e_b(t)$ are multiplied together by means of the multiplier. The expected value or statistical average of the output of the multiplier, indicated by $\mathrm{E}(e_a e_b)$ or $\overline{e_a e_b}$ is

$$E(e_a e_b)=E\left[\left(\int_{-\infty}^{\infty} h_a(\alpha, t)[x(t-\alpha)+y(t-\alpha)+z_a(t-\alpha)]d\alpha\right)\left(\int_{-\infty}^{\infty} h_b(\beta, t)[x(t-\beta)-y(t-\beta)+z_b(t-\beta)]d\beta\right)\right].$$

Since $h_a(\alpha,t)$ and $h_b(\beta,t)$ are considered to be nonrandom functions,

$$E(e_a e_b)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_a(\alpha, t)h_b(\beta, t)E([x(t-\alpha)+y(t-\alpha)+z_a(t-\alpha)][x(t-\beta)-y(t-\beta)+z_b(t-\beta)])d\alpha d\beta \tag{5}$$

where the order of integration and averaging have

been interchanged. The expected value or statistical mean of the indicator output, $\bar{I}$, is given by

$$\bar{I}=H_d(0,t)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h_a(\alpha,t)h_b(\beta,t)\,E([x(t-\alpha)+y(t-\alpha)+z_a(t-\alpha)][x(t-\beta)-y(t-\beta)+z_b(t-\beta)])d\alpha d\beta \quad (6)$$

where $H_d(0,t)$ is the zero frequency response of the filter and indicator.

The expected value of the time varying signal function in (6) is composed of the statistical autocorrelation and cross correlation functions of the various random signals present. Thus

$$E[x(t-\alpha)x(t-\beta)]\equiv R_x(\beta-\alpha) \quad (7a)$$

where $R_x(\beta-\alpha)$ is the statistical autocorrelation function of $x(t)$. Similarly

$$E[x(t-\alpha)y(t-\beta)]\equiv R_{xy}(\beta-\alpha) \quad (7b)$$

with similar results for other combinations, where $R_{xy}(\beta-\alpha)$ is the cross correlation function of the functions $x(t)$ and $y(t)$.

From the foregoing, (6) may be written

$$\bar{I}=H_d(0,t)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h_a(\alpha,t)h_b(\beta,t)[R_x(\beta-\alpha)-R_y(\beta-\alpha)-R_{xy}(\beta-\alpha)+R_{xz_b}(\beta-\alpha)+R_{yx}(\beta-\alpha)+R_{yz_b}(\beta-\alpha)+R_{z_ax}(\beta-\alpha)-R_{z_ay}(\beta-\alpha)+R_{z_az_b}(\beta-\alpha)]d\alpha d\beta. \quad (8)$$

If the respective signals are uncorrelated all cross correlation functions become zero, and (8) reduces to (correlation between $z_a(t)$ and $z_b(t)$ is discussed later)

$$\bar{I}=H_d(0,t)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h_a(\alpha,t)h_b(\beta,t)\,[R_x(\beta-\alpha)-R_y(\beta-\alpha)]d\alpha d\beta. \quad (9)$$

The above correlation functions are related to the spectral densities of the various signals. Let $S_x(f)$ represent the spectral density of the signal $x(t)$. Then $S_x(f)$ and $R_x(\tau)$ are related (by definition) through the Fourier transform pair

$$S_x(f)=\int_{-\infty}^{\infty}R_x(\tau)e^{-j2\pi f\tau}d\tau \quad (10a)$$

$$R_x(\tau)=\int_{-\infty}^{\infty}S_x(f)\,e^{j2\pi f\tau}df \quad (10b)$$

with similar relations for the y signal.

Substituting the above values for the correlation functions into (9) together with the relation $\tau=\beta-\alpha$ yields

$$\bar{I}=H_d(0,t)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h_a(\alpha,t)\,e^{-j2\pi f\alpha}h_b(\beta,t)e^{j2\pi f\beta}\,[S_x(f)-S_y(f)]df d\alpha d\beta. \quad (11)$$

Now the impulse response function, $h_a(\alpha,t)$, is related to the complex frequency response function, $H_a(j2\pi f,t)$ by

$$\int_{-\infty}^{\infty}h_a(\alpha,t)e^{-j2\pi f\alpha}d\alpha=H_a(j2\pi f,t)=|H_a(j2\pi f,t)|e^{j\theta_a} \quad (12)$$

where $H_a(j2\pi f,t)$ is the complex frequency gain function of amplifier A and θ_a is the associated phase angle. Similarly

$$\int_{-\infty}^{\infty}h_b(\beta,t)e^{j2\pi f\beta}d\beta=H_b^*(j2\pi f,t)=|H_b(j2\pi f,t)|e^{-j\theta_b}. \quad (13)$$

In general, θ_a and θ_b are functions of time, though for simplicity in notation they are not explicitly written as such. Thus (11) becomes

$$\bar{I}=H_d(0,t)\int_{-\infty}^{\infty}|H_a(j2\pi f,t)||H_b(j2\pi f,t)|[\cos(\theta_a-\theta_b)+j\sin(\theta_a-\theta_b)]\,[S_x(f)-S_y(f)]df. \quad (14)$$

It may be shown that $H(j2\pi f,t)$ has an imaginary part that is an odd function of frequency.

Using this fact, (14) becomes

$$\bar{I}=H_d(0,t)\int_{-\infty}^{\infty}|H_a(j2\pi f,t)||H_b(j2\pi f,t)|\,[\cos(\theta_a-\theta_b)]\,[S_x(f)-S_y(f)]\,df. \quad (15)$$

Periodic signals such as CW signals or even periodic wide-sense random processes have power spectral densities expressed by means of the Dirac delta function, δ [Davenport, 1958; ch. 6]. For example, a CW signal, $V\cos(\omega_0 t+\theta)$ has a power spectral density

$$S_{cw}(f)=\frac{V^2}{4}\,\delta(f-f_0)+\frac{V^2}{4}\,\delta(f+f_0). \quad (16)$$

The total power, therefore, is

$$\int_{-\infty}^{\infty}S_{cw}(f)df=\frac{V^2}{2} \quad (17)$$

as is well known.

The output, $\overline{I}$, would become zero (a null would exist) when

$$\int_{-\infty}^{\infty} |H_a(j2\pi f,t)||H_b(j2\pi f,t)|\cos(\theta_a-\theta_b)S_x(f)\,df$$

$$=\int_{-\infty}^{\infty} |H_a(j2\pi f,t)||H_b(j2\pi f,t)|\cos(\theta_a-\theta_b)S_y(f)\,df. \tag{18}$$

It is of considerable practical importance to consider those cases for which a null is obtained independent of the variations in the amplifiers and filter response functions. Since (18) contains no filter response function, the null is independent of the filter characteristics. From (18) it is also obvious that if the spectral densities of the $x(t)$ and $y(t)$ signals were equal over the amplifier pass bands, that is, if

$$S_x(f)=S_y(f), \tag{19}$$

a null is then obtained regardless of the time varying aspects of the amplifiers. Thus the important result is obtained that a drift or variation in the gain or phase response of the amplifiers and filter would not influence the null condition.

Further, if the amplifiers were constructed so that the frequency response determining components were stable, the amplifier characteristics could be written

$$|H_a(j2\pi f,t)||H_b(j2\pi f,t)|\cos(\theta_a-\theta_b)$$

$$=G_a(t)G_b(t)|H'_a(j2\pi f)||H'_b(j2\pi f)|\cos(\theta_a-\theta_b) \tag{20}$$

where $G_a(t)$ and $G_b(t)$ represent the time variable or unstable parts of the gains of the respective amplifiers and the primed H's are new time-stable frequency-dependent parts of the amplifier gains. The θ's are now also independent of time.

The output of the system may then be written, since $G_a(t)G_b(t)$ is independent of f over the bandpass,

$$\overline{I}=H_a(0,t)G_a(t)G_b(t)\int_{-\infty}^{\infty} |H'_a(j2\pi f)||H'_b(j2\pi f)|$$

$$\cos(\theta_a-\theta_b)[S_x(f)-S_y(f)]df. \tag{21}$$

This becomes zero when

$$\int_{-\infty}^{\infty} |H'_a(j2\pi f)||H'_b(j2\pi f)|\cos(\theta_a-\theta_b)S_x(f)df$$

$$=\int_{-\infty}^{\infty} |H'_a(j2\pi f)||H'_b(j2\pi f)|\cos(\theta_a-\theta_b)S_y(f)df \tag{22}$$

and is independent of the amplifier drift or variation in gain. Such amplifiers are not too hard to realize by making the unstable frequency-dependent parts with pass bands broad compared to the stable frequency-dependent parts.

With such amplifiers a time-stable null condition expressed by (22) may be met with signals having spectral densities with widely different frequency characteristics. For example $y(t)$ may be a sinusoidal signal

$$y(t)=Y_0\cos(\omega_0 t+\theta). \tag{23}$$

Then (22) becomes

$$\int_{-\infty}^{\infty} |H'_a(j2\pi f)||H'_b(j2\pi f)|\cos(\theta_a-\theta_b)S_x(f)df$$

$$=\int_{-\infty}^{\infty} [|H'_a(j2\pi f)||H'_b(j2\pi f)|\cos(\theta_a-\theta_b)]$$

$$\left[\frac{Y_0^2}{4}\delta(f-f_0)+\frac{Y_0^2}{4}\delta(f+f_0)\right]df$$

$$=|H'_a(j2\pi f_0)||H'_b(j2\pi f_0)|\cos[\theta_a(f_0)-\theta_b(f_0)]\frac{Y_0^2}{2}. \tag{24}$$

This is again independent of the amplifier gain variation.

4. Measurement of Spectral Densities

From the foregoing relationships, spectral densities of one source may be obtained in terms of that of another source.

Assume, for example, the value $S_y(f)$ is known and is also known to be essentially constant over the pass band, and it is desired to obtain the value of $S_x(f)$ which is also known to be constant over the pass band.

With reference to (18), a null condition is obtained when $S_x(f)=S_y(f)$. The two spectral densities are thus equated.

Let two different $x(t)$ signals be measured. Then

$$S_{x1}(f)=S_{y1}(f) \tag{25}$$

$$S_{x2}(f)=S_{y2}(f). \tag{26}$$

The ratio

$$\frac{S_{y2}(f)}{S_{y1}(f)}=A \tag{27}$$

may be measured by means of the precision attenuator (see fig. 1), and hence

$$S_{x2}(f)=AS_{x1}(f). \tag{28}$$

In this case $S_{x1}(f)$ and $S_{x2}(f)$ need not be equal and only relative values of $S_y(f)$ are needed. The relative values of $S_y(f)$ may be obtained by means of the attenuator.

A CW signal could just as well be used in place of $S_y(f)$. Then

$$S_{x2}(f)=\frac{V_2^2}{V_1^2}S_{x1}(f)=AS_{x1}(f) \tag{29}$$

where

$$A=\frac{V_2^2}{V_1^2}$$

and is again measured by the attenuator.

If $S_{x2}(f)$ and $S_{x1}(f)$ vary differently with frequency, the relationship between them may be obtained by using (22) to relate $S_{x1}(f)$ with $S_{y1}(f)$ and $S_{x2}(f)$ with $S_{y2}(f)$. Noting also that $S_{y2}(f)=AS_{y1}(f)$, the desired relationship becomes

$$\int_{-\infty}^{\infty}|H_a'(j2\pi f)||H_b'(j2\pi f)|\cos(\theta_a-\theta_b)S_{x2}(f)df$$
$$=A\int_{-\infty}^{\infty}|H_a'(j2\pi f)||H_b'(j2\pi f)|\cos(\theta_a-\theta_b)S_{x1}(f)df. \quad (30)$$

If the spectral density of $S_{x1}(f)$ is constant in value over the pass band, (30) becomes

$$\frac{\int_{-\infty}^{\infty}|H_a'(j2\pi f)||H_b'(j2\pi f)|\cos(\theta_a-\theta_b)S_{x2}(f)df}{\int_{-\infty}^{\infty}|H_a'(j2\pi f)||H_b'(j2\pi f)|\cos(\theta_a-\theta_b)df}$$
$$=AS_{x1}(f). \quad (31)$$

In practice, this is what is usually measured. This gives the equivalent constant spectral density that would yield the same total power as the actual spectral density function when both are integrated over the actual pass band response function. This means the relative pass band response function of the actual system used should be given when stating the equivalent $S_x(f)$.

A system that meets the conditions imposed by (22) lends itself very well to a measurement of the relative pass band response function,

$$K|H_a'(j2\pi f)||H_b'(j2\pi f)|\cos(\theta_a-\theta_b).$$

Here K is a constant and represents the fact that only relative and not absolute values are needed as a function of frequency. The above function is obtained by using CW signals for both the $x(t)$ and $y(t)$ signals. The $y(t)$ signal may be varied in level to maintain a null as $x(t)$ is varied in frequency but kept at a constant level. The relative values of the attenuator setting as a function of frequency give the desired relationship.

Other ways of using the system may readily come to mind. One other will be briefly mentioned. Another uncorrelated signal $v(t)$ with spectral density $S_v(f)$ may be combined with the $x(t)$ signal. The relative amounts of this extra signal may be measured by an attenuator external to the $x(t)$ signal in much the same way as is done with the $y(t)$ signal. In this case, the $y(t)$ signal need not be varied and, (19) could be written

$$S_y(f)=S_{x1}(f)+S_{v1}(f)$$
$$=S_{x2}(f)+S_{v2}(f)=S_{x3}(f)+S_{v3}(f). \quad (32)$$

If only relative values of $S_v(f)$ are known a useful relation is

$$\frac{S_{x3}(f)}{S_{x1}(f)}=1+\left(\frac{S_{x2}}{S_{x1}}-1\right)\left(\frac{A_{31}-1}{A_{21}-1}\right) \quad (33)$$

where

$$A_{31}=\frac{S_{v3}}{S_{v1}}$$

$$A_{21}=\frac{S_{v2}}{S_{v1}}.$$

4.1. Broad Spectral Signals Versus CW Signals

As indicated, either broad spectral signals or a CW signal may be used as the $y(t)$ signal to compare various random $x(t)$ signals. When CW signals are used, it is important that the frequency of the CW signal be constant. Also, practical usage dictates finite averaging times at the output. This means that a CW $y(t)$ signal would produce less random fluctuation than a random $y(t)$ signal.

Broad spectral $y(t)$ signals, however, place less stringent requirements on the constancy of the band pass characteristics, especially if their spectral densities are similar to those of $x(t)$. A broad spectral nonrandom signal would produce less random fluctuation than a random signal and still have those advantages of broad spectral signals.

4.2. Residual Noise

If the signals $z_a(t)$ and $z_b(t)$ are correlated an additional term

$$|H_a(j2\pi f,t)||H_b(j2\pi f,t)||S_{z_az_b}(f)|\cos(\theta_a-\theta_b+\phi_{z_az_b})$$

would have occurred. Here, $\phi_{z_az_b}$ is the relative phase of the correlated part of the amplifier noise signal in (15). This represents a contribution in the output due to correlation between the two amplifier noise signals. While there is usually very little correlation between the two amplifier noise signals there can be contributions by the two amplifier noises that have common origins and are therefore correlated. Whether the effect of this correlated residual noise can be ignored depends on the relative strength of signals being measured to the residual noise and the desired accuracy of measurement.

This residual noise may possibly be reduced by proper design of the amplifier, or by using special networks that prevent coupling or that cause shifts in the phase of the correlated parts of the residual noise such that the quantity $\theta_a-\theta_b+\phi_{z_az_b}$ in (17) equals $\pi/2$, thus making the cosine factor zero.

Also, the residual noise may be simply measured by the use of two sources of different known levels. The principle may be simply demonstrated by assuming (this is not a necessary assumption) all signals have constant spectral densities within the amplifier pass bands and the amplifiers have equal phase shifts ($\theta_a=\theta_b$). Using (15) with the assumption that no signals are correlated except $z_a(t)$ and

$z_b(t)$ and that the null condition holds, then

$$S_x(f)-S_y(f)+|S_{z_az_b}(f)|\cos\phi_{z_az_b}=0. \quad (34)$$

For two different levels of $S_x(f)$, this yields

$$|S_{z_az_b}(f)|\cos\phi_{z_az_b}=\frac{AS_{x1}-S_{x2}}{1-A} \quad (35)$$

where

$$A=\frac{S_{y2}(f)}{S_{y1}(f)}.$$

5. Relative Error of a Single Measurement

Any practical system will average the output over a finite time. Because this time is finite, random fluctuations will appear in the output giving rise to errors when the output indicator is read.

For simplicity in evaluating this error, the following assumptions in addition to those above (with one exception) will be made.

1. Each amplifier has a square band pass of width B.

2. The random processes giving rise to $x(t)$, $z_a(t)$, and $z_b(t)$ are considered to be Gaussian, and the respective signals are independent and thus uncorrelated. Also, the various signal sources have zero means.

3. The signal $x(t)$ has a spectral density $S_x(f)$ that is constant over the band.

4. The signal, $y(t)=Y_0\cos(\omega_0 t)$, is now a deterministic (nonrandom) CW signal centered in the pass band of the amplifier. The former theory still applies. The present system merely represents sample functions of the ensemble with time origins fixed relative to the CW signal.

5. The two amplifier noise signals, $z_a(t)$ and $z_b(t)$, are noncorrelated. Their respective spectral densities $S_{z_a}(f)$ and $S_{z_b}(f)$ are constant over the band.

6. The exception to the former assumptions is that the response functions of the amplifiers and filter are constant with time.

It is quite common to represent a narrow band of noise of band width B around a central frequency by [Davenport, p. 158]

$$X(t)=X_c(t)\cos\omega_0 t-X_s(t)\sin\omega_0 t$$
$$=R(t)\cos[\omega_0 t-\phi(t)] \quad (36)$$

where $X_c(t)$, $X_s(t)$, $R(t)$, and $\phi(t)$ are random fluctuations that are slowly varying (due to the narrow band) with respect to the central frequency, $f_0=\omega_0/2\pi$. Similarly

$$z_a(t)=Z_{ac}(t)\cos\omega_0 t-Z_{as}(t)\sin\omega_0 t \quad (37)$$

$$z_b(t)=Z_{bc}(t)\cos\omega_0 t-Z_{bs}(t)\sin\omega_0 t. \quad (38)$$

It can be shown [Davenport, p. 158] that

$$\overline{X_c^2}=\overline{X_s^2}=2\int_0^\infty S_x(f)df=2S_x(0)B \quad (39)$$

$$\overline{Z_{ac}^2}=\overline{Z_{as}^2}=2S_{z_a}(0)B \quad (40)$$

$$\overline{Z_{bc}^2}=\overline{Z_{bs}^2}=2S_{z_b}(0)B \quad (41)$$

$$\overline{X_cX_s}=\overline{Z_{ac}Z_{as}}=\overline{Z_{bc}Z_{bs}}=0 \quad (42)$$

where $S_x(0)$ denotes the constancy of S_x with frequency.

Also, it can be shown [van der Ziel, 1954] that

$$\overline{X_c^2(t)X_c^2(t+\tau)}=2R_{xc}^2(\tau)+\overline{(X_c^2)}^2 \quad (43)$$

where

$$R_{xc}(\tau)=\overline{X_c(t)X_c(t+\tau)}=\int_{-\infty}^{\infty}S_x(f)e^{j(\omega-\omega_0)\tau}df$$
$$=2S_x(0)B\frac{\sin\pi B\tau}{\pi B\tau} \quad (44)$$

since $S_x(0)$ is constant. Similar relationships hold for $X_s(t)$, $Z_c(t)$, and $Z_s(t)$.

It will also be assumed that the two amplifier noise signals are of equal strength. Hence

$$S_{z_a}(0)=S_{z_b}(0)\equiv S_z(0). \quad (45)$$

The inputs to the two amplifiers will be

Input a

$$e_a(t)=[X_c(t)+Y_0+Z_{ac}(t)]\cos\omega_0 t$$
$$-[X_s(t)+Z_{as}(t)]\sin\omega_0 t \quad (46)$$

Input b

$$e_b(t)=[X_c(t)-Y_0+Z_{bc}(t)]\cos\omega_0 t$$
$$-[X_s(t)+Z_{bs}(t)]\sin\omega_0 t. \quad (47)$$

For mathematical convenience, the filter is considered to consist of two parts; (1) a zonal filter (i.e., a filter that has a system function that is unity over the pass band and zero elsewhere) that effectively integrates [Freeman, p. 226] the output of the multiplier over a period long compared to that of the central frequency f_0, and (2) a filter that is effective at much lower frequencies.

The output of the multiplier, M, is effectively time averaged over a cycle of the center frequency, ω_0, by the zonal filter. Thus

$$M(t)=\langle e_a(t)e_b(t)\rangle=\frac{1}{2}\{[X_c(t)+Y_0+Z_{ac}(t)]$$
$$[X_c(t)-Y_0+Z_{bc}(t)]$$
$$+[X_s(t)+Z_{as}(t)][X_s(t)+Z_{bs}(t)]\} \quad (48)$$

where $\langle e_a(t)e_b(t)\rangle$ represents the temporal average.

To obtain the spectral density of M, its auto-

correlation function is needed. Now

$$R_M(\tau)=\overline{M(t)M(t+\tau)}. \tag{49}$$

If the value of $M(t)$ and $M(t+\tau)$ are used as given in (48) together with other relationships expressed above, it may be shown that

$$R_M(\tau)=\frac{1}{4}\left\{\frac{B^2\sin^2\pi B\tau}{(\pi B\tau)^2}\,[16S_x^2(0)+16S_x(0)S_z(0)+8S_z^2(0)]+4Y_0^2S_z(0)B\,\frac{\sin\pi B\tau}{\pi B\tau}+16S_x^2(0)B^2-8Y_0^2S_x(0)B+Y_0^4\right\}. \tag{50}$$

Since $R_M(\tau)$ is an even function of τ

$$S_M(f)=2\int_0^\infty R_M(\tau)\cos(2\pi f\tau)d\tau. \tag{51}$$

Substituting the value of $R_M(\tau)$ as given by (50) and integrating

$$S_M(f)=\frac{1}{4}[4S_x(0)B-Y_0]^2\delta(f)+\begin{cases} Y_0^2S_z(0)+2[2S_x^2(0)+2S_x(0)S_z^2(0)+S_z^2(0)][B-|f|] & \text{for } 0\le|f|<\frac{B}{2} \\ 2[2S_x^2(0)+2S_x(0)S_z(0)+S_z^2(0)][B-|f|] & \text{for } \frac{B}{2}<|f|<B \\ 0 & \text{for } |f|\ge B. \end{cases} \tag{52}$$

The mean value of the indicator deflection, $\bar{I}$, is given by

$$\bar{I}=\frac{1}{2}H_d(0)[4S_x(0)B-Y_0^2] \tag{53}$$

which may be obtained from (48) by taking the average over the ensemble and multiplying by the d-c response function $H_d(0)$. The same value can be obtained by integrating (52) over the vicinity of zero frequency, taking the square root, and multiplying by $H_d(0)$. At null, of course, $\bar{I}$ is zero which occurs when

$$Y_0^2=4S_x(0)B. \tag{54}$$

In adjusting the output to a null, the fluctuations in the output cause an uncertainty in the value of Y_0 for a null condition. This means that the measured value of Y_0 and consequently the measured value of $S_x(0)$ are random variables. These are designated Y_{0m} and $S_{xm}(0)$ respectively and are related through (54); that is

$$Y_{0m}^2=4S_{xm}(0)B. \tag{55}$$

The relative error, ϵ, in a single measurement is now defined as

$$\epsilon=\frac{\sqrt{(S_{xm}(0)-\overline{S_{xm}(0)})^2}}{\overline{S_{xm}(0)}}=\frac{\sqrt{\overline{\Delta S_{xm}^2}(0)}}{\overline{S_x(0)}} \tag{56}$$

where

$$S_{xm}(0)-\overline{S_{xm}(0)}=\Delta S_x(0)$$

and

$$\overline{S_{xm}(0)}=S_x(0). \tag{57}$$

It is assumed the operator (or servomechanism) in adjusting the attenuator makes an error, ΔY, (which is a random variable) that just compensates for the fluctuation, ΔI, in $I(t)$, where

$$\Delta I=I(t)-\bar{I}$$

or

$$\overline{\Delta I^2}=\overline{I^2(t)}-(\bar{I})^2. \tag{58}$$

This means, using (53), that

$$\Delta I=-\frac{1}{2}H_d(0)[4S_x(0)B-(Y_0+\Delta Y)^2]. \tag{59}$$

Simplifying by the use of (54)

$$\Delta I=H_d(0)Y_0\Delta Y \tag{60}$$

where ΔY^2 has been neglected compared to ΔYY_0. Using (58) it can be shown that

$$\overline{\Delta I^2}=\int_{-\infty}^{\infty}|H_d(j2\pi f)|^2S_M(f)df-(\bar{I})^2. \tag{61}$$

When $H_d(j2\pi f)$ is nonzero except at zero and near-zero frequencies, it follows that essentially

$$\overline{\Delta I^2}=S'_M(0)\int_{-\infty}^{\infty}|H_d(j2\pi f)|^2df. \tag{62}$$

Now, from (52), and using (54)

$$S'_M(0)=4S_x^2(0)B\left[1+2\gamma+\frac{\gamma^2}{2}\right] \tag{63}$$

where

$$\gamma=\frac{S_z(0)}{S_x(0)} \tag{64}$$

and the prime denotes $S_M(0)$ less the term involving the delta function.

The quantity $1+\gamma$ is equivalent to the *operating* noise figure. This would reduce to the *standard* noise figure when $x(t)$ is equivalent to a signal that originates from a resistor at a temperature of 290 °K.

Using various relations as given above, the follow-

ing chain of relations may be obtained

$$\epsilon^2=\frac{\overline{\Delta S_x^2(0)}}{S_x^2(0)}=\frac{\overline{4\Delta Y^2}}{Y_0^2}=\frac{\overline{4\Delta I^2}}{H_d^2(0)Y_0^4}$$

$$=\frac{4}{H_d^2(0)Y_0^4}\left[\int_{-\infty}^{\infty}|H_d(j2\pi f)|^2S_M(f)df-(\bar{I})^2\right]$$

$$=\frac{16S_x^2(0)B}{Y_0^4}\left(1+2\gamma+\frac{\gamma^2}{2}\right)\int_{-\infty}^{\infty}\left|\frac{H_d(j2\pi f)}{H_d(0)}\right|^2df$$

$$=\frac{1}{B}\left(1+2\gamma+\frac{\gamma^2}{2}\right)\int_{-\infty}^{\infty}\left|\frac{H_d(j2\pi f)}{H_d(0)}\right|^2df. \qquad (65)$$

For a simple RC filter of time constant T, the integral is equal to $\frac{1}{2T}$. For a critically damped system the integral would be ½ that given above.

Using the simple RC filter, the equation for ϵ becomes

$$\epsilon=\left(\frac{1}{2}+\gamma+\frac{1}{4}\gamma^2\right)^{1/2}\frac{1}{\sqrt{BT}}. \qquad (66)$$

For example, if
$B=5\times10^3$ c/s
$\gamma=1$
$T=3.5$ sec

then

$\epsilon=10^{-2}$ or 1 percent.

If $\gamma=10$, then T must be increased to 80 sec to obtain the same value of ϵ. If $\gamma=0$, then T need only be 1.0 sec for the same value of ϵ.

6. Conclusions

The system described has several desirable characteristics.

1. The normally used high-speed switch has been eliminated. This removes the problems of switch noise, insertion loss, and instability.
2. The effect of amplifier gain drift has been essentially eliminated.
3. The system is operated under null conditions and the indicator null position is therefore independent of amplifier gain settings.
4. A precision, continuously variable piston attenuator may be used.
5. Signals of widely varying strengths may be compared.
6. The system is capable of measuring its own effective pass band response to a high degree of precision.
7. Problems of impedance variations in a dissipative attenuator have been removed.
8. The system is quite insensitive to "hum" and other such signals that might possibly modulate the normally present signals. This follows since the null condition is independent of a time variable system response. This time variation if produced by "hum" or other signals is what often gives troubles in systems, especially those using linear or envelope detectors and switching frequencies commensurate with the undesirable modulating frequencies.

Some undesirable aspects are:

1. One must have two low noise amplifiers.
2. The signal power is now divided into two channels.
3. A determination of the effect of correlation between the internal amplifiers may be necessary in some situations. This may be done by comparing two standard sources of different levels.
4. An appropriate multiplier is needed. Only the output at zero frequency is of interest and this fact removes many problems. Also, an ideal multiplier is not necessarily needed. The main consideration is that no output at zero frequency should exist when only one input is energized.

The author expresses his gratitude for the valuable discussions and encouragement given by M. C. Selby and M. G. Arthur.

7. References

Davenport, W. B., Jr., and W. L. Root, An Introduction to the Theory of Random Signals and Noise, ch. 6, p. 158 (McGraw-Hill Book Co., Inc., New York, N.Y., 1958).
Dicke, R. H., The measurement of thermal radiation at microwave frequencies, Rev. Sci. Instr. **17,** No. 7, 268 (1946).
Fink, H. J., A new absolute noise thermometer at low temperatures, Can. J. Phys. **37,** 1397 (1959).
Freeman, J. J., Principles of Noise, p. 226; p. 274 (John Wiley & Sons, Inc., New York, N.Y., 1958).
van der Ziel, A., Noise, p. 312 (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1954).

(Paper 66C4–108)

A Precision Noise-Power Comparator

M. G. ARTHUR, C. M. ALLRED, MEMBER, IEEE, AND M. K. CANNON, MEMBER, IEEE

Abstract—This paper describes a prototype noise-power comparator based upon a theory given by Allred. Operating at 3 MHz, it is a null-type instrument, the principal components of which are a hybrid four-port, a stable CW generator, a precision attenuator, a dual-channel amplifier and band-pass filter, and an analog multiplier. Unlike other radiometers, no rapid switching of the noise power or reference voltage is performed.

The instrument can compare noise powers having effective noise temperatures in the range from below liquid-nitrogen temperature to greater than 30,000°K. Two noise generators having known spectral densities are used to calibrate the comparator. The uncertainty of comparison is less than 1 per cent at 75°K and decreases to 0.2 per cent at 29,000°K.

Introduction

THIS PAPER DESCRIBES a prototype noise-power comparator, based upon a theory given by Allred.[1] The comparator, in conjunction with suitable standards of noise power, is used to measure the noise-spectral density in a 7.5-kHz band of frequencies centered at 3 MHz. Its measurement-range capability is greater than 26 dB; it can measure generators having equivalent-noise temperatures from 75°K to 30,000°K, which, for a 7.5-kHz bandwidth, corresponds to available signal powers from −141 dBm to −115 dBm.

At the present stage of development, the uncertainty in a measurement is less than 1 per cent for generators at 75°K, decreasing to 0.2 per cent for generators above 500°K.

The principal advantage of this comparator is that the adverse effects of changes in amplifier gain are essentially eliminated without resorting to the use of high-speed switching.

The manner in which the comparator is used is analogous to that of the familiar uncalibrated potentiometer. A noise generator is connected to the input port of the comparator, and a null indicator is connected to its output port. A precision piston attenuator is adjusted until a null is obtained. A second noise generator is connected in the place of the first noise generator, and the attenuator is readjusted to again obtain a null. The ratio of powers delivered by the two generators to the comparator is obtained from the difference in precision-attenuator settings.

Description of the Comparator

A block diagram of the comparator is shown in Fig. 1. The essential components are:

1) A hybrid four-port.
2) A stable CW generator.
3) A precision attenuator.
4) Two separate channels of amplification and filtering.
5) An analog multiplier.
6) An averaging network.

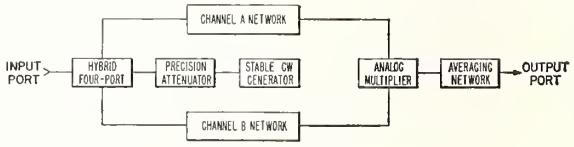


Fig. 1. Block diagram of comparator.

Hybrid Four-Port

The hybrid four-port, shown in Fig. 2, is the coaxial equivalent of a magic tee. A flange is mounted on the cross arm of a coaxial tee opposite the stem of the tee. Several slots are cut into the flange, symmetrically with respect to the tee junction, and the barrel of a precision piston attenuator is attached to the flange. Continuous wave power is coupled into this hybrid four-port via the magnetic field that passes through the slots to the cross-center conductor of the tee.

Precision Attenuator

The precision attenuator is a waveguide-below-cutoff piston attenuator operating with the TE_{11} mode. Figure 3 shows how it is mounted with relation to the hybrid four-port. The piston is translated with a lead screw and carriage mounted on a heavy lathe bed. End-gage blocks and a precision micrometer are used to measure changes in the position of the piston.

The precision barrel that was on hand for use in the prototype comparator has an attenuation constant of 21.304 dB per inch of piston travel at 3 MHz. The drive mechanism and micrometer are such that relative attenuation settings as small as 0.0005 dB can be resolved.

CW Generator

The crystal controlled 3-MHz CW generator delivers one watt of power to the input-matching network of the precision attenuator. The frequency stability is one part in 10^6, and the output voltage is stabilized to five parts in 10^5 by means of a high-gain degenerative feedback loop.

Channel A, Channel B Networks

A block diagram of Channel A and Channel B networks is shown in Fig. 4. These networks consist of a

Manuscript received July 18, 1964.
The authors are with the National Bureau of Standards, Boulder, Colo.
[1] Allred, C. M., A precision noise spectral density comparator, *J. Res. NBS*, vol 66C, Oct 1962, pp 323–330.

Fig. 2. Hybrid four-port.

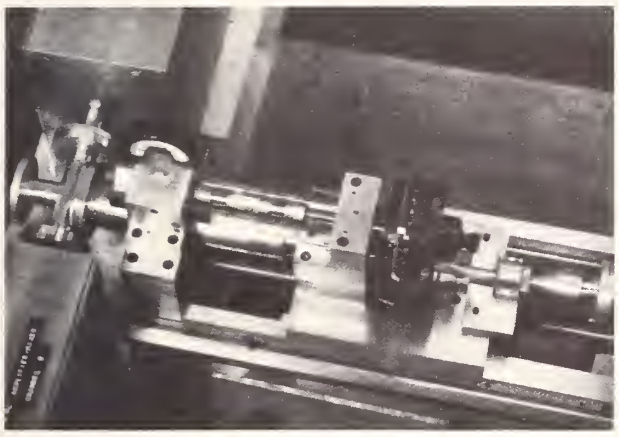


Fig. 3. Piston attenuator assembly.

low noise 3-MHz RF amplifier, a linear-frequency converter, several 500-kHz IF amplifier stages and gain controls, and a 500-kHz narrow band-pass filter. Although these circuits are designed and built with special care, no new principles are involved.

The overall gain of each channel is approximately 135 to 180 dB, depending upon gain control settings. An automatic level control servo system, operating from voltage obtained from the analog-multiplier circuit, maintains a nearly constant output voltage from these networks.

Figure 5 shows the frequency and phase response of the comparator, and the equivalent rectangular bandwidth. The relative phase difference between the output voltages from the two channels does not exceed four degrees over the pass band.

Analog Multiplier

The analog multiplier is constructed from four 5-mA thermal converters, arranged to form a quarter-square multiplier. Figure 6 shows a schematic diagram of the multiplier circuit. Because of differences in the transfer characteristics of the four thermal converters, the multi-

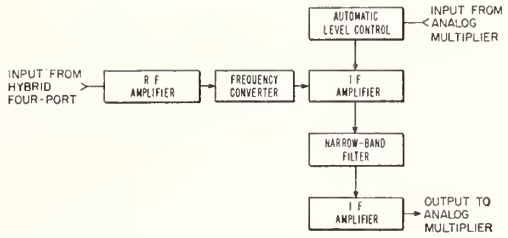


Fig. 4. Block diagram of channel A and channel B networks.

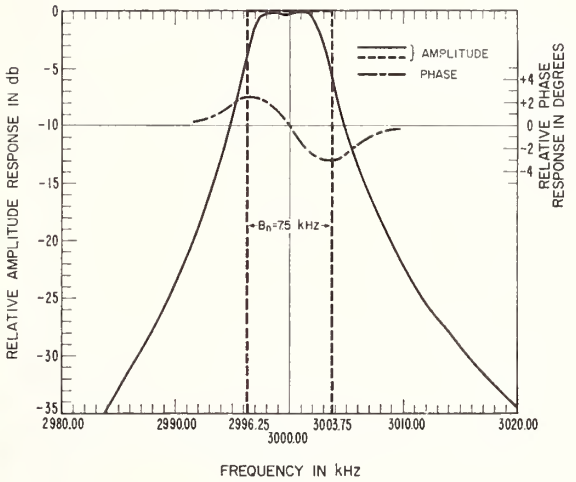


Fig. 5. Amplitude and phase response of comparator.

plier is operated at a constant-input level. The level-sense thermal converters and the automatic level-control servo system regulate the input currents to the multiplier at approximately 4 mA to within 0.1 per cent.

To further stabilize the operating point of the thermal converters, they are housed in a heavy copper compartmentalized box, shown in Fig. 7, which is temperature controlled by circulating temperature-regulated water around four sides of the box.

Averaging Network

The averaging network may take on any of three configurations.

1) It may be a simple RC filter network. In this case, the output voltage of the filter approximates the average of the multiplier-output voltage, over a time approximately equal to the time constant of the filter. A sensitive high impedance, dc microvoltmeter displays the filter-output voltage.

2) It may consist of a sensitive recording voltmeter. In this case, the average of the multiplier output voltage is obtained by averaging graphically the recorded voltage over a suitable period of time.

3) It may consist of a voltage-to-frequency converter and digital-counter combination that effectively integrates the multiplier-output voltage over a preselected period of time.

385-582 O - 70 - 13

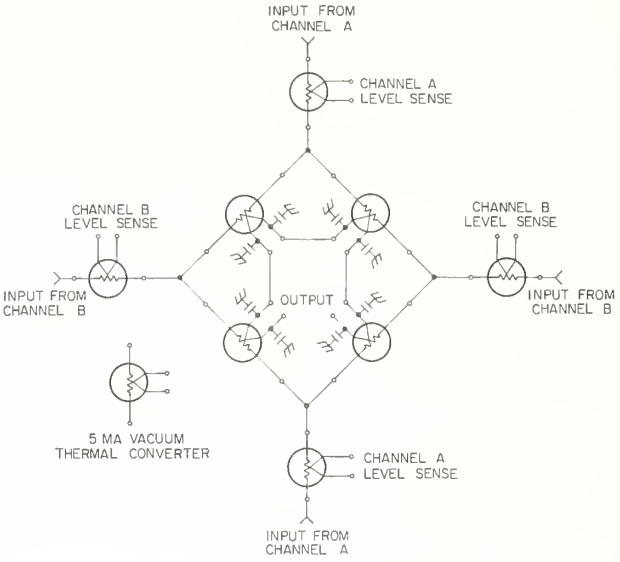


Fig. 6. Analog multiplier circuit.

Fig. 7. Analog multiplier.

Each method has its particular advantages and disadvantages. A combination of 1) and 2) may be used with additional advantages.

Circuit Adjustment

The circuits and components of the prototype comparator are designed and adjusted to represent as nearly as practicable the circuits outlined by Allred.[1]

The comparator is designed for use with noise generators having nominally 50-ohm source impedance. The hybrid four-port is nearly symmetrical and ideal in its operation. The Channel A and Channel B networks are adjusted to provide zero relative phase difference at their output ports for a 3-MHz input signal. This also results in a nearly 180° relative-phase difference for the 3-MHz CW signal entering the hybrid four-port via the precision attenuator. The narrow band-pass filters in these networks are constructed to be as identical as possible, and their response functions are essentially independent of gain fluctuations of the amplifiers. The nonideal behavior of the analog multiplier is minimized by holding its operating point constant.

Comparator Operation

In the basic general theory of operation given by Allred,[1] certain assumptions were stated to simplify the development of the theory; consequently, all the operational parameters involved in the prototype comparator were not taken into account. A totally complete theory is very lengthy and will not be given here.[2] However, a brief heuristic description of the essential operation of the prototype follows.

With a noise generator connected to the input port of the comparator, the total effective input power P_e, lying within the total response band of the comparator and entering this port, is proportional to the sum of T_n, the effective noise temperature of the generator, and T_e, the effective input noise temperature of the comparator (due to its internal noise) referred to this port. Thus

$$P_e = a(T_n + T_e) \tag{1}$$

The total effective input power P_h, entering the port of the hybrid four-port to which the precision attenuator is attached and lying within the total response band of the comparator, is proportional to the sum of AP_{cw}, the attenuated power from the CW generator, and T_h, the effective input noise temperature of the comparator referred to this input port. Thus

$$P_h = b(AP_{\mathrm{cw}} + T_h). \tag{2}$$

At null, these two input powers are related by a factor c. Thus

$$P_e = cP_h. \tag{3}$$

The effective input noise temperatures T_e and T_h will be constant to the extent that the noise factors of the amplifier networks are stable. The quantity c will be constant with time and input-power level if the parameters of the hybrid four-port, the narrow band-pass filters, and the analog multiplier are stable, and if the amplifier networks are linear.

From (1), (2), and (3),

$$T_n + T_i = \frac{bc}{a} AP_{\mathrm{cw}}, \tag{4}$$

where

$$T_i = T_e - \frac{bc}{a} T_h. \tag{5}$$

[2] To be published.

For two known noise generators of equal source impedances and spectral-density distributions, but with noise temperatures of T_{n_1} and T_{n_2}, the null condition gives

$$T_{n_1} + T_i = A_1 \frac{bc}{a} P_s \tag{6}$$

and

$$T_{n_2} + T_i = A_2 \frac{bc}{a} P_s. \tag{7}$$

Solving (6) and (7) for T_i gives

$$T_i = \frac{\frac{A_1}{A_2} T_{n_2} - T_{n_1}}{1 - \frac{A_1}{A_2}}. \tag{8}$$

The ratio A_1/A_2, in dB, is obtained directly from the precision attenuator.

Having evaluated T_i, the noise temperature T_x of an unknown noise generator is measured in terms of the noise temperature T_r of a reference standard noise generator by use of the relation

$$T_x = T_r \frac{A_x}{A_r} - \frac{1 - \frac{A_x}{A_r}}{1 - \frac{A_1}{A_2}} \left(T_{n_2} \frac{A_1}{A_2} - T_{n_1} \right). \tag{9}$$

Here, either T_{n_1} or T_{n_2} may be equal to T_r, thus reducing to two the number of known temperatures of noise generators required to perform a measurement.

Comparator Performance

Sensitivity

For the purposes of this paper, the sensitivity of the comparator is defined as the change in average output voltage from the analog multiplier, near null, per unit relative change in input-noise temperature. The units of sensitivity are millivolts per degree Kelvin per degree Kelvin (mV/°K/°K).

The two principal factors that affect the sensitivity of the comparator are 1) the reduction of gain by the automatic-level control with an increase of input noise power, and 2) the increased relative contribution of the net effective input noise temperature T_i of the comparator to the total input noise power at lower noise-generator temperatures. The changes in multiplier-output voltage for 1) a change in effective noise temperature ΔT_n of a generator and 2) a change in attenuation ΔA of the precision attenuator, are shown in Fig. 8. The empirically-determined sensitivity of the comparator, as a function of the noise temperature of a noise generator, is shown in Fig. 9. The comparator sensitivity changes by less than a factor of three over the temperature range indicated.

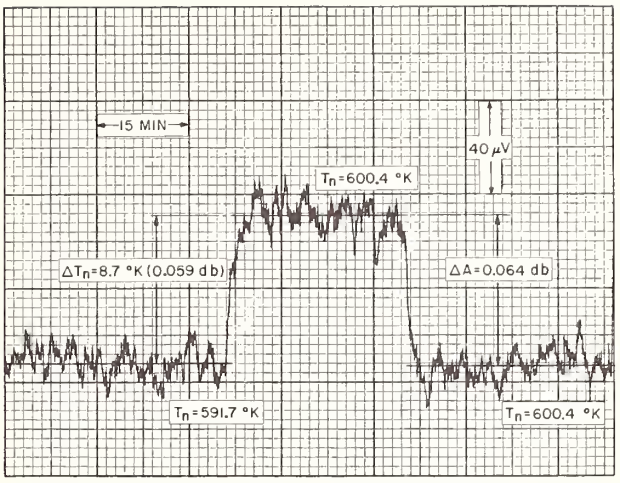


Fig. 8. Multiplier output voltage.

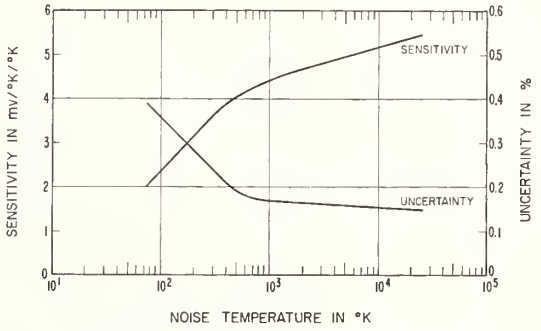


Fig. 9. Sensitivity and uncertainty curves of comparator.

Uncertainty of a Measurement

Several factors affect the uncertainty of a comparison measurement:

1) The rms fluctuations of the averaging network output voltage.
2) The sensitivity of the comparator.
3) The stability of the analog multiplier.
4) The stability of the CW generator.
5) The stability of the net effective input noise temperature of the comparator.
6) The stability of the frequency-response determining networks in the comparator.

Uncertainty is here defined as the uncertainty, in millivolts, in the mean output voltage from the averaging network divided by the sensitivity of the comparator. When multiplied by 100, this quantity is expressed in per cent.

With a recording voltmeter to provide an averaging time of approximately thirty minutes, the empirically-determined uncertainty as a function of generator-noise temperature is also shown in Fig. 9. The increase in uncertainty with a decrease in noise temperature is

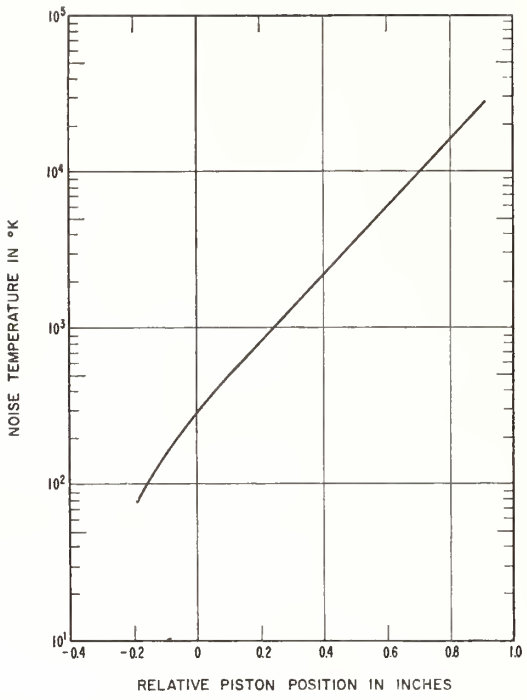


Fig. 10. Calibration curve of comparator.

principally due to the nonzero effective input noise temperature of the comparator.

Calibration Curve

A calibration curve for the comparator is shown in Fig. 10. On the semilog plot, the curve is nearly linear above 500°K and becomes curved where the internal noise becomes influential. From this curve, the net effective internal noise temperature T_i of the comparator appears to be approximately 40°K.

Conclusion

The prototype 3-MHz noise-power comparator has been shown to have operational characteristics as predicted by theory and can intercompare noise generators having effective noise temperatures from below 75°K to greater than 30,000°K. The accuracy of comparison is better than 1 per cent at 75°K and improves to better than 0.2 per cent above 500°K. These results indicate that a high-quality comparator can be built for use in a calibration service for noise power.

Acknowledgment

The authors gratefully acknowledge the contributions to this work by L. D. Driver, R. L. Martin, E. E. Baldwin, W. H. Long, and J. J. Norton.

Reprinted from IEEE TRANSACTIONS
ON *INSTRUMENTATION AND MEASUREMENT*
Volume IM-13, Number 4, December, 1964
pp. 301-305

Measurement of Effective Temperatures of Microwave Noise Sources

J. S. WELLS, W. C. DAYWITT, AND C. K. S. MILLER

Summary—This report describes a system now in operation for the purpose of calibrating microwave noise sources at three selected frequencies in the range 8.2 to 12.4 Gc. Included are a discussion of the reference standard noise source, an error analysis of the standard source and the comparison system, and evidence of system performance. Results of measurements indicate that the excess noise ratio of a specific commercial argon filled noise source is 15.6 db at 9.8 Gc.

A calibration service is now available for tube-in-mount noise sources at the three selected frequencies 9.0 Gc, 9.8 Gc, and 11.2 Gc.

I. Introduction

RECENT WORK in microwave noise measurement techniques in the Radio Standards Laboratory, Boulder, Colo., has led to the completion of a hot load standard noise source and an improved microwave radiometer. These are now incorporated in a microwave noise calibration system.

This report includes a description of recent improvements in the radiometer, the reference standard noise source, and an analysis of possible errors involved in making a measurement. The results of experiments to determine the effective noise temperature of one type of commonly used argon noise source are presented. This source is used as a working standard for the calibration service in the frequency range 8.2 to 12.4 Gc.

Manuscript received June 14, 1963; revised January 22, 1964. This paper was previously published in the 1962 IRE INTERNATIONAL CONVENTION RECORD, pt. 3, pp. 220–238.

The authors are with the National Bureau of Standards, Boulder, Colo.

II. Radiometer

The comparison system consists of a modified radiometer of the type developed by Dicke [1]. The principal modification of the Dicke radiometer is the use of one arm for comparison purposes and the other for establishing a reference noise level. A block diagram of the system is shown in Fig. 1. A photograph of the comparison system is shown in Fig. 2. An earlier version of this system and its operation have been previously described [2]; therefore, a detailed description of its components and operation will not be given.

In brief, the power from an unknown noise source and the power from a standard noise source are compared through a single arm of the radiometer, Fig. 1, and the ratio of the power levels is determined from an attenuation difference read on a precisely calibrated variable attenuator. The following procedure is used to determine the attenuation difference. The unknown source is connected to the radiometer input and the precision attenuator is set to one of the calibrated points in the vicinity of 10.6 db. A null output of the radiometer is then obtained and is achieved by readjusting the balance attenuator, which remains fixed during the remainder of the measurement. The hot load standard is then sub-

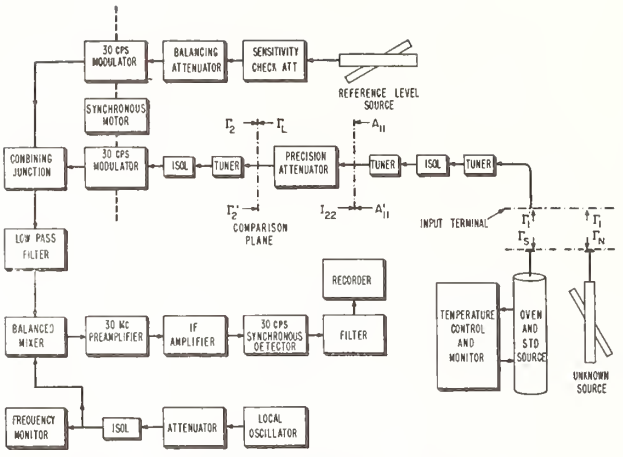


Fig. 1—Block diagram of the system.

Fig. 2—The radiometer.

stituted for the unknown source and the precision attenuator readjusted to null output. The attenuation difference is then the difference between the attenuation readings on the precision attenuator at the two null conditions. The temperature of the unknown source is then determined from the attenuation difference, the temperature of the attenuator, and the effective temperature of the standard.

III. Reference Standard

The reference standard is based on Nyquist's theorem and in practice consists of a high temperature waveguide terminated by a suitably matched resistive element. The waveguide itself is one continuous piece with a heat sink in the form of a water jacket enclosing the flanged end. This design eliminates two flange joints and the resulting reflections which had existed in the previous model [2], and since flange joints lack exact reproducibility, the new design permits greater uniformity. The use of a homogeneous material for the entire length of waveguide also simplifies the calculation of the attenuation as a function of temperature.

The choice of materials for a high temperature waveguide is somewhat limited. Magnetic materials such as inconel and nickel were eliminated due to some uncertainty regarding their behavior in the vicinity of the Curie point. The materials given further consideration were restricted to the noble metals because of their relative stability and resistance to oxidation and corrosion. Most of these were eliminated for various reasons. Silver has too low a melting point; palladium and rhodium oxidize at relatively low temperatures; ruthenium, rhodium, and iridium are difficult to work, and osmium produces poisonous fumes when heated [3].

It was finally decided to build two standards, as shown in Fig. 3. One is of gold and the other is of a platinum-13 per cent rhodium alloy. Both gold and platinum are too soft for repeated use at high temperatures. The platinum-rhodium alloy is quite satisfactory mechanically and will maintain its shape over a long period; however, the uncertainty in its attenuation is larger than that of a gold waveguide. The attenuation of the gold waveguide is calculable, but its softness and lower melting point are disadvantages. The gold waveguide was terminated by a silicon-carbide load, and the platinum-rhodium alloy waveguide was terminated by a zinc-titanate load. It was necessary to use something other than a silicon-carbide load in the platinum alloy waveguide, since platinum and silicon combine to form a eutectic which melts around 800°C [4].

The temperature controller and oven used to heat the waveguide and its resistive termination are shown in Fig. 4. The temperature controller was designed by E. Campbell of the Boulder Laboratories and was found to be very effective. A horizontal orientation of the oven was found to be more desirable than a vertical orientation for several reasons. Possible cooling of the load by convection currents is reduced. In addition, the thermocouples which measure the temperature of the termination are brought out of the rear of the oven and thus the possibility of error due to kinking is minimized. With the oven in the horizontal position the flange of the output terminal of the standard may be connected directly to a reflectometer for a more accurate measurement of the reflection coefficient. The load may be positioned within the waveguide for optimum performance by use of a ceramic rod inserted from the rear of the oven. The desired temperature distribution is achieved by controlling the current through three separate heater coils wound on a single core inside the oven. While in theory it would seem desirable to servo all three coils, proportional control of the current in the center coil has proved satisfactory. The sensing thermal junction is located longitudinally in the proximity of the hot load and radially near the inside of the core. The current through the auxiliary coils on either side of the central coil is adjusted manually to minimize the gradient along the length of the load. The load itself contains a thermocouple that is used to monitor its temperature.

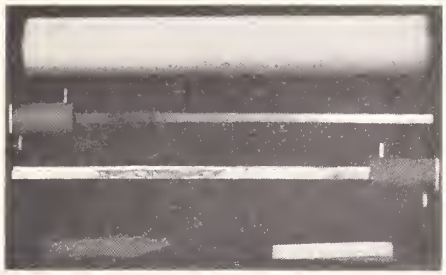


Fig. 3—Heat distributor, loads, and high temperature waveguides.

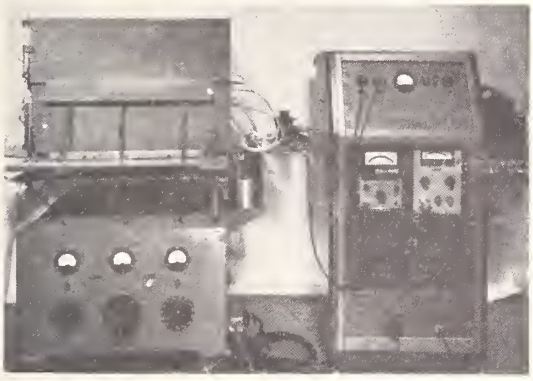


Fig. 4—Temperature controller and oven.

Fig. 5—Inorganic growth in an inconel waveguide.

A combination water flow and temperature indicator is attached to the front of the oven. It was found necessary during the initial heating to keep the jacketed section of the waveguide considerably above room temperature to prevent water from condensing on the interior of the waveguide. Some earlier erroneous results are attributable to such condensation. Another source of error during the developmental stage was power absorption by an inorganic growth which formed in the inconel waveguide then in use, as pictured in Fig. 5. The bulk of this material was lost in the process of measuring its power absorption, hence a proposed chemical analysis could not be attempted. Later efforts to duplicate this growth were unsuccessful.

IV. Calculation of Output Temperature

A. General Analysis

The noise power entering the radiometer from the standard may be written as

$$P = kBT_s \frac{(1 - |\Gamma_s|^2)(1 - |\Gamma_1'|^2)}{|1 - \Gamma_s\Gamma_1'|^2}$$

where the Γ's, reflection coefficients, are interpreted in Fig. 9 (Appendix I), T_s is the effective temperature of the standard, k is Boltzmann's constant, and B the bandwidth of the system. T_s will be less than the measured value of the load temperature, due to the attenuation of the noise power by the walls of the high temperature waveguide. However, the resistivity of the walls also contributes to the noise output, and this must be taken into account.

The analysis used here to determine the output temperature of the standard is essentially that outlined by Sees [5]. His expression can be written

$$T_s = 10^{-A_g'/10}\left[T_m + \int_0^b a(z)T_g(z) \cdot \exp\left(\int_0^z a(z')dz'\right)dz\right], \quad (1)$$

where:

T_s = the output noise temperature in °K
$T_g(z)$ = the temperature of the waveguide in °K
T_m = the measured temperature of the load in °K
A_g' = the total attenuation of the waveguide in decibels
z = the axial distance in inches from the tip of the load (for convenience the reference point was taken to coincide with the tip of the load) to the waveguide region under consideration
b = the length of the waveguide from the tip of load to the output flange in inches
$a(z)$ = the attenuation in nepers per unit length, $= (a_0/\rho_0^{\frac{1}{2}})(\rho(z))^{1/2}$ where ρ is the resistivity of the waveguide. The subscript "$_0$" refers to 0°C. The quantities a_0 and ρ_0 are constants, depending on the size and composition of the waveguide.

Eq. (1) is obtained from

$$T_s = T_m \exp\left[-\int_0^b a(z')dz'\right] + \int_0^b a(z)T_g(z) \cdot \exp\left[-\int_z^b a(z')dz'\right]dz.$$

The physical interpretation of (1) is made more evident with this alternate form, where the first term in the alternate expression is proportional to the power generated by the hot load and attenuated by the waveguide

between 0 and b, and the second term is proportional to the sum of the power generated in each infinitesimal segment dz of waveguide at z and attenuated by the remaining $(b-z)$ inches of waveguide.

To make practical use of (1), it is convenient to make several approximations. With the section of the waveguide between the hot load and the output flange considered to be divided into sufficiently small segments, the quantities a and T_g can be considered constant over each segment. Then, when (1) is applied to each segment, the output of the kth segment may be expressed as

$$T_{k+1} = 10^{-A_k'/10}\left[T_k + \int_{z_k}^{z_{k+1}} a_k T_{gk} \cdot \exp\left(\int_{z_k}^{z} a(z')dz'\right)dz\right], \quad (2)$$

which becomes

$$T_{k+1} = T_k - 0.23A_k'(T_k - T_{gk}) \quad (3)$$

provided that second and higher order terms in A_k' (the total attenuation of the kth section) are discarded and the average values of a and T_g are taken. A_k' is related to a_k by $A_k' = 4.34\ a_k \Delta z_k$, where $4.34\ a_k$ is the average attenuation in decibels per unit length for the kth segment and Δz_k is the length of this segment.

To apply (3), the temperature distribution along the waveguide must be measured and $a(z)$ calculated as a function of this distribution. The expression for the attenuation per unit length (db/in) along the waveguide is given in Appendix II as

$$C_1(\rho/f)^{1/2}\frac{(f^2 + C_2)}{(f^2 - C_3)^{1/2}} \quad (4)$$

where f is the frequency in Gc, ρ is the resistivity in $\mu\Omega$-cm, and C_1, C_2, C_3 are constants. Eq. (4) is used to calculate A_k' for each segment. Roeser and Wensel's [6] values for resistivity as a function of temperature for platinum and platinum-rhodium alloys appear in Table I. Values for resistivity as a function of temperature for gold are taken from Northrup's work [7].

The uncertainty in the output temperature of each section, δT_{k+1}, is found by taking the total differential of (3) and is

$$\delta T_{k+1} = |\delta T_k| + 0.23(T_k - T_{gk})|\delta A_k'| + 0.23A_k'|\delta(T_k - T_{gk})|. \quad (5)$$

The total uncertainty in T_s is then found from successive applications of (5).

B. Sample Calculation

The analysis in Section IV-*A* will now be used to calculate the correction to the final output noise temperature of the hot load standard and also the accompanying uncertainty. Fig. 6 shows a typical temperature distribution for a gold waveguide as measured by temporarily positioning a thermocouple at integral values of z inside the waveguide and reading the thermal EMF after equilibrium had been attained. The tip of the load is located at $z=0$. When the system is in steady state, the gradient can be held to within 1°C per load length. The temperature at $z=11$ inches is dependent upon the temperature and rate of flow of water through the water jacket. Also indicated in Fig. 6 are the assigned values of T_{gk} and A_k' for each segment.

T_s and δT_s are calculated from (3) and (5) and Fig. 6. This calculation shows that T_s, the temperature of the reference standard, is reduced by 5.9°K, due to the resistivity of the waveguide; and that the uncertainty in T_s, δT_s, is 4.1°K.

TABLE I

RESISTIVITY OF WAVEGUIDE MATERIALS AS A FUNCTION OF TEMPERATURE

Temp. °C	Platinum 9.83 $\mu\Omega$-cm* ρ/ρ_0	90 per cent Platinum 10 per cent Rhodium 18 $\mu\Omega$-cm* ρ/ρ_0	87 per cent Platinum 13 per cent Rhodium 19.0 $\mu\Omega$-cm* ρ/ρ_0	Gold ρ
0	1.000	1.000	1.000	
100	1.392	1.166	1.156	2.97 $\mu\Omega$-cm
200	1.773	1.330	1.308	3.83
300	2.142	1.490	1.456	4.72
400	2.499	1.646	1.601	
500	2.844	1.798	1.744	6.62
600	3.178	1.947	1.885	
700	3.500	2.093	2.030	
800	3.810	2.234	2.157	9.94
900	4.109	2.370	2.287	
1000	4.396	2.503	2.414	12.52
1100	4.671	2.633	2.538	
1200	4.935	2.761	2.660	
1300	5.187	2.887	2.780	
1400	5.427	3.011	2.898	
1500	5.655	3.133	3.014	

Data for Platinum alloys given by Roeser and Wensel [6].
Data for Gold given by Northrup [7].
* Resistivities at zero degree centigrade.

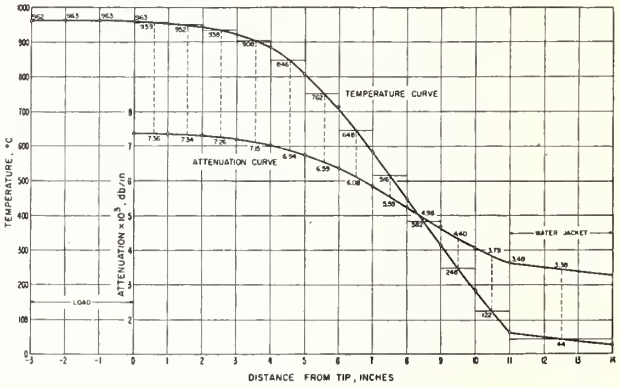


Fig. 6—Typical waveguide temperature distribution.

V. Error Analysis

A. Introduction and General Analysis

Several factors limit the accuracy to which the noise temperature of an unknown source can be measured. These may be grouped into two categories:

1) *Temperature uncertainties:* These temperature uncertainties depend on three factors, the accuracy with which temperatures can be measured, the accuracy to which attenuation or resistivity of the high temperature guide is known, and the approximations made in determining the noise temperature at the output terminal of the standard source.
2) *Uncertainties in the analysis of the comparison system:* One might question how closely an *indication* of equality relates the effective noise temperatures of two sources. The variables here are the attenuator error, which includes mismatch, resolution and resettability. Further included in system uncertainties are system resolution and the impedance matches at the input of the radiometer and output of the noise power generators. It is also convenient to put a small uncertainty in the balance equation in this group. This arises due to changes in the noise power originating in and reflected from the attenuator [8] at different attenuation levels.

In order to conveniently combine the above uncertainties into an expression for the total uncertainty of the noise temperature of the standard, the subsequent procedure has been followed: a first-order balance equation, which is essentially the one used for computational purposes, is indicated; this equation is then differentiated to determine the weights assigned to the various uncertainties. The temperature uncertainties are treated at this point.

In a matched system with limitless resolution and stability, a balance equation could be easily obtained by measuring the ratio of two excess noise powers and expressing the ratio in terms of the measured attenuation. The first-order balance equation based on these assumptions is

$$\frac{kB(T_{ne} - T_a)}{kB(T_s - T_a)} = 10^{\Delta A/10}$$

which may be written in the form

$$\Delta A = 10 \log_{10} \frac{kB(T_{ne} - T_a)}{kB(T_s - T_a)} = 4.34 \log \frac{(T_{ne} - T_a)}{(T_s - T_a)} \cdot \quad (6)$$

The quantities in the above equations are defined as follows:

ΔA = the attenuation difference in db, as read from the calibrated attenuator
k = Boltzmann's constant
B = the bandwidth of the system
T_a = the temperature of the attenuator in °K
T_s = the load thermocouple temperature modified by the effective attenuation of the guide in °K
T_{ne} = the effective temperature of the noise tube and mount to be determined (the mount includes a termination at room temperature at one end).

By differentiating (6) and rearranging, one obtains

$$\delta T_{ne} = -\left(\frac{T_{ne} - T_a}{T_s - T_a} - 1\right)\delta T_a + \left(\frac{T_{ne} - T_a}{T_s - T_a}\right)\delta T_s + 0.23(T_{ne} - T_a)\delta(\Delta A). \quad (7)$$

From (7), the expression for the maximum uncertainty becomes

$$|\delta T_{ne}| \leq \left(\frac{T_{ne} - T_a}{T_s - T_a} - 1\right)|\delta T_a| + \left(\frac{T_{ne} - T_a}{T_s - T_a}\right)|\delta T_s| + 0.23(T_{ne} - T_a)\,|\delta(\Delta A)|\,. \quad (8)$$

Although the above treatment is based on assumptions which are not strictly justified, it is none the less a convenient method for separating the different sources of uncertainty and does not lead to error in the treatment of temperature uncertainties which follows. In order to evaluate $\delta(\Delta A)$ a more detailed expression will be derived following a discussion of the temperature uncertainties.

B. Uncertainty in the Temperature of the Radiometer

δT_a is the uncertainty in measuring the ambient temperature of the attenuator section of the radiometer. The room is temperature controlled and T_a does not vary appreciably. A thermocouple was taped to the side of the resistive element in an attenuator (which was substituted for the calibrated attenuator) to measure heating due to radiation when the thermal noise standard was connected to the radiometer. An increase in temperature of less than 1°C was observed 30 minutes after noise power from the reference standard had been applied. This is the normal time duration for a measurement.

As may be seen by examining (8), a δT_a of 1°K in the equation causes an uncertainty which is small compared to that contributed by the other sources. Hence, a conservative estimate is

$$\delta T_a \leq 2°\text{K}.$$

C. Uncertainty in the Temperature of the Standard

It is convenient to define δT_s as $\delta T_m + \Sigma \delta T_k$, where δT_m is the uncertainty in the measurement of the load temperature and is due to a small inaccuracy in the calibration of the thermocouple and to its physical placement at a small distance from the point of interest.

Error due to the latter is reduced by minimizing the temperature gradient across the load and waiting sufficiently long to ensure thermal equilibrium. The quantity $\delta T_m/T_m$ can be as much as ± 0.25 per cent. However, precautions are taken to minimize this uncertainty by annealing the platinum thermocouple at approximately 1450°C for one hour. The utmost care is exercised to avoid contamination in construction of the thermocouples. Insulation and protection tubes of high purity alumina (better than 99.0 per cent) are used for greater stability and accuracy. The portion of the thermocouple not enclosed in protective ceramic tubing is prevented from kinking by enclosing it in thick walled plastic tubing. The $\Sigma\delta T_k$ contributions have been discussed. The uncertainty in the effective temperature is

$$\delta T_s = \delta T_m + \Sigma\delta T_k \leq 4.1°\text{K}.$$

D. Uncertainties Associated with the Comparison System

In practice, one does not have a matched system. In order to properly evaluate the resulting uncertainties, a more sophisticated balance equation which includes the effects of the various mismatches must be used. The phases of the impedance mismatches appearing in this higher-order balance equation are not known, but the new equation does permit one to establish the limits of uncertainty which have been neglected in the computational equation. As the derivation of the complete balance equation is somewhat tedious, it has been relegated to Appendix I. The resulting expression, as it appears there, is

$$\begin{aligned} 10\log_{10}\frac{(T_{ne}-T_a)}{(T_s-T_a)(1-|\Gamma_s|^2)} &= \Delta A + 20\log_{10}\left|\frac{1-C_{11}\Gamma_n}{1-C_{11}'\Gamma_s}\right| \\ &\quad + 20\log_{10}\left|\frac{1-\Gamma_2\Gamma_L}{1-\Gamma_2'\Gamma_L}\right|\left|\frac{1-I_{22}A_{11}}{1-I_{22}A_{11}'}\right| \\ &\quad + 10\log_{10}(1+\epsilon), \end{aligned}$$

where the second term on the right is a source mismatch uncertainty, the third is an attenuator mismatch uncertainty, and the fourth term arises when one considers the effect of noise contributed by and reflected from the mismatch attenuator. (The Γ's, A_{ii}'s, and I_{ii}'s are reflection coefficients and scattering elements which are defined in the Appendix). The last three terms, along with a term representing the uncertainty in the calibration of the attenuator, are all included in $\delta(\Delta A)$. The magnitudes of these uncertainties will now be considered.

1) *Precision Attenuator Mismatch Uncertainty:* The maximum value of the precision attenuator mismatch uncertainty is given by

$$\delta(\Delta A)_\epsilon = 20\log_{10}\frac{|1\pm|I_{22}A_{11}|||1\pm|\Gamma_2\Gamma_L||}{|1\mp|I_{22}A_{11}'|||1\mp|\Gamma_2'\Gamma_L||}.$$

When the equality $\log_{10} x = 0.434\log_e x$ and the approximation $\log_e(1+x)\approx x$, are inserted into this expression, it becomes

$$\delta(\Delta A)_\epsilon \approx 8.68\{|I_{22}|(|A_{11}|+|A_{11}'|) + |\Gamma_L|(|\Gamma_2|+|\Gamma_2'|)\}. \quad (9)$$

Multiple-stub tuners are used to reduce the reflection coefficients $|\Gamma_L|$ and $|I_{22}|$ to as small a value as possible in order to minimize $\delta(\Delta A)$. The uncertainty in measuring these parameters is small compared to their absolute value and may be ignored. It was found necessary to use two isolators in cascade following the tuners to insure that the measured reflections remained the same when the units were connected to the radiometer.

The magnitudes of the reflection coefficients are measured at the reference planes which are indicated in Fig. 1 and discussed in Appendix I. $|A_{11}|$, a scattering element at the input of the attenuator, is measured with Γ_L replaced by a matched load ($\Gamma_L \leq 0.001$). Typical curves of reflection coefficient as a function of frequency are presented in Fig. 7, with the IF signal and image frequency responses indicated by cross hatching. The primes of (9) refer to the values of the quantities obtained with the attenuator set at approximately zero db, as is the case when the hot load is supplying the power at the radiometer input. Using Fig. 7 one can calculate $\delta(\Delta A)_\epsilon$ with the result

$$\delta(\Delta A)_\epsilon \leq 0.016 \text{ db}.$$

2) *Source Mismatch Uncertainty:* The second factor in the expression for ΔA may be considered to be an uncertainty due to multiple reflections resulting from two mismatched sources supplying noise power to a mismatched input terminal at different times. This source mismatch factor may be written

$$\delta(\Delta A)_m \leq 20\log_{10}\frac{[1\pm|C_{11}\Gamma_n|]}{[1\mp|C_{11}'\Gamma_s|]}$$

or,

$$\delta(\Delta A)_m \approx 8.68\{|C_{11}\Gamma_n| + |C_{11}'\Gamma_s|\}.$$

A typical value is

$$\delta(\Delta A)_m \leq 0.005 \text{ db}.$$

3) *Balance Uncertainty:* Using typical values of the parameters in (19) of Appendix I to calculate ϵ, one finds that

$$10\log_{10}(1+\epsilon) \approx 0.006 \text{ db}.$$

The experimental resolution of the radiometer is found by changing the precision attenuator setting a small measured value and noting the resulting deflection on the recorder. Attenuation changes corresponding to the minimum detectable deflection must be included as a balance uncertainty. With a power level corresponding to a hot load temperature of 1000°C at the radiometer input, unbalances of less than 0.005 db are discern-

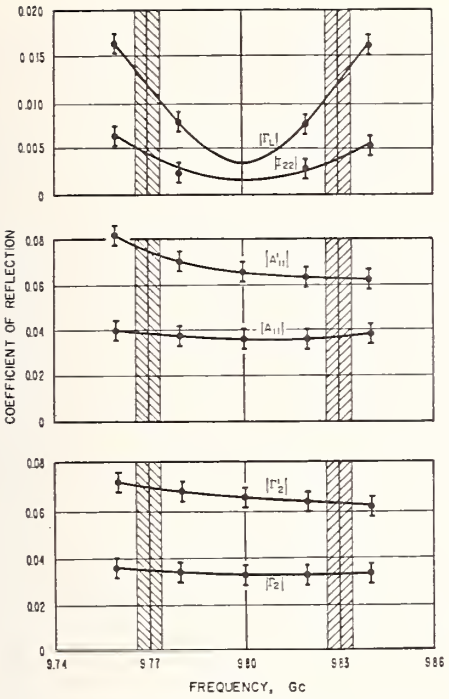


Fig. 7—Typical curves of reflection coefficient as a function of frequency.

ible at the output, and unbalances of 0.002 have been detected. The limit of the resolution is certainly less than 0.01 db.

Adding these two uncertainties the total balance uncertainty is

$$\delta(\Delta A)_u \leq 0.016 \text{ db.}$$

4) *Attenuator Calibration Uncertainty:* The limits of resolution and repeatability were established by means of observations made during the precise calibration of the precision attenuator and are estimated to be ± 0.002 db. The calculated mismatch error was less than 0.005 db. Therefore, in summing both contributions, the uncertainty in the precision attenuator calibration is

$$\delta(\Delta A)_c \leq 0.007 \text{ db.}$$

The requirement for keeping the error in the excess noise ratio to within 0.1 db necessitated a more precise calibration of the attenuator than was commonly available. The method decided upon was a subcarrier technique [9]. In order to assure the accuracy of the attenuator, it is calibrated periodically. The need for assurance was pointed up when it was noted that a previous calibration was changed by 0.04 db at the higher levels of attenuation during an 18 month period. The maximum associated attenuation uncertainty is then obtained by adding the uncertainties of 1, 2, 3, and 4. The result is

$$\delta(\Delta A) \leq 0.044 \text{ db.}$$

E. Maximum Total Uncertainty

The maximum total uncertainty given by (8) will now be calculated. For representative values it is assumed that T_s is 1236°K, that T_a is 296°K, and that T_{ne} is 10,796°K. Then $(T_{ne}-T_a)/(T_s-T_a)$ is 11.17, and δT_s is found to be 4.1°K. $\delta(\Delta A)$ was found from Section V-D to be 0.044 db and δT_a from Section V-*B* was found to be 2°K. Substituting these values into (8) gives $|\delta T_{ne}| \leq (0.23)(10{,}500)(0.044)+(10.17)(2)+(11.17)(4.1) \approx 172$°K.

The excess noise ratio (ENR) is defined to be

$$\text{ENR} = 10 \log_{10}\left(\frac{T_{ne} - T_0}{T_0}\right),$$

where $T_0 = 290$°K. From this equation it is easily found that

$$\delta(\text{ENR}) = 10 \log_{10}\left(1 + \frac{\delta T_{ne}}{T_{ne} - T_0}\right).$$

Substituting the above values into this equation there results

$$\delta(\text{ENR}) = 0.070 \text{ db.}$$

These results represent the maximum total uncertainty in the measurement of effective noise temperature and excess noise ratio of a waveguide noise source given in this analysis.

VI. Experimental Results

The performance of the comparison system may be determined by examining Fig. 8 which contains recordings of the radiometer output and a temperature recording made simultaneously with part of Strip No. 3. Corresponding points in time are indicated by pairs of markers designated *X*, *Y*, and *Z*. The stability and sensitivity at an excess noise temperature of 10,500°K are indicated on Strip No. 1. The distance from the bottom of the chart to the point "1a" represents six hours. The power level was changed 0.02 db at "1a" and returned to its original level at "1b."

The power level in Strip No. 2 corresponds to an excess temperature of 1213°K at the radiometer input which is the nominal operating temperature of the hot load. The increase in fluctuation of the amplitude is due to the deterioration of the signal-to-noise ratio at the lower signal input temperature and the consequent increase in gain required to maintain the same nominal deflection sensitivity. Strip No. 2 shows changes of 0.02 db at "2a" and 0.04 db in the opposite direction at "2b."

Strips Nos. 3 and 4 are indicative of the types of records made during an actual measurement. At the beginning of the measurement the precision attenuator was adjusted to 10.62 db. A small unbalance was noted

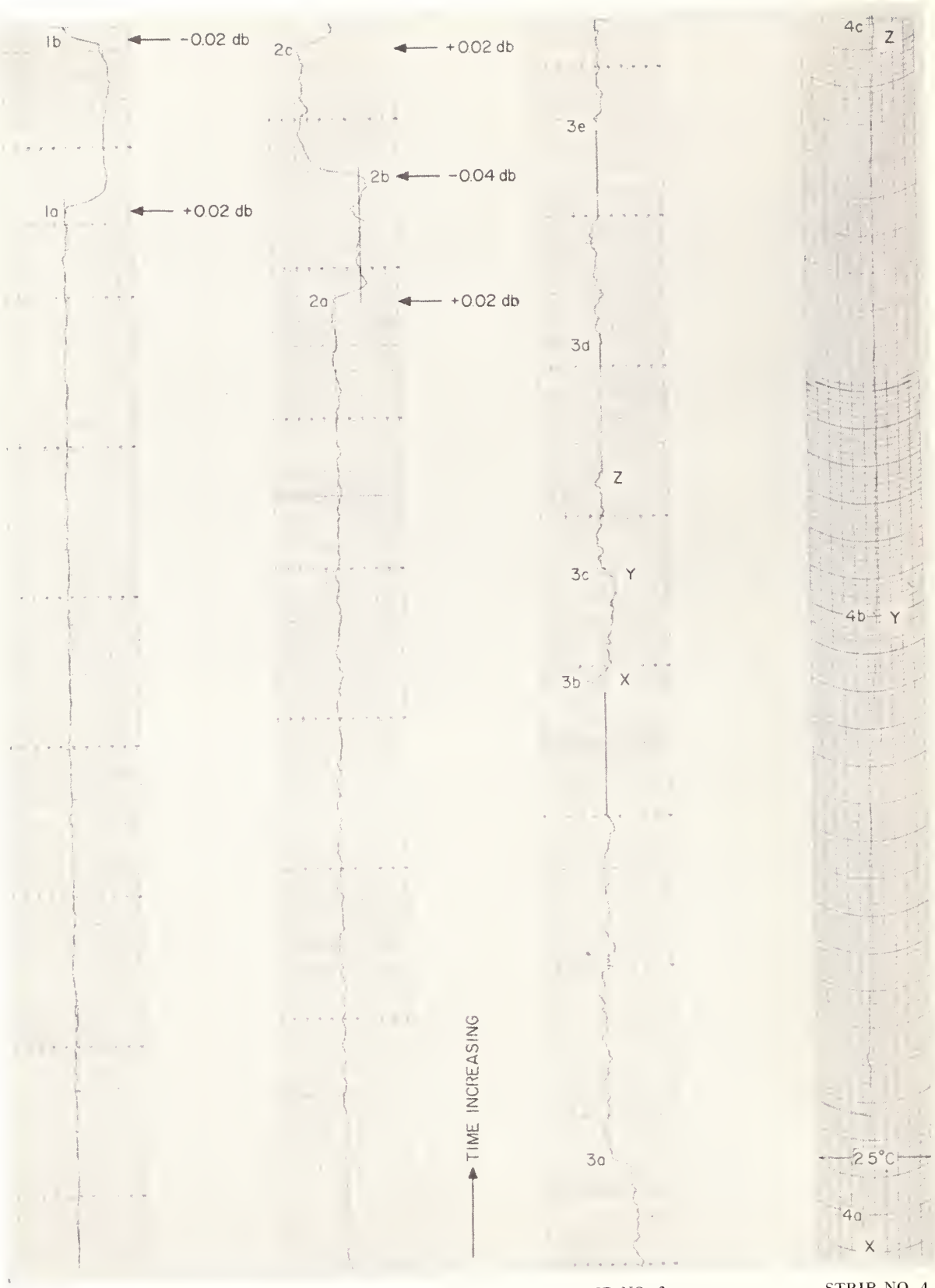


Fig. 8—Records of over-all system performance.

and the balancing attenuator was used to obtain a null output as indicated at "3a." After sufficient time to establish the reference level, the recorder input was short circuited and the unknown waveguide source was replaced by the hot load standard. The precision attenuator was changed from 10.62 db to 0.09 db and the output was recorded again at "3b." Note that the temperature of the reference standard as indicated on Strip No. 4 increases from point X to point Y. At point Y an unbalance is detectable on Strip No. 3 and the attenuator repositioned to 0.10 db, and the system uninterrupted for 30 minutes, during which time the temperature is stable as indicated. At "3d," the unknown waveguide noise source is substituted for the hot load standard and the attenuator returned to 10.62 db. Flange reproducibility and attenuator resettability appear good. The measurement is complete at "3e."

The results of two sets of 10 consecutive runs such as illustrated on Strip No. 3 are shown in Table II to indicate the system repeatability and self consistency. One set was taken using the gold and silicon carbide model of the standard noise source. The other set was obtained using the platinum alloy and zinc-titanate model. The center calibration frequency was 9.8 Gc and the average excess noise ratios were 15.59 db for the gold combination and 15.58 db for the platinum combination. In either case the uncertainty was less than 0.1 db. T_m and T_s are defined in Section IV-A, T_a, ΔA, and T_{ne} in Section V-A and ENR in Section V-E.

TABLE II

System Consistency Indicated by Consecutive Test Runs

Using the Gold Wageguide

T_m°C	T_s°C	T_a°C	ΔA db	T_{ne}°K	ENR in db
958.4	952.6	25.0	10.54	10,800	15.593*
961.2	955.4	25.9	10.53	10,800	15.592
966.6	960.7	26.0	10.51	10,810	15.596
964.5	958.6	25.9	10.50	10,770	15.580
976.4	970.6	26.1	10.46	10,800	15.592
984.5	978.5	26.0	10.43	10,820	15.598
986.0	980.0	26.3	10.41	10,780	15.584
999.5	993.5	25.7	10.35	10,790	15.587
1000.0	994.0	25.8	10.35	10,790	15.589
1010.6	1004.5	25.9	10.30	10,780	15.586

Average T_{ne}=10,790°K
ENR=15.59 db

Using the Platinum-Rhodium Waveguide

T_m°C	T_s°C	T_a°C	ΔA db	T_{ne}°K	ENR in db
947.3	934.5	22.5	10.60	10,770	15.578*
972.0	959.0	25.2	10.50	10,780	15.581
976.6	963.6	24.9	10.48	10,780	15.585
995.6	982.4	24.3	10.39	10,780	15.582
1008.0	994.7	23.3	10.34	10,800	15.593
1019.5	1006.0	24.4	10.28	10,770	15.578
1021.1	1007.6	24.5	10.28	10,780	15.585
1047.4	1033.7	23.9	10.17	10,800	15.591
1059.2	1045.4	25.3	10.11	10,780	15.584
1060.0	1046.2	23.3	10.10	10,760	15.577

Average T_{ne}=10,780°K
ENR=15.58+ db

* The fifth place is retained for averaging only.

To obtain a better idea of the reproducibility of gas-tube noise sources, several different comparisons were made with the radiometer operated at a level corresponding to an excess temperature of 10,500°K. Strip No. 1 in Fig. 8 indicates the stability and sensitivity at this level.

Four each of two different specially constructed noise tubes, one type using a heated cathode, the other operating with a cold cathode, were compared in the same mount, which remained fixed to the comparison system. The spread in excess noise ratio was less than 0.02 db for both sets. Two tubes with nearly identical excess temperatures were then put in different mounts, and the tube-in-mount units were compared. The output of these units differed by as much as 0.05 db. This seems to suggest that the repeatability among tubes is better than that for mounts. Factors, such as a flaw in the glass envelope and a deviation from nominal mount insertion angle were found to cause even larger differences.

VII. Conclusion

Because of the very good agreement between measurements made with the two hot load noise sources it was decided that the outputs from both sources were sufficiently calculable, and that either could be used as a reference noise source. From experimental results discussed, it is apparent that the system has sufficient sensitivity for the purpose of comparing sources, and the above analysis shows that the system is capable of the desired accuracy. It was concluded from these considerations that a calibration service for WR-90 waveguide noise sources could be offered.

Due to small differences in output power of one tube in different mounts it was concluded that for a calibration service to be offered with the desired accuracy it would be mandatory for the units submitted for calibration to consist of a tube-in-mount.

Consequently, a calibration service is being offered at the selected frequencies of 9.0 Gc, 9.8 Gc, and 11.2 Gc to the accuracy of ± 250°K of the effective noise temperature, which corresponds to ± 0.1 db of the excess noise ratio for waveguide noise sources with effective noise temperatures in the order of 11,000°K.

Appendix I

Derivation of Balance Equation

Fig. 9 depicts how the radiometer is considered to be divided for the purpose of this analysis. With reference to Fig. 1, the input terminal of the radiometer is coincidental with the input plane of Fig. 9, and the terminal surface between the output of the precision attenuator and the following tuner is coincidental with the comparison plane of Fig. 9. The attenuator section is considered to be at constant ambient temperature T_a in °K,

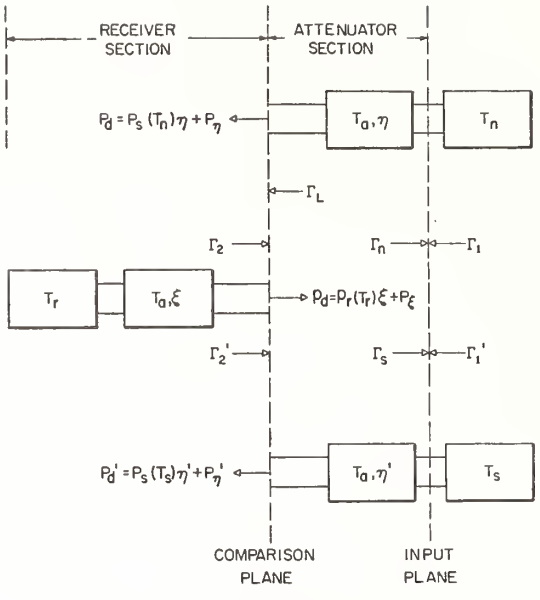


Fig. 9—Schematic showing planes for derivation of balance equation.

with an efficiency η from the input to the output of the attenuator section. The input portion of the receiver section is assumed to be at the constant temperature T_a with an efficiency ξ from the output to the input and the output portion at a constant temperature T_r in °K. The input portion of the receiver section consists of the tuner and isolator immediately following the comparison plane. All Γ's appearing in Fig. 9 are reflection coefficients.

When used for the measurement described in this paper, the radiometer may be considered as a device which indicates the equality of noise power propagating toward the comparison plane under two different conditions. This noise power is composed of power originating in a noise generator at temperature T which is transmitted by the network, $P_s(T)$, power originating to the left of the comparison plane (Fig. 9) which is reflected back into the radiometer by impedance discontinuities denoted by either Γ_2 or Γ_2'. Let P denote power originating to the right of the comparison plane and p denote power originating to the left of this plane. The subscripts $_r$, and $_d$ refer to the reflected and delivered power, respectively. The primed variables refer to the situation in which the thermal noise standard is connected to the input terminal of the radiometer; the same variables without primes refer to the situation in which the waveguide noise source is connected to the input terminal. When one noise source is compared to another and the system is returned to a null, the power entering the receiver section in either case must be the same. Since noise powers from independent sources can be added linearly, we get

$$P_d + p_r = P_d' + p_r' \tag{10}$$

where

$$p_r = \frac{p_d |\Gamma_2|^2}{1 - |\Gamma_2|^2}.$$

A similar expression holds for P_r', in which all the quantities except Γ_L become primed. Fig. 9 may be used to relate P_d and p_d to the measured quantities of interest.

To find P_d, it is necessary to find P_s, the noise power delivered from a source at temperature T to a passive load (*i.e.*, a load at 0°K); P_n, the power contributed by the network; and η, the efficiency of the network.

Combining the expression due to Nyquist for the average voltage squared,

$$\overline{e^2} = 4kTB \operatorname{Re}(Z),$$

with the expression for the power delivered to a mismatch load from a mismatch generator,

$$P = \frac{e^2}{4Z_0} \frac{|1 - \Gamma_g|^2(1 - |\Gamma_h|^2)}{|1 - \Gamma_g\Gamma_h|^2}$$

(k is Boltzmann's Constant, T is the temperature of the source, B is the bandwidth of the receiver, Re (Z) is the real part of the source impedance, Z_0 is the characteristic impedance of the line, Γ_g is an arbitrary source reflection coefficient, and Γ_h an arbitrary load reflection coefficient), the result is

$$P = kTB \frac{(1 - |\Gamma_g|^2)(1 - |\Gamma_h|^2)}{|1 - \Gamma_g\Gamma_h|^2}. \tag{11}$$

Using (11), the power delivered by a source at temperature T°K connected to the input plane is

$$P_s(T) = \frac{kTB(1 - |\Gamma_n|^2)(1 - |\Gamma_1|^2)}{|1 - \Gamma_n\Gamma_1|^2}. \tag{12}$$

P_η, the power generated by the attenuator section, will be determined. With the receiver section at a temperature T_a the expression for the power delivered to the right of the comparison plane is

$$P_L(T_a) = \frac{kT_aB(1 - |\Gamma_2|^2)(1 - |\Gamma_L|^2)}{|1 - \Gamma_2\Gamma_L|^2}.$$

However, when the networks on both sides of the comparison plane are at the same temperature T_a,

$$P_L(T_a) = P_s(T_a)\eta + P_\eta(T_a)$$

or solving for $P_\eta(T_a)$

$$P_\eta(T_a) = P_L(T_a) - P_s(T_a)\eta. \tag{13}$$

In general

$$P_d = P_s(T)\eta + P_\eta(T_a). \tag{13a}$$

The expression for the efficiency η [10] of the network when applied to this system is

$$\eta = \frac{|C_{21}|^2(1 - |\Gamma_L|^2)}{|1 - C_{22}\Gamma_L|^2(1 - |\Gamma_1|^2)}. \tag{14}$$

Where C_{22} is the appropriate element of the scattering matrix for the attenuator section, which consists of a matched isolator and an attenuator. The element C_{21} is the transmission scattering coefficient for the attenuator section and is determined by the usual method of cascading matrices. The quantities I_{ij} and A_{ij} are the elements of the scattering matrices for the isolator and attenuator, respectively.

$$|C_{21}|^2 = \frac{|I_{21}A_{21}|^2}{|1 - I_{22}A_{11}|^2} \tag{15}$$

When (12)–(14), along with the relations

$$(1 - \Gamma_1\Gamma_n)(1 - C_{22}\Gamma_L) = (1 - C_{11}\Gamma_n)(1 - \Gamma_2\Gamma_L),$$

and

$$(1 - \Gamma_1'\Gamma_s)(1 - C_{22}'\Gamma_L) = (1 - C_{11}'\Gamma_s)(1 - \Gamma_2'\Gamma_L)$$

are substituted into an expression (13a) for the power delivered to the receiver section and when the temperature of the unknown source is T_n, the result is

$$P_d = kB\left[\frac{(T_n - T_a)(1 - |\Gamma_n|^2)|C_{21}|^2}{|1 - C_{11}\Gamma_n|^2|1 - \Gamma_L\Gamma_2|^2} + \frac{T_a(1 - |\Gamma_2|^2)}{|1 - \Gamma_2\Gamma_L|^2}\right](1 - |\Gamma_L|^2) \tag{16a}$$

where C_{11} is a scattering coefficient for the attenuator section. Similarly, for the standard source T_s,

$$P_d' = kB\left[\frac{(T_s - T_a)(1 - |\Gamma_s|^2)|C_{21}'|^2}{|1 - C_{11}'\Gamma_s|^2|1 - \Gamma_L\Gamma_2'|^2} + \frac{T_a(1 - |\Gamma_2'|^2)}{|1 - \Gamma_2'\Gamma_L|^2}\right](1 - |\Gamma_L|^2). \tag{16b}$$

Note the common factor in the denominator of the two terms comprising the expressions for P_d and P_d'. In a similar fashion the power represented by p consisting of power from a receiver at temperature T_r and the network ξ at T_a is given by

$$p_d = kB[(T_r - T_a)F\xi + T_a(1 - |\Gamma_L|^2)] \cdot \frac{(1 - |\Gamma_2|^2)}{|1 - \Gamma_2\Gamma_L|^2}$$

$$p_d' = kB[(T_r - T_a)F\xi + T_a(1 - |\Gamma_L|^2)] \cdot \frac{(1 - |\Gamma_2'|^2)}{|1 - \Gamma_2'\Gamma_L|^2}. \tag{17}$$

The details of the matching factor F need not be known since ξ is of the order of 10^{-12}, and hence this term may be neglected. When (15)–(17) are combined with (10) the resulting balance equation may be shown to be

$$\frac{(T_n - T_a)}{(T_s - T_a)}\frac{(1 - |\Gamma_n|^2)}{(1 - |\Gamma_s|^2)} = \frac{|A_{21}'|^2}{|A_{21}|^2}\frac{|1 - C_{11}\Gamma_n|^2}{|1 - C_{11}'\Gamma_s|^2} \cdot \frac{|1 - I_{22}A_{11}|^2|1 - \Gamma_2\Gamma_L|^2}{|1 - I_{22}A_{11}'|^2|1 - \Gamma_2'\Gamma_L|^2}(1 + \epsilon) \tag{18}$$

where

$$\epsilon \approx \frac{T_a|1 - C_{11}'\Gamma_s|^2|1 - \Gamma_2'\Gamma_L|^2}{(T_s - T_a)|C_{21}'|^2(1 - |\Gamma_s|^2)} \cdot \left\{\frac{1}{|1 - \Gamma_2'\Gamma_L|^2} - \frac{1}{|1 - \Gamma_2\Gamma_L|^2}\right\}. \tag{19}$$

The term including $\xi \approx 10^{-12}$ has been neglected in the approximation for ϵ.

Let

$$\frac{|A_{21}'|^2}{|A_{21}|^2} \equiv 10^{\Delta A/10}$$

and let T_{ne} be defined as follows:

$$(T_n - T_a)(1 - |\Gamma_n|^2) \approx T_n(1 - |\Gamma_n|^2) - T_a \equiv T_{ne} - T_a.$$

Using the two definitions given above it is easily shown that (18) can be written as

$$10\log_{10}\frac{(T_{ne} - T_a)}{(T_s - T_a)(1 - |\Gamma_s|^2)} = \Delta A + 20\log_{10}\left|\frac{1 - C_{11}\Gamma_n}{1 - C_{11}'\Gamma_s}\right| + 20\log_{10}\left|\frac{(1 - I_{22}A_{11})(1 - \Gamma_2\Gamma_L)}{(1 - I_{22}A_{11}')(1 - \Gamma_2'\Gamma_L)}\right| + 10\log_{10}(1 + \epsilon). \tag{20}$$

Appendix II

Calculation of the Attenuation of a High Temperature Waveguide

The field intensity in a rectangular waveguide of finite conductivity is known [11] to vary as $e^{j(\omega t - \gamma z)}$, where γ, the propagation constant, is defined as

$$\gamma = \gamma_0\left[1 + (1 - j)\frac{\mu_m\delta}{\mu\gamma_0^2}\left(\frac{\pi^2}{a^3} + \frac{k^2}{2b}\right)\right], \tag{21}$$

and

$$k^2 = \omega^2 \mu \epsilon$$

$$\gamma_0^2 = k^2 - \left(\frac{\pi}{a}\right)^2,$$

$$\mu_m = \mu_0(1 + \chi_m),$$

$$\delta = \left(\frac{2}{\sigma\mu\omega}\right)^{1/2} = \left(\frac{2\rho}{\mu\omega}\right)^{1/2}.$$

a = large dimension of the waveguide
b = small dimension of the waveguide
ω = angular frequency in radians/sec.
σ = conductivity of the waveguide walls
μ_m = permeability of the waveguide walls
μ = permeability of the medium in the waveguide
ϵ = dielectric constant of the medium in the waveguide
ρ = resistivity of the waveguide walls
δ = skin depth in the waveguide walls

The units for the above definitions are all mks units. The imaginary part of the propagation constant γ_ϵ causes a diminution of the field intensity and is related to the attenuation along the waveguide in db, A', by the definition

$$A' = 20(\log_{10} e)\gamma_i \Delta Z, \tag{22}$$

or

$$A'/\Delta z = 20\,(\log_{10} e)\gamma_i \text{ db/m},$$

where Δz is an incremental length along the waveguide.

$$\gamma_i = \text{Im}\,(\gamma) = \frac{\mu_m \delta}{\mu \gamma_0}\left(\frac{\pi^2}{a^3} + \frac{k^2}{2b}\right)$$

$$= \mu_m \frac{\left(\frac{\pi^2}{a^2} + \frac{k^2}{2b}\right)}{\left(k^2 - \frac{\pi^2}{a^2}\right)^{1/2}} \frac{(2\rho)^{1/2}}{(\omega\mu^3)^{1/2}}.$$

Then,

$$A'/\Delta z = 20(\log_{10} e)\gamma_i = \frac{20(\log_{10} e)\mu_m}{\mu}$$

$$\cdot \frac{\left(\frac{\pi^2}{a^3} + \frac{k^2}{2b}\right)}{\left(k^2 - \frac{\pi^2}{a^2}\right)^{1/2}} \frac{(2\rho)^{1/2}}{(\mu\omega)^{1/2}},$$

which can be written as

$$A'/\Delta z = C_1\left(\frac{\rho}{f}\right)^{1/2} \frac{f^2 + C_2}{(f^2 - C_3)^{1/2}},$$

if the constants C_1, C_2, and C_3 are adjusted so that $A'/\Delta z$ reads in db/inches, when the frequency is expressed in gigacycles per second and the resistivity is in micro-ohm-centimeters, their values are found to be $C_1 = 36.2 \times 10^{-5}$, $C_2 = 38.2$, and $C_3 = 43.0$ for WR-90 waveguide size.

Acknowledgment

The authors wish to acknowledge A. J. Estin for his contributions concerning the analysis of the reference standard during an earlier collaboration. The fruitful discussions with G. F. Engen which have helped further the state-of-the-art and his suggestions leading to a more compact error expression are greatly appreciated. Thanks are due to H. W. Lance for suggestions regarding the paper and to R. E. Larson for his interest and support of the project. The technical services of W. Foote and E. Campbell were indispensable in the completion of the present phase of the work.

References

[1] R. H. Dicke, "The measurement of thermal radiation at microwave frequencies," *Rev. Sci. Instr.*, vol. 17, no. 7, p. 268–275; July, 1946.
[2] A. J. Estin, C. L. Trembath, J. S. Wells, and W. C. Daywitt, "Absolute measurement of temperatures of microwave noise sources," IRE TRANS. ON INSTRUMENTATION, vol. I-9, pp. 209–213; September, 1960.
[3] A. W. Grosvenor, Ed. "Basic Metallurgy," American Society of Metals, Cleveland, Ohio; 1954.
[4] E. D. Zysk, "Platinum Metal Thermocouples," *Symp. on Temp. and Meas. Control in Science and Industry*, Columbus, Ohio; 1961.
[5] J. E. Sees, "Fundamentals in Noise Source Calibrations at Microwave Frequencies," Naval Research Lab., Washington, D. C., Rept. No. 5051; 1958.
[6] W. F. Roeser and H. T. Wensel, "Temperature—Its Measurement and Control in Science and Industry," Reinhold Publishing Corp., New York, N. Y., pp. 1293–1326; 1941.
[7] E. F. Northrup, "Resistivity of pure gold in temperature range 20°C to 1500°C," *J. Franklin Inst.*, vol. 177, no. 3, pp. 287–292; 1914.
[8] J. M. Richardson, private communication.
[9] G. E. Schafer and R. R. Bowman, "Preliminary report on a modulated subcarrier technique of measuring microwave attenuation," *Proc. IEE*, vol. 109, Part B, suppl. no. 23, pp. 783–786; May, 1962.
[10] R. W. Beatty, "Mismatch errors in the measurement of ultra-high-frequency and microwave variable attenuators," *J. Res. NBS*, vol. 52, pp. 7–9; January, 1954.
[11] D. M. Kerns and R. W. Hedberg, "Propagation constant in rectangular waveguide of finite conductivity," *J. Appl. Phys.*, vol. 25, p. 1550; 1954.

General References

D. A. Bell, "Electrical Noise," D. Van Nostrand Company, Ltd., London, England; 1960.
R. F. Vines and E. M. Wise, "The Platinum Metals and Their Alloys," The International Nickel Company, Inc., New York, N. Y.; 1941.

Reprinted from IEEE TRANSACTIONS
ON *INSTRUMENTATION AND MEASUREMENT*
Volume IM-13, Number 1, March, 1964

5.a. **Absolute measurement of temperatures of microwave noise sources,** A. J. Estin, C. L. Trembath, J. S. Wells, and W. C. Daywitt, IRE Trans, Instr. I-9, No. 2, 209-213 (September 1960).

5.b. **Sensitivity of a correlation radiometer,** John J. Faris, J. Res. NBS 71C, No. 2, 153-170 (Apr.-June 1967).

The correlation radiometer is analyzed to determine the sensitivity that can be obtained under various operating conditions. The radiometer using a sine wave comparison signal is analyzed and compared with the usual radiometer that employs a random noise for the comparison signal. It is found that the radiometer employing the sine wave comparison signal is the more sensitive of the two circuits, particularly in the case that the effective temperature of the input noise signal is greater than the effective input temperature of the amplifiers. It is shown that if nonidentical amplifiers are used in the correlation circuit, the properties of the radiometer are determined by the portion of the amplifier response functions in the frequency interval that the two response functions overlap. The effect of amplifier gain fluctuations are considered, and although the correlation scheme reduces the effect of gain fluctuations, it is shown that they still do contribute to the output fluctuations of the radiometer.

Calculations are included showing that the effect of a differential phase shift between the two channels is a reduction in radiometer sensitivity. The same conclusion is reached concerning the effect of a differential time delay.

Finally, it is shown that if the comparison signal and the input signal have the same statistical properties, the requirements on the multiplier are less stringent than if the two signals have different statistical properties.

5.c. **A waveguide noise-tube mount for use as an interlaboratory noise standard,** C. K. S. Miller, W. C. Daywitt, and E. Campbell, ACTA IMEKO III, 371-382 (1964).

5.d. **The sensitivity of the Dicke radiometer,** David F. Wait, J. Res. NBS 71C (Eng. & Instr.), No. 2, 127-152 (Apr.-June 1967).

The literature is reviewed concerning the sensitivity of the Dicke radiometer, excluding gain fluctuations. Discrepancies are pointed out and a new derivation of sensitivity using a Fourier transform method is used to resolve these discrepancies and to extend the results to radiometers with lossy switches.

Experimentally it is shown that radiometers using a half-wave square-law, linear-law, intermediate-law, or envelope detectors all have a sensitivity equal to the theoretical full-wave square-law detector (within the $\pm$ 20 percent uncertainty of the experiment).

385-582 O - 70 - 14

6. Attenuation and Phase

Papers

Abstracts

Insertion Loss Concepts

ROBERT W. BEATTY, SENIOR MEMBER, IEEE

Summary—The term "insertion loss" has become such a popular one and is applied to so many different concepts that it is becoming unsuitable for precise use. Certain other concepts connected with insertion loss such as attenuation and mismatch loss, are in need of clarification. A logical set of definitions and equations are presented as a suggested approach to more precise terminology. The effect of complex characteristic impedance on these equations is given in an Appendix. Specific suggestions are given for changes in the usage of terms, which if acceptable and adopted would improve the situation.

Introduction

THE TERM "insertion loss" has become such a popular one and is applied to so many different concepts that it is becoming unsuitable for precise use. For example, it is used to designate 1) the residual loss of an isolator, variable attenuator, or other component which ideally would have no residual loss, 2) the loss of power received by a nonreflecting load from a nonreflecting generator when a two-arm waveguide junction is inserted between them, 3) the loss of power received by an arbitrary load from an arbitrary generator when a two-arm waveguide junction is inserted between them, and 4) the loss in power received by an arbitrary load from an arbitrary generator when a two-arm waveguide junction is inserted in place of a lossless transducer which transfers all of the available power to the load.

Certain other concepts connected with insertion loss such as attenuation and mismatch loss are in need of clarification. For example, there is a tendency in some quarters for the term attenuation to be used to refer only to dissipative loss. Note, however, that the word attenuation denotes a reduction or weakening, and that in a circuit, a reduction in power can be caused either by dissipation or reflection. Therefore attenuation should not be limited to dissipative losses. The term "attenuation" has also been used to refer to the minimum loss obtainable by connecting lossless tuners and adjusting them for a bilateral nonreflecting match. This practice tends to confuse attenuation with the con-

Manuscript received August 19, 1963; revised February 12, 1964. This paper was presented at the URSI-IRE Spring Meeting in Washington, D. C., May 2, 1962.

The author is with the Radio Standards Laboratory, National Bureau of Standards, Boulder, Colo.

cept called "intrinsic insertion loss"[1] or "intrinsic attenuation."[2]

The term mismatch loss has been used to refer to the ratio of the power received by a load having an impedance equal to the complex conjugate of the generator impedance (the available power) to the power received by a specified different load later connected to the same generator. However it has been also used[3] to refer to the ratio of power received by a nonreflecting load (matched to the characteristic impedance of the waveguide[4]) to the power received by a specified different load later connected to the same generator. This latter concept has also been called "reflection loss."[5]

Although the above problems arise mainly in connection with microwave measurements, it is felt that the concepts and recommendations given are applicable and worthy of consideration at lower frequencies whenever waveguide and transmission lines are employed.

The purpose of this paper is not only to discourage practices which tend to confuse, but also to provide a logical and natural basis for new and revised definitions for selected loss concepts, and to suggest a more precise terminology. The names are intended to be descriptive, and where possible to conform to past usage. If these suggestions are found useful, they could make unnecessary the above confusing practices.

Loss Concepts—General

It is proposed in the following to introduce a term called "substitution loss" which is intended to be the most general loss concept involving a two-arm waveguide junction or two port. Other losses such as transducer loss, insertion loss and attenuation may be derived from it as special cases. Instead of actually deriving them in this way, they will be defined in terms of power ratios, so that each may also be considered independently of the others, if desired.

It is also proposed to introduce a term called "comparison loss" which is intended to be a general loss concept involving different loads connected to the same generator. Mismatch losses can be derived from it as special cases.

In deriving the equations for the various losses considered in this paper from the basic definitions, any set of network parameters could be chosen. In this paper, formulas are presented in terms of scattering coefficients,[6] which are particularly useful in the analysis of circuits in which waveguide (of any type) is employed. It is first assumed that the characteristic impedances of the waveguides are real, as would be the case if the waveguide were lossless. Later, this restriction is removed and, in the Appendix, corresponding equations for the case of complex characteristic impedances are presented.

Substitution Loss

Consider as in Fig. 1 the load power under two separate conditions. Initially a two-arm waveguide junction (two-port or two terminal-pair network) is connected between the generator and the load. Its scattering coefficients ${}^{i}S_{11}$, ${}^{i}S_{12}$, ${}^{i}S_{21}$, and ${}^{i}S_{22}$ have the front superscript i to denote the fact that this is the junction initially connected. When this junction is removed and a different waveguide junction is substituted in its place, net power transmitted[7] to the load changes from ${}^{i}P_T$ to ${}^{f}P_T$, where the front superscript f denotes the final condition. The substitution loss is defined as

$$L_s = 10 \log_{10} \frac{{}^{i}P_T}{{}^{f}P_T} . \tag{1}$$

In terms of the scattering coefficients of the two waveguide junctions and the reflection coefficients Γ_G and Γ_L of the generator and the load, the substitution loss is[8]

$$L_s = 20 \log_{10} \cdot \left| \frac{{}^{i}S_{21}[(1 - {}^{f}S_{11}\Gamma_G)(1 - {}^{f}S_{22}\Gamma_L) - {}^{f}S_{12}{}^{f}S_{21}\Gamma_G\Gamma_L]}{{}^{f}S_{21}[(1 - {}^{i}S_{11}\Gamma_G)(1 - {}^{i}S_{22}\Gamma_L) - {}^{i}S_{12}{}^{i}S_{21}\Gamma_G\Gamma_L]} \right| . \tag{2}$$

It has been assumed in deriving (2) that the generator and load characteristics are the same finally as they

[1] K. Tomiyasu, "Intrinsic insertion loss of a mismatched microwave network," IRE Trans. on Microwave Theory and Techniques, vol. MTT-3, pp. 40–44; January, 1955.

[2] R. W. Beatty, "Intrinsic attenuation," IEEE Trans. on Microwave Theory and Techniques, vol. MTT-11, pp. 179–182; May, 1963.

[3] This usage has occurred in manufacturer's literature. For example, see the catalog of Sperry Microline 1960 Test Instruments, p. 55.

[4] The term "Waveguide" is used in this paper in the general sense recommended in IRE standards, thus including coaxial and two-wire transmission lines, for example.

[5] P. J. Selgin, "Electrical Transmission in Steady State," McGraw-Hill Book Co. Inc., New York, N. Y., 1st ed., pp. 112–120; 1946.

[6] The scattering coefficients are defined to be those coefficients appearing in the scattering equations

$$b_1 = S_{11}a_1 + S_{12}a_2, \quad \text{and} \quad b_2 = S_{21}a_1 + S_{22}a_2$$

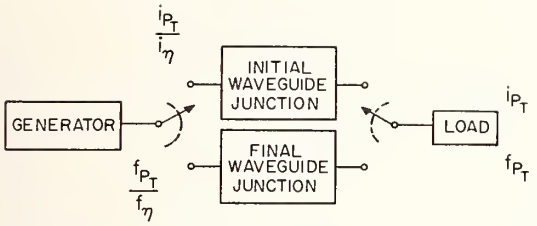
where the b's and a's denote the complex voltage wave amplitudes, and the subscripts 1 and 2 refer to terminal surfaces in the waveguides as shown in Fig. 2. Waves incident on the waveguide junction are designated by the a's and waves emergent from the waveguide junction are designated by the b's. Instead of the more usual normalization scheme, one has been chosen in which the characteristic impedances of the waveguides appear explicitly in some of the loss equations. This has been done for the sake of generality. The more general expressions given enable one more easily to deal with waveguide junctions having dissimilar arms, such as rectangular waveguide and coaxial line, for example. If desired, simpler expressions corresponding to the usual normalization scheme can be obtained by substituting

$$S_{12} = \sqrt{\frac{Z_{01}}{Z_{02}}}\, S_{12}', \quad \text{and} \quad S_{21} = \sqrt{\frac{Z_{02}}{Z_{01}}}\, S_{21}',$$

where the primes denote the normalization scheme which is widely used, although seldom discussed or acknowledged.

[7] This is the power dissipated in the load and is also known as the power delivered to the load.

[8] This is the same as the change of insertion loss ΔL in (3) of R. W. Beatty, "Mismatch errors in the measurement of ultrahigh-frequency and microwave variable attenuators," *J. Research NBS*, vol. 52, pp. 7–9; January, 1954.



i_η = EFFICIENCY OF INITIAL WAVEGUIDE JUNCTION

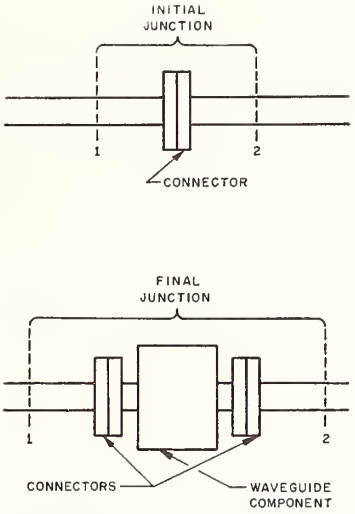
f_η = EFFICIENCY OF FINAL WAVEGUIDE JUNCTION

SUBSTITUTION LOSS = $10 \log_{10} \frac{i_{P_T}}{f_{P_T}}$

i_{P_T} = POWER TRANSMITTED TO (DISSIPATED BY) LOAD WHEN INITIAL WAVEGUIDE JUNCTION IS INSERTED

f_{P_T} = POWER TRANSMITTED TO LOAD WHEN FINAL WAVEGUIDE JUNCTION IS INSERTED

Fig. 1—Substitution loss.



INSERTION OF A WAVEGUIDE COMPONENT INTO A WAVEGUIDE SYSTEM IS A SUBSTITUTION PROCESS

Fig. 2—Insertion of a waveguide component into a waveguide system.

were in the initial condition. The expression applies to cases in which the two arms of the waveguide junctions may be dissimilar or have different propagating modes, as well as to the more usual case in which they are identical. It is apparent that substitution loss is a function of the generator and load impedances or reflection coefficients and therefore is associated with the values of these impedances at the time that it is determined.

The substitution loss as defined above may range from $-\infty$ to $+\infty$, in principle, and upon assuming negative values, could be regarded as a gain, rather than a loss. This continues to hold true even when we exclude "active networks" such as amplifiers. If one specifies certain characteristics for the initial waveguide junction or for the generator and load, the range over which the loss varies may be restricted to positive values, in some cases, to be discussed.

Actually the substitution loss most closely corresponds to what one can measure, since even if initially no waveguide junction is placed between generator and load, one must still have a waveguide joint or connector. When very accurate measurements are to be made, it is not permissible to neglect the reflection and dissipative loss from the connector, so that it must be considered as the initial waveguide junction. Thus as shown in Fig. 2, one always measures a substitution loss when attempting to measure the insertion loss or attenuation of a waveguide junction.

A step attenuator, in which one attenuator is removed and another inserted in its place, is an excellent example of the need for the concept of substitution loss. A smoothly variable attenuator in operation may be regarded as though one removes an initial attenuator (corresponding to the initial setting) and substitutes in its place another attenuator (corresponding to a final setting), although the variable attenuator is not physically removed from the circuit. Thus it is analytically equivalent[9] to a step attenuator and the substitution loss concept applies.

Transducer Loss[10]

As a special case of substitution loss, consider that the initial waveguide junction is a perfect transducer, *i.e.*, transmits all of the available[11] generator power P_A to the load. In order to do this, the perfect transducer must be lossless and in addition transform the load impedance to equal the complex conjugate of the generator impedance. Removal of the perfect transducer and substitution of another waveguide junction as in Fig. 3 will always decrease the power P_T transmitted to the load, unless the second junction is also a perfect transducer, or an active device such as an amplifier. The transducer loss is

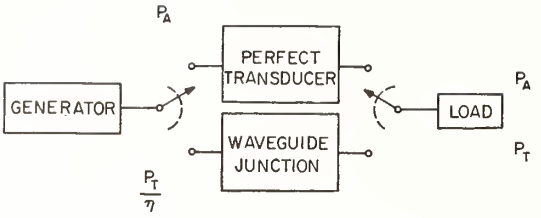
$$L_{TD} = 10 \log_{10} \frac{P_A}{P_T} . \tag{3}$$

This loss may be expressed in terms of the scattering coefficients S_{11}, S_{12}, S_{21} and S_{22} of the final waveguide junction, the reflection coefficients Γ_G and Γ_L of the generator and load, and the real characteristic impedances Z_{01} and Z_{02} of the input and output waveguide arms of the junction. The expression is as follows.

[9] See also p. 7 of the reference cited in Beatty.[8]

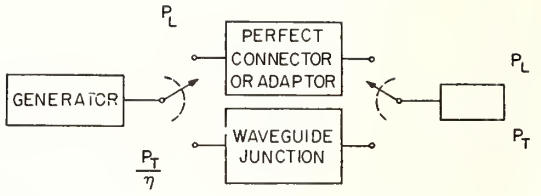
[10] The concept of transducer loss is defined in a similar sense in "Reference Data for Radio Engineers," International Telephone and Telegraph Corp., New York, N. Y., 4th ed., ch. 20, p. 569; 1956.

[11] A variable load connected to a fixed generator will absorb (dissipate) maximum power when the load impedance is the complex conjugate of the generator impedance. This maximum power is called the available power of the generator.



$$\text{TRANSDUCER LOSS} = 10 \log_{10} \frac{P_A}{P_T}$$

Fig. 3—Transducer loss.



$$\text{INSERTION LOSS} = 10 \log_{10} \frac{P_L}{P_T}$$

Fig. 4—Insertion loss.

$$L_{TD} = 10 \log_{10} \left[\frac{Z_{02}}{Z_{01}} \cdot \frac{|(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L|^2}{(1 - |\Gamma_G|^2)|S_{21}|^2(1 - |\Gamma_L|^2)} \right]. \quad (4)$$

It follows from the above definition and (3) that the transducer loss of a *passive* waveguide junction cannot be negative. It is a measure of how closely the performance of a waveguide junction approaches that of a perfect transducer between a given generator and load. When it is zero, there is no difference between them, and the waveguide junction is a perfect transducer.

Although it is felt to be a useful concept for purposes of analysis and calculation, attempts to measure the transducer loss will be in error by the amount by which the initial junction fails to be a perfect transducer, unless this quantity can be accurately determined or estimated.

One can apply this concept to a waveguide junction in which the waveguide arms are dissimilar, or propagating different modes, or both. Thus it would be of value for example, in situations where a rectangular waveguide to coaxial waveguide adapter was used. It cannot be used to specify a property of the junction unless the generator and load characteristics are also specified.

Insertion Loss

Another special case of substitution loss is the insertion loss as illustrated in Fig. 4. It is different from transducer loss in that the initial waveguide junction, although lossless, does not provide a conjugate impedance to match to the generator. One supposes that the initial waveguide junction is an ideal connector or adaptor; *i.e.*, it is nonreflecting ($S_{11}=S_{22}=0$) and lossless, and its phase shift (in a nonreflecting system) is an integral number of 2π radians. The effect of this initial junction is thus almost as if it were not there, for the reflection coefficient at its input is the same as the reflection coefficient of the load.

Insertion of the final waveguide junction between the generator and load will usually change the net power to the load. The ratio of initial load power P_L to the final load power P_T, expressed in decibels, is the insertion loss,

$$L_I = 10 \log_{10} \frac{P_L}{P_T}. \quad (5)$$

It may be expressed as follows:[12]

$$L_I = 10 \log_{10} \left[\frac{Z_{02}}{Z_{01}} \cdot \left| \frac{(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L}{S_{21}(1 - \Gamma_G\Gamma_L)} \right|^2 \right], \quad (6)$$

where the symbols have the usual meanings. If, in the initial connection, the load does not provide a conjugate match to the generator impedance and consequently does not absorb all of the available power, there is a possibility that the insertion loss might be negative. This could be the case, for example, if the waveguide junction under consideration were a perfect transducer.

The concept of insertion loss may be useful when one can safely neglect the effect of an imperfect waveguide connector or adaptor. This idealized condition may perhaps be more closely approached in practice than that of the perfect transducer.

Although one could regard the substitution loss as a change in either transducer loss or insertion loss, it is felt to be more fundamental to regard those as special cases of substitution loss which is more general and involves fewer implicit assumptions.

It is clear from inspection of (2), (4) and (6), that insertion loss, like transducer loss and substitution loss, depends in general not only upon the characteristics of the waveguide junction under consideration, but also upon the reflection coefficients of the generator and load

[12] This is similar to (1) of the reference cited in Beatty.[8]

of the system. Thus, these quantities are not in general suitable by themselves for designating a property of a waveguide junction for specification purposes.

Attenuation

A quantity which is useful for specification purposes is the attenuation of a waveguide junction. It is defined to be the insertion loss of the junction when the system is nonreflecting ($\Gamma_G = \Gamma_L = 0$). It is the ratio of initial to final load powers, expressed in decibels, when the waveguide junction is substituted for a perfect connector or adaptor, and the generator and load impedances have been adjusted so that the system is nonreflecting. Using the symbols of Fig. 5, the attenuation may be written

$$A = 10 \log_{10} \frac{P_0}{P_{T_0}} \cdot \tag{7}$$

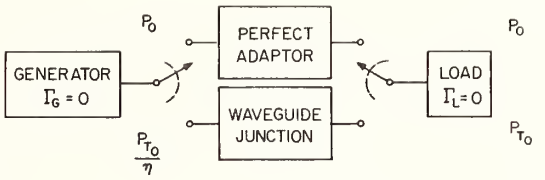
Substitution of the nonreflecting conditions into (4) yields the expression.

$$A = 10 \log_{10} \left(\frac{Z_{02}}{Z_{01}} \cdot \frac{1}{|S_{21}|^2} \right). \tag{8}$$

It is apparent that the attenuation is a property of the waveguide junction and is independent of the generator and load impedances. This concept can be applied to waveguide junctions having dissimilar arms, or similar arms with similar or dissimilar propagating modes. One should note that although this is a useful concept for purposes of analysis or calculation, the specified conditions, *i.e.*, an initially connected perfect connector or adaptor, and a nonreflecting system, cannot be strictly realized in practice, and that measured results will therefore be in error from this cause to some degree.

In the special but common case when the waveguide junction has identical waveguide arms and the same mode propagates in each, it is natural to choose $Z_{01} = Z_{02}$, and the attenuation power ratio equals the reciprocal of $|S_{21}|^2$. This concept has been named "insertion loss" in the Institute of Radio Engineer's Standards on Antennas and Waveguides: Waveguide and Waveguide Component Measurements.[13] However it is also there recognized that a more general definition of insertion loss does not specify a nonreflecting system or identical waveguide arms or propagating modes. Unfortunately this opens the door to confusion, since one is permitted under this "standard" to call both concepts (which may correspond to different quantities) by the same name. In order to avoid confusion, the author recommends that they be called by different names, *i.e.*, the ones given in this paper.

[13] "IRE standards on antennas and waveguides: waveguide and waveguide component measurements, 1959," Proc. IRE, vol. 47, pp. 568–582; April, 1959.



$$\text{ATTENUATION LOSS} = 10 \log_{10} \frac{P_0}{P_{T_0}}$$

Fig. 5—Attenuation.

Other Loss Concepts

One can define various other losses in terms of power ratios, using available power, incident power, reflected power, net power (delivered power), etc. One can also specify different initial waveguide junctions, such as one having the equivalent circuit of an ideal transformer. The scope of this paper has been limited to the above four concepts which are regarded by the author as especially fundamental and valuable in microwave circuit analysis.

Components of Losses

It is convenient and enlightening to separate each of the above losses into two components, one associated with the dissipation of energy within the waveguide junction, and the other with the reflection of energy. Since the loss concepts are each special cases of the substitution loss, their components are corresponding special cases of the components of substitution loss. In the following, this will be kept in mind, at the same time defining the components in terms of power ratios, where applicable.

Components of Substitution Loss

If one refers back to Fig. 1 and (1) and designates the efficiency of a waveguide junction by η [of (21)], the substitution loss is separated into components as follows:

$$L_s = 10 \log_{10} \frac{{}^{i}P_T}{{}^{f}P_T} = 10 \log_{10} \left(\frac{{}^{i}\eta \; {}^{i}P_1}{{}^{f}\eta \; {}^{f}P_1} \right)$$

$$= 10 \log_{10} \left(\frac{{}^{i}\eta}{{}^{f}\eta} \right) + 10 \log_{10} \left(\frac{{}^{i}P_1}{{}^{f}P_1} \right). \tag{9}$$

The first term involving only the efficiencies is the component associated with dissipation and the second term involving the net power inputs to the junctions is the component associated with reflection. They are written in terms of the coefficients previously used as follows:

$$(L_s)_D = 10 \log_{10}\left(\frac{{}^i\eta}{{}^f\eta}\right)$$

$$= 10 \log_{10}\left(\left|\frac{{}^iS_{21}}{{}^fS_{21}}\cdot\frac{1-{}^fS_{22}\Gamma_L}{1-{}^iS_{22}\Gamma_L}\right|^2\cdot\frac{1-|{}^f\Gamma_1|^2}{1-|{}^i\Gamma_1|^2}\right) \quad (10)$$

and

$$(L_s)_R = 10 \log_{10}\left(\frac{{}^iP_1}{{}^fP_1}\right)$$

$$= 10 \log_{10}\left(\left|\frac{1-\Gamma_G{}^f\Gamma_1}{1-\Gamma_G{}^i\Gamma_1}\right|^2\frac{1-|{}^i\Gamma_1|^2}{1-|{}^f\Gamma_1|^2}\right). \quad (11)$$

The substitution loss is given by adding the above equations, obtaining

$$L_s = 20 \log_{10}\left|\frac{{}^iS_{21}}{{}^fS_{21}}\cdot\frac{1-{}^fS_{22}\Gamma_L}{1-{}^iS_{22}\Gamma_L}\cdot\frac{1-\Gamma_G{}^f\Gamma_1}{1-\Gamma_G{}^i\Gamma_1}\right|. \quad (12)$$

It is apparent that (12) is equivalent to (2) if the following equation is substituted.

$$\Gamma_1 = \frac{(S_{12}S_{21}-S_{11}S_{22})\Gamma_L + S_{11}}{1-S_{22}\Gamma_L}. \quad (13)$$

In (10)–(13) and throughout the paper, Γ_1 is the reflection coefficient at terminal surface number 1 equal to the ratio of b_1 to a_1.

Components of Other Losses

The components of the transducer loss, insertion loss, and attenuation associated with dissipation and with reflection, are shown in the table of Fig. 6. In each case it is noted that the initial waveguide junction is lossless (${}^i\eta = 1$). It will also be shown that the components of loss associated with mismatch can be regarded as a change of mismatch loss, to be defined.

Comparison Loss

When two different loads are alternately connected to the same stable generator, the ratio of the power ${}^iP_1$ dissipated in the load initially connected to the power ${}^fP_1$ dissipated in the load finally connected is (expressed in decibels)

$$L_c = 10 \log_{10}\frac{{}^iP_1}{{}^fP_1}$$

$$= 10 \log_{10}\left(\left|\frac{1-\Gamma_G{}^f\Gamma_1}{1-\Gamma_G{}^i\Gamma_1}\right|^2\frac{1-|{}^i\Gamma_1|^2}{1-|{}^f\Gamma_1|^2}\right). \quad (14)$$

This ratio is widely used[14] in analysis of mismatch errors in power measurements. If one is given the power dissipated in the load initially connected, and this power is expressed in decibels referred to some convenient level, then one subtracts L_c to obtain the power dissipated in the different load finally connected.

[14] R. W. Beatty and A. C. Macpherson, "Mismatch errors in microwave power measurements," Proc. IRE, vol. 41, pp. 1112–1119; September, 1953.

LOSS	COMPONENT ASSOCIATED WITH: MISMATCH	COMPONENT ASSOCIATED WITH: DISSIPATION
TRANSDUCER LOSS	$10 \log_{10}\frac{\lvert 1-\Gamma_G\Gamma_1\rvert^2}{(1-\lvert\Gamma_G\rvert^2)(1-\lvert\Gamma_1\rvert^2)}$	$10 \log_{10}\frac{1}{\eta}$
INSERTION LOSS	$10 \log_{10}\left[\left\lvert\frac{1-\Gamma_G\Gamma_1}{1-\Gamma_G\Gamma_L}\right\rvert^2\frac{1-\lvert\Gamma_L\rvert^2}{1-\lvert\Gamma_1\rvert^2}\right]$	$10 \log_{10}\frac{1}{\eta}$
ATTENUATION	$10 \log_{10}\frac{1}{1-\lvert S_{11}\rvert^2}$	$10 \log_{10}\frac{1-\lvert S_{11}\rvert^2}{\lvert S_{21}\rvert^2}$
NOTE THAT $\eta = \frac{Z_{01}}{Z_{02}}\cdot\frac{\lvert S_{21}\rvert^2(1-\lvert\Gamma_L\rvert^2)}{\lvert 1-S_{22}\Gamma_L\rvert^2-\lvert(S_{12}S_{21}-S_{11}S_{22})\Gamma_L+S_{11}\rvert^2}$		

Fig. 6—Components of losses.

Usually ${}^i\Gamma_1$ is specified, but one must then know in addition Γ_G and ${}^f\Gamma_1$ in order to determine L_c.

If the specified load is a "matched load," then L_c is the loss of power due to mismatch. Since more than one interpretation of the term "matched load" is possible, one can define more than one mismatch loss.

Mismatch Loss

In the following discussion, it is helpful to recall a pertinent ordinary meaning of the word match: "to adapt, fit, or suit (one thing to another)." In electrical circuits, the maximum transfer of power from a generator to a load requires matching of the impedances in the circuit. When only resistances are involved, the maximum power is obtained from a generator when the load resistance is identical with (matches) the generator internal resistance. However when the generator impedance is complex, maximum power requires that the load impedance be the complex conjugate of the internal impedance of the generator. Such a condition is called a conjugate match.

When the electrical circuits involve waveguides and transmission lines, it is found that standing waves due to reflection of energy are eliminated when the impedance of the load terminating the transmission line equals (matches) the characteristic impedance of the line. If the internal impedance of the generator also equals the characteristic impedance of the line, then the system is nonreflecting, or matched in both directions. However if the generator does not match the line, then the load can match either the line or the generator, but not both.

One can also consider impedance matching of networks on an image or an iterative basis, so that the terms "match" or "mismatch" taken by themselves are not specific enough for precise use.

There are two types of mismatch loss which might be considered useful. One is the usual mismatch loss,

here called the conjugate mismatch loss, which is useful when one is given the available power P_A from a given generator, and wishes to determine the power P_1 that it will deliver to some specified load. If the powers are given in decibels relative to some convenient level, one subtracts the conjugate mismatch loss from the available power to obtain the power absorbed by the load.

If one is given, instead, the power P_0 that the generator will deliver to a nonreflecting load, and wishes to determine the power P_1 delivered to some specified load, one subtracts the Z_0-mismatch loss from P_0. This is a useful concept because power meters are often designed to be nonreflecting so that one is often given or can easily determine P_0.

Each of these types of mismatch loss can be considered as a special case of the more general comparison loss. This approach has been used by Selgin[5] in discussing similar loss concepts having different names, and is consistent with the approach used in this paper of defining attenuation, insertion loss, and transducer loss in terms of the more general substitution loss.

Conjugate Mismatch Loss

The conjugate mismatch loss is the ratio, expressed in decibels, of the available power from the generator to the power absorbed by the load connected to the generator. In terms of impedances, using the usual symbols, it has the familiar form

$$M_c = 10 \log_{10} \frac{|Z_G + Z_1|^2}{4R_G R_1} . \tag{15}$$

If the same expression is desired in terms of reflection coefficients and if the characteristic impedance is real, we then obtain

$$M_c = 10 \log_{10} \frac{|1 - \Gamma_G \Gamma_1|^2}{(1 - |\Gamma_G|^2)(1 - |\Gamma_1|^2)} , \tag{16}$$

where Γ_1 is the reflection coefficient of the mismatched load. The conjugate mismatch loss is, from the above definition, never negative, and is identical with the component of transducer loss associated with reflection (see Fig. 6). This is almost universally called "mismatch loss" in the literature.

Z_0-Mismatch Loss

The Z_0-mismatch loss is the ratio (expressed in decibels) of the power that would be absorbed by a nonreflecting load if it were connected to the generator, to the power absorbed by the load which is connected to the generator. This loss may be written

$$M_{Z_0} = 10 \log_{10} \frac{|1 - \Gamma_G \Gamma_1|^2}{1 - |\Gamma_1|^2} . \tag{17}$$

It is possible for this loss to be negative if the internal impedance of the generator does not match the characteristic impedance Z_0 of the waveguide or transmission line. For example, if the load under consideration provides a conjugate match and absorbs all of the power, the load which matches the Z_0 of the waveguide will absorb less power, and hence M_{Z_0} must be negative.

In the special case when the generator does match the line ($\Gamma_G = 0$), the Z_0-mismatch loss is

$$(M_{Z_0})_{\Gamma_G = 0} = 10 \log_{10} \frac{1}{1 - |\Gamma_1|^2} . \tag{18}$$

If, in addition, the load under consideration consists of a two-arm waveguide junction terminated in a nonreflecting load, then $\Gamma_1 = S_{11}$ and (18) becomes identical with the component of attenuation associated with reflection (see Fig. 6).

The conjugate mismatch loss and the Z_0-mismatch loss can both be called "mismatch loss."[15] In order to distinguish between them, the author recommends that the modifiers "conjugate" and "Z_0-" be used, as above. As shown below, they are not in general the same except in cases in which the generator is nonreflecting.

Difference Between Conjugate and Z_0-Mismatch Losses

It has been observed that when the internal impedance of the generator matches the characteristic impedance of the transmission line ($\Gamma_G = 0$), the conjugate mismatch loss and the Z_0-mismatch loss are identical and are given by (18). When this is not the case, the difference between these two losses is

$$M_{C-Z_0} = M_C - M_{Z_0} = 10 \log_{10} \frac{1}{1 - |\Gamma_G|^2} . \tag{19}$$

It is the ratio (expressed in decibels) of the available power from a generator to the power that a nonreflecting load would absorb when connected to this generator.

Return Loss

A measure of the loss in power returned to the source by a reflecting load is given by the return loss L_R. It is written[16]

$$L_R = 20 \log_{10} \frac{1}{|\Gamma_1|} . \tag{20}$$

In the case of total reflection, the return loss is zero, whereas it becomes infinite when the reflection approaches zero. This is not to be confused with mismatch losses which were previously defined.

[15] Selgin[5] has called the Z_0-mismatch loss by the name "reflection loss." However, it has also been named "mismatch loss" in recent manufacturer's literature, see for example the 1960 Sperry Microline Catalog, p. 55, and Hewlett-Packard Application Note, no. 38, sec. III, p. 1; 1960.

[16] R. W. Beatty and W. J. Anson, "Table of Magnitude of Reflection Coefficient Versus Return Loss,"

$$\left(L_R = 20 \log_{10} \frac{1}{|\Gamma|} \right),$$

NBS, Tech. Note 72; September 19, 1960.

Losses Associated with Dissipation

A number of losses have been defined which are associated with dissipation of energy within a two-arm waveguide junction. The case usually considered is that in which energy is fed into one arm of the junction, and some of this energy passes out the other arm.

The ratio expressed in decibels of the net power input to the net power output is the dissipative loss L_D of the junction. It may be written

$$L_D = 10 \log_{10} \frac{1}{\eta}, \tag{21}$$

where η is the efficiency. It is a measure of the dissipation occurring within the waveguide junction. When there is no dissipation, all of the net power input is transmitted to the load, and there is no loss in transmission.

This dissipative loss is the same as either the dissipative component of transducer loss, insertion loss or attenuation, whichever load condition applies. The efficiency is[17]

$$\eta = \frac{Z_{01}}{Z_{02}} \cdot \frac{|S_{21}|^2(1 - |\Gamma_L|^2)}{|1 - S_{22}\Gamma_L|^2 - |(S_{12}S_{21} - S_{11}S_{22})\Gamma_L + S_{11}|^2}, \tag{22}$$

showing that it depends in general not only upon the properties of the waveguide junction, but also upon the reflection coefficient Γ_L of the load.

In the special case when $\Gamma_L = 0$ the efficiency equals

$$\eta_A = \frac{Z_{01}}{Z_{02}} \cdot \frac{|S_{21}|^2}{1 - |S_{11}|^2}. \tag{23}$$

Comparison with Fig. 6 shows that (23) is related to the component of attenuation associated with dissipation. It is evident that attenuation as defined in this paper is not strictly dissipative, but involves reflection as well.

Conclusions

In the author's opinion the restriction of the term attenuation to purely dissipative losses should be discouraged and also the tendency to speak of mismatch loss without specifying what is meant by mismatch should be similarly discouraged.

On the constructive side, a basis was given for precisely defining a selected number of loss concepts, such as substitution loss, transducer loss, insertion loss and attenuation. The components of these losses associated with dissipation and with reflection were defined and equations were presented. The author's recommendations for the changed usage of terms are summarized in Fig. 7.

Thus it is hoped that use of a more precise terminology where losses are concerned will be encouraged, as this will reduce confusion and lead to more precise thinking and to more precise measurements as well.

Common Usage	Recommended
Insertion Loss:	*Instead of Insertion Loss, Use:*
1) Referring to the minimum or initial loss of a variable attenuator, or isolator; or to loss of a component which is ideally lossless.	1) Residual loss (numerically equal to attenuation below).
2) Referring to loss[13] of a component designed to produce a loss, such as an attenuator.	2) Attenuation (equal to the insertion loss in a nonreflecting system).
Attenuation:	*Instead of Attenuation, Use:*
1) Referring only to dissipative loss.	1) Dissipative loss.
2) Attenuation of a variable attenuator; the difference between a given setting and the zero setting.	2) Incremental attenuation.

Fig. 7—Some recommended changes in terminology.

Appendix

It is usual to choose the characteristic impedances of lossless waveguides to be pure real numbers. However it has been shown[18] that the effect of losses is often to add a small reactive component X_0, while the resistive component R_0 may be slightly different than the characteristic impedance of a lossless waveguide of the same dimensions.

One can take into account the effect of a complex characteristic impedance by straightforward circuit analysis, obtaining the following equations. Reflection coefficients are related to impedances by equations of the form

$$\Gamma = \frac{Z - Z_0}{Z + Z_0}.$$

Substitution Loss: See (2). There is no change.

[17] The derivation is straightforward, starting with the familiar equation for power $P = \mathrm{Re}\,(i^* v)$, where $Z_0 i = a - b$, and $v = a + b$, and a and b are the incident and emergent wave amplitudes. For example,

$$P_1 = \frac{|a_1|^2}{Z_{01}}(1 - |\Gamma_1|^2) \quad \text{and} \quad P_L = \frac{|b_2|^2}{Z_{02}}(1 - |\Gamma_L|^2).$$

The efficiency is the ratio of P_L to P_1 and Γ_1 is obtained from (13).

[18] See for example Eq. (62), p. 23, of "Microwave Transmission Circuits," George L. Ragan, M.I.T. Rad Lab. Ser., McGraw-Hill Book Co. Inc., New York, N. Y., vol. 9; 1948.

Transducer Loss:

$$L_{TD_1} = 10 \log_{10}\left[\frac{R_{02}}{R_{01}} \cdot \frac{|(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L|^2}{|S_{21}|^2(1 - |\Gamma_G|^2)(1 - |\Gamma_L|^2)}\right]$$

$$+ 10 \log_{10}\left[\frac{1 + \left(\frac{X_{02}}{R_{02}}\right)^2}{\left(1 + j\frac{X_{01}}{R_{01}} \cdot \frac{\Gamma_G - \Gamma_G^*}{1 - |\Gamma_G|^2}\right)\left(1 + j\frac{X_{02}}{R_{02}} \cdot \frac{\Gamma_L - \Gamma_L^*}{1 - |\Gamma_L|^2}\right)}\right]. \tag{24}$$

Insertion Loss:

$$L_{I_1} = 10 \log_{10}\left[\frac{R_{02}}{R_{01}} \frac{|(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L|^2}{|S_{21}|^2 \cdot |1 - \Gamma_G\Gamma_L|^2}\right]$$

$$+ 10 \log_{10}\left[\frac{1 + \left(\frac{X_{02}}{R_{02}}\right)^2}{1 + \left(\frac{X_{01}}{R_{01}}\right)^2} \cdot \frac{1 + j\frac{X_{01}}{R_{01}} \frac{\Gamma_L - \Gamma_L^*}{1 - |\Gamma_L|^2}}{1 + j\frac{X_{02}}{R_{02}} \frac{\Gamma_L - \Gamma_L^*}{1 - |\Gamma_L|^2}}\right]. \tag{25}$$

Attenuation:

$$A_1 = 10 \log_{10}\left(\frac{R_{02}}{R_{01}} \cdot \frac{1}{|S_{21}|^2}\right) + 10 \log_{10}\left[1 + \left(\frac{X_{02}}{R_{02}}\right)^2\right]$$

$$- 10 \log_{10}\left[1 + \left(\frac{X_{01}}{R_{01}}\right)^2\right]. \tag{26}$$

Comparison Loss: See (14). There is no change.

Conjugate Mismatch Loss:

$$M_c = 10 \log_{10}\left[\frac{|1 - \Gamma_G\Gamma_1|^2}{(1 - |\Gamma_G|^2)(1 - |\Gamma_1|^2)}\right]$$

$$+ 10 \log_{10}\left[1 + \left(\frac{X_{01}}{R_{01}}\right)^2\right]$$

$$+ 10 \log_{10}\left[\frac{1 + j\frac{X_{01}}{R_{01}} \cdot \frac{\Gamma_1 - \Gamma_1^*}{1 - |\Gamma_1|^2}}{1 + j\frac{X_{01}}{R_{01}} \cdot \frac{\Gamma_G - \Gamma_G^*}{1 - |\Gamma_G|^2}}\right]. \tag{27}$$

Z_0-Mismatch Loss:

$$M_{Z_0} = 10 \log_{10}\frac{|1 - \Gamma_G\Gamma_1|^2}{1 - |\Gamma_1|^2}$$

$$- 10 \log_{10}\left[1 + j\frac{X_{01}}{R_{01}} \frac{\Gamma_1 - \Gamma_1^*}{1 - |\Gamma_1|^2}\right]. \tag{28}$$

Note that the first terms of the equations are of the same form as those given for pure real Z_0, and that the subsequent terms take into account the effects of complex Z_0.

The subsequent terms contain factors of the form

$$1 + j\frac{X}{R} \frac{\Gamma - \Gamma^*}{1 - |\Gamma|^2},$$

and of the form $1+(X/R)^2$. The quantity X/R appears to the first order in one and to the second order in the other. Its effect can be greater in the first term provided that the reflection coefficient Γ has a sufficiently large component along the imaginary axis.

For example, lossy cable having a nichrome center conductor may have an attenuation of 44 decibels per 100 feet at 3 Gc. The ratio of α to β[19] and of X_0 to R_0 is then approximately 0.0027. If $\Gamma = j0.5$, a term of the first form above would have an effect of 0.016 db, and a term of the second form would have an effect of less than 0.001 db. Thus if lossy cable were used on one side of the insertion point, but not on the other, the effect on the attenuation would be less than 0.001 db, and if used on both sides, there would be no effect.

These effects are thus negligible in many cases, but could become important at lower frequencies if the ratio of α to β increases, and if the precision of the measurements increases.

Acknowledgment

The author has had the benefit of suggestions from many National Bureau of Standards Senior Staff Members, whose assistance is gratefully acknowledged.

[19] α and β are respectively the real and imaginary components of the propagation constant of the waveguide.

Reprinted from the PROCEEDINGS OF THE IEEE
Vol. 52, No. 6, June, 1964

RF Attenuation

DAVID RUSSELL AND WILBUR LARSON

Abstract—A tutorial review of RF attenuation measurement methods and standards is presented.

Accepted and proposed definitions and attenuator models are discussed. Commonly used standards operating from dc through most waveguide bands are compared with the "ideal" interlaboratory standard. Characteristics of fixed resistive, waveguide below-cutoff, and rotary-vane standards are included.

Measurement methods are classified and described, including comments concerning convenience and accuracy of various methods, and references are given which cover most of the basic important research in the field.

I. Introduction

A CONSIDERABLE amount of material has been published on the subject of attenuation measurement and standards. In an attempt to provide a compilation of methods and standards, the present paper is divided into five areas. Attenuator models and definitions of terms are included to clarify existing confusion on exactly what is being measured. A comprehensive listing of attenuator types and the features of those most commonly used form a major segment of this discussion. Interlaboratory standards are classified by mode of operation (coaxial or waveguide) rather than by frequency. Qualities of the ideal interlaboratory standard are compared with existing standards. Various methods of attenuation measurement are described, including methods of measurement which do not require reference to a standard attenuator. The final section describes the many errors of systematic and random nature that affect attenuation measurements. Suggestions are offered for reduction of some and for estimating the magnitude of others not easily reduced. Some of the material in this paper, particularly in Section III-B on power ratio standards and Sections V-A, V-B, and V-C on measurement methods, is adapted from NBS Monograph 97 [1].

II. Attenuation Models and Definitions

The term attenuation means a decrease in power from one point in a transmission system to another, usually resulting from either dissipative or reflective losses. This can be a characteristic of a transmission line expressed in decibels or nepers per unit length, or an attribute of a particular device. To adequately describe attenuation measurement and devices to be measured it is necessary to choose a model for an attenuator that will closely approximate what is encountered in practice. With such a model one can define precisely what is to be measured, errors that may be expected, and limitations of the device in question.

Manuscript received March 31, 1967; revised April 10, 1967.
The authors are with the National Bureau of Standards, Boulder, Colo.

The customary model for an attenuator is a two-port[1] as shown in Fig. 1 (b) where access to the attenuator is achieved by ideal waveguide connectors. The fact that this condition and the presentation of proper impedances to the attenuator cannot be achieved in practice has led to a proliferation of definitions [2]–[6], [42] for Insertion Loss, Attenuation, Transducer Loss, Substitution Loss, Incremental Attenuation, Change in Insertion Loss, Intrinsic Attenuation, and others. Because so many different concepts have arisen, much confusion exists. A consistent set of definitions, necessary for a discussion of errors in attenuation measurements, is provided in this paper. Although they may not be universally acceptable, they do serve the purposes of this paper very well. In most coaxial cases the two-port has a connecting mechanism with losses and reflections as shown in Fig. 1(a). Access to the device is impossible in the manner described above and shown in Fig. 1(b). The model shown in Fig. 1(c) approaches the actual situation found in practice and is the model referred to in the definitions for Substitution Loss and Attenuation. These are the most useful concepts for attenuation measurements.

A. Insertion Loss

The most widely used definitions for insertion loss are contained in two IRE Standards [3], [4]. The earlier 1953 Standard [3] gives a quite general definition[2] using arbitrary source and load impedances. The 1959 IRE Standard [4] gives two definitions for insertion loss, one in which system mismatch is not specified and another in which the system is nonreflecting.[3] The definitions contradict each other since

[1] The term "two-port" as used here includes linear structures containing passive or active elements which may be nonreciprocal. Multiport devices such as directional couplers become "two-ports" when unused ports are terminated with unchanging impedances.

[2] *Insertion Loss:* 1) The loss in load power due to the insertion of apparatus at some point in a transmission system. It is measured as the difference between the power received at the load before insertion of the apparatus and the power received at the load after insertion. 2) The ratio, expressed in decibels, of the power received at the load before insertion of the apparatus, to the power received at the load after insertion.

[3] *Insertion Loss (of a Waveguide Component):* The change in load power, due to the insertion of a waveguide component at some point in a transmission system, where the specified input and output waveguides connected to the component are reflectionless looking in both directions from the component (match-terminated). This change in load power is expressed as a ratio, usually in decibels, of the power received at the load before insertion of the waveguide component to the power received at the load after insertion. Note 1: A more general definition of insertion loss does not specify match-terminated connecting waveguides, in which case the insertion loss would vary with the load and generator impedances. In this Standard, match-terminated connecting waveguides will be assumed unless otherwise specified. Note 2: When the input and output waveguides connected to the component are not alike or do not operate in the same mode, the change in load power is determined relative to an ideal reflectionless and lossless transition between the input and output waveguides.

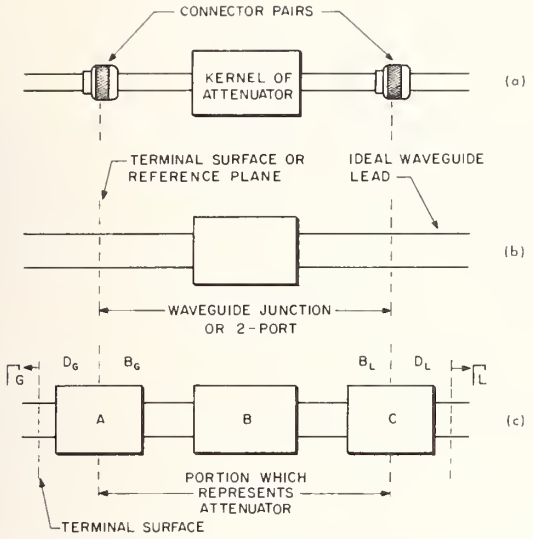


Fig. 1. Two models to represent an attenuator. (a) An attenuator. (b) A two-port representation. (c) Use of three cascaded two-ports to represent an attenuator connected in a system. Connector pairs are represented by two-ports A and C. The individual connectors are not separately represented by two-ports but are designated by D_G, B_G, B_L, and D_L.

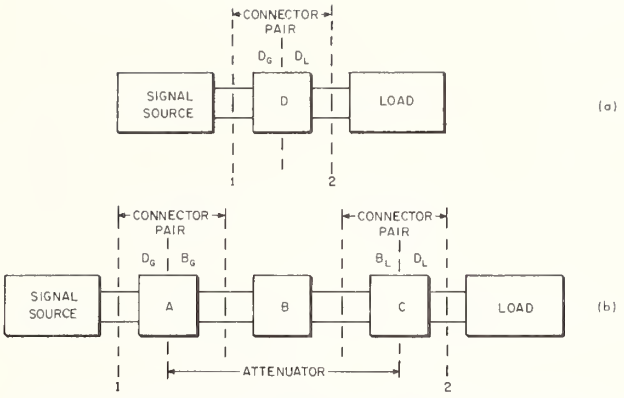


Fig. 2. Representation of a fixed attenuator inserted into a waveguide system. (a) Initial waveguide junction and connector pair. (b) Final waveguide junction where the attenuator consists of a central portion or kernel to which is attached connectors that will mate with the connectors belonging to the waveguide system to form the three cascaded two-ports shown.

the insertion loss of an attenuator will be different for each case. The measurement procedure for all the definitions is to open the system, insert the attenuator, and note the relative power absorbed by the load (or detecting device) before and after insertion. The insertion loss in decibels is then computed from these two values. If the attenuator is variable and remains in the system, the initial and final powers absorbed by the load for two settings are used; this determination is more properly called "Change in Insertion Loss."

Any of the definitions are entirely adequate for a single, unique system, but if the loss (or gain) measurement is to be transferred from one laboratory to another, more must be specified about system conditions in the "mismatched" case.

In practice, many manufacturers use "insertion loss (nonreflecting)" for what is described in Section II-C as attenuation. This includes the attenuation at the zero dial setting of a variable attenuator and the forward loss in a nonreciprocal device such as an isolator.

B. Substitution Loss [5]

In this case two conditions must be described, an initial system configuration as shown in Fig. 2(a) and a final condition as shown in Fig. 2(b). The quantity to be measured is still the relative power absorbed by the load before and after insertion, but in calculating substitution loss one includes the effect of system mismatches and connector losses. These losses are functions of differing interface combinations and not of the connector body, which should be lumped with attenuator loss. This definition is perfectly general and applies to variable as well as fixed attenuators. The main difference between substitution loss and insertion loss (general) is that in the former, the initial condition (joining connector) is specified, but in the latter nothing whatever is said about it. It should be noted that substitution loss depends on both the attenuator and the system. Because so many variables must be measured before the calculation of substitution loss can be made, simplifying assumptions are usually applied. A special case of substitution loss is known as attenuation.

C. Attenuation

This is defined as the insertion loss in a nonreflecting system ($\Gamma_g = \Gamma_l = O$),[4] and is the same as the IRE 1959 definition of insertion loss (matched) [4]. It is ideally a property of the attenuator only and does not depend on the system. In actual measurement these initial conditions can only be approached, due to imperfections in connectors or adapters and the uncertainties in reducing system reflections to zero. These imperfections and uncertainties are much more pronounced in coaxial systems than in waveguide except when precision coaxial connectors are employed. Since attenuation cannot be exactly measured a more practical term has been defined and may come into more general use where uncertainties less than 0.01 dB, 0.001 dB are important.

It should be noted that the differences between attenuation and standard attenuation become negligible when high-precision connectors are used. These connectors are not widely used at present because of their high cost.

D. Standard Attenuation

This has been defined [5] as the insertion loss of a linear two-port device in a nonreflecting system which is initially connected together at the insertion point by a standard[5]

[4] Where Γ_g and Γ_l are defined as the reflection coefficients of the generator and load, respectively. See Fig. 1(c) for reference plane.

[5] Standard connector or waveguide point. A "standard connector" is one which is made precisely to standard specifications. It can be a general connector, such as the type N or TNC, or a high-precision connector (see elsewhere in this issue [73]). Standard connector pairs usually have low but measurable loss and reflections [7], [8].

connector pair, adapter, or waveguide joint, the nonreflecting condition being obtained in the standard waveguide sections to which the standard connectors or waveguide joints are attached. The standard attenuation is the ratio expressed in decibels of the powers absorbed by the load before and after the insertion of the two-port device under test. Eventually, the concepts in this definition should probably be incorporated in the definition for attenuation (Section II-C) and thus avoid still another term.

E. Incremental Attenuation

Incremental attenuation is the change in attenuation of an adjustable attenuator between a reference setting (usually zero) and any other setting. The same restraints on system conditions apply as for attenuation and standard attenuation. The term "differential attenuation" is sometimes used instead of incremental attenuation and usually refers to two nonzero settings.

F. Intrinsic Attenuation [2]

This is more of a theoretical quantity than a practical one and relates to the absolute minimum attenuation which can be obtained from a mismatched two-port network. This is achieved by tuning the input and output with appropriate susceptances to achieve a bilateral match. It is almost always possible to achieve a combined nonreflecting network with "lossless" tuners. Since these "lossless" tuners will have losses of a few tenths of a decibel unless specially designed and constructed, they add to the apparent or measured intrinsic attenuation. Unless the initial mismatch is substantial (VSWR > 2), the apparent net loss will not be improved.

G. Transducer Loss [5]

This is a special case of substitution loss where it is assumed that the initial system contains a perfect transducer at the insertion point (lossless, nonreflecting connectors) which transmits all available generator power to the load. Removal of the perfect transducer and substitution of another will always decrease the transmitted power unless the second transducer is also perfect. It follows that transducer loss is a measure of how closely a given junction approaches the performance of a perfect transducer between a generator and load. This concept can be of value in describing the losses of adapters such as waveguide-to-coaxial transitions. Since the initial junction is never perfect, the loss cannot be exactly measured, but the concept may be useful for analysis and calculation.

H. Other Definitions

Other definitions have been proposed and are presently under consideration by various IEEE subcommittees. Until the deliberations of these committees have resulted in agreement, their definitions of attenuation and insertion loss, in particular, must be regarded as tentative. The tentative definitions now being considered by the IEEE subcommittees will not be quoted in detail because they are subject to further change. However, the definitions mentioned in this paper have, for the most part, appeared many times in the literature. The references, particularly two articles by Beatty [5], [8] provide an extensive treatment of attenuator models and definitions of attenuation.

III. Types of Attenuators

This section deals with specific types of attenuators encountered in the measurement field. An understanding of measurement methods requires that the reader be familiar with the devices to be measured and know which devices qualify as standards in a given measurement technique. These attenuators can all be analyzed with the aid of the attenuator models described elsewhere [1], [8].

A. Broadband Dissipative Attenuators

For the frequency region from dc to 18 GHz several types of resistive attenuators have been developed. The frequencies from dc to 1000 MHz are most adequately covered by T or π section attenuators consisting of series cylindrical deposited film elements and shunt annular deposited elements [9], [10]. Because of the symmetry that can be achieved in a coaxial structure, these can be designed to have a very flat frequency response and low VSWR. The films are usually vacuum-deposited and stabilized by high-power pulsing. Other techniques of applying a thin resistive layer by bonding or "painting" and baking have been utilized, but the necessity of extremely uniform and very thin layers eliminates this method for high-frequency work. The exact techniques for designing compensated structures have not been well detailed in the literature, because most of this information is proprietary with the manufacturers of these devices. A disadvantage of the usual construction techniques lies in the fact that the shunt and series elements must be made separately and then assembled. This leads to discontinuities at the joining planes and a corresponding frequency sensitivity independent of current distribution on the elements themselves. One-piece deposited T or π structures would be desirable but have not appeared commercially. An approach which eliminates separate elements involves a resistive "card" and is described later in this section.

To solve the problem of frequency sensitivity above 1 GHz, the "lossy line" attenuator was developed [11]–[14]. This energy-dissipating device uses a metallized resistive film for the inner conductor (it can be used on the outer conductor also). The films used are thin compared with the depth of penetration. This ensures a nearly uniform current distribution through the thickness of the film at microwave frequencies. Certain limitations must be placed upon this attenuator. The design equations show an increasing attenuation frequency dependence with decreasing frequency and increasing resistive film loss per unit length. Proper design procedure requires approximately a wavelength of lossy material at the lowest frequency of operation and the addition of lower-loss, compensating films at the ends of the lossy film section for higher-value attenuators. Typical limits for most commercially available attenuators are 2 GHz to 18 GHz in frequency and 0 to 20 dB in attenuation.

Another type of dissipative attenuator coming into general use has a "card" in a strip-line configuration [15]. It is essentially a flat, rectangular piece of insulating material,

typically a ceramic for temperature stability, with a deposited thin resistive coating on the surface. One longitudinal edge or both edges of the card are grounded (connected to a conducting side plane), the input and output signals are coupled with directly connected probes to the center of the short sides, and the entire assembly is housed between conducting planes that are insulated from the card surface. Recent improvements have been described [16] involving multilayer films and a modified housing which decrease frequency sensitivity and increase power-handling capability. With an appropriate transition from coax to the element and housing these devices work well from dc to 18 GHz with low frequency sensitivity.

Directional couplers have been used occasionally as attenuation standards in both uniconductor waveguide and coaxial systems. They are attractive because their properties are primarily determined by mechanical configuration and only to a second order by the resistance stability of their terminations. They can be built to operate over large bandwidths with exceptionally low VSWR but may not come into general use because of their additional cost and complexity.

B. Power Ratio Standards

These devices were developed for the dc and audio-frequency range but are useful at much higher frequencies where techniques are available to make use of their characteristics. One very useful device is the barretter which has a response closely approximating square law even into the microwave-frequency region. Such a device will accurately measure power or signal ratios by referring to a substituted dc or audio voltage or ratio. Attenuation is calculated from these power ratios [37], [40], [46].

1) *Dc Power Ratio Standards:* Resistive voltage dividers [17] are available commercially but are not as widely used as the dc potentiometer. The potentiometer is a versatile instrument that is readily available in most laboratories and has been perfected to a high degree.

Dc power is normally determined by measuring the current through a resistor of known value. The uncertainty of this determination may be kept small. To determine the current, the voltage across the resistor is obtained. With precision potentiometers and accurately known resistors, one can readily make dc power ratio measurements with an uncertainty of 0.005 dB or better over a range of 20 dB. Precision equipment with lower uncertainties [17] is available to allow determination of dc power ratios with an uncertainty of 0.0003 dB over the same range.

2) *Audio-Frequency Power Ratio Standards:* Power ratio measurements in the audio range (typically 30 Hz to 20 kHz) are performed in the same manner as at dc using inductive voltage dividers or voltage transformers. Typical units have a ratio uncertainty of 0.001 percent plus 0.0006 percent of the ratio. For ratios near unity, this formula gives an uncertainty of 0.00014 dB rising to 0.052 dB for a ratio of 0.001 or 60 dB.

The inductive voltage divider is usually a very stable device and can be calibrated [18] to achieve very high accuracy. Calibration of inductive voltage dividers may be accomplished with capacitance standards [19], or "complementary" measurement techniques [20] more recently developed.

C. Waveguide Below-Cutoff Attenuators [21 a–n]

Since this is the type of standard employed in the IF substitution method of measuring attenuation, which is *very* widely used, it is given emphasis here. It will probably continue to be used for some time, although additional refinements will no doubt be made.

The waveguide below-cutoff (WBCO) attenuator is a nonresistive continuously variable attenuation standard. It is often called a piston attenuator. The incremental attenuation of this device may be closely predicted from only a knowledge of its dimensions. If the waveguide section is uniform, nearly lossless, and excited in only one mode by a sinusoidal signal, the field will decay exponentially along the guide. An actual device has a moving probe or coil to sample the decaying field and an indicating mechanism to relate probe position to output level variation. If certain constraints are applied regarding frequency of operation, resistivity of the waveguide, and ratio of the frequency of operation to the frequency of cutoff, the attenuator will constitute a nearly ideal standard. Its only obvious disadvantages are a relatively high minimum insertion loss and an attenuation rate that is a slowly varying function of frequency.

The WBCO attenuator that is most commonly encountered is constructed of circular guide for ease of precise fabrication and because excitation symmetry is not critical. It operates in the TE_{11} mode. It is true that the dominant TE_{11} mode in this guide is not far separated in frequency from higher, undesired modes, but these can be suppressed by metallic strips or dielectric mode filters. In this type of attenuator the TM_{01} mode is most likely to cause difficulty since its attenuation rate is close to that of the TE_{11} mode. Mode filters have been constructed that will suppress the TM modes by over 60 dB with a corresponding reduction in the TE_{11} mode of less than 0.5 dB [21-e].

A good approximation for the propagation constant of the TE_{11} mode is given by (γ is in nepers per meter) [21-k]:

$$\gamma = \frac{S_{11}}{r}\sqrt{1-\left(\frac{2\pi r f\sqrt{\varepsilon}}{S_{11}c}\right)^2}-\frac{1}{r\sqrt{\pi\mu f\sigma}}\left[1+j\left(1-\frac{1}{1-\frac{1}{r\left\{2-\left(\frac{2\pi r f\sqrt{\varepsilon}}{S_{11}c}\right)^2\right\}\sqrt{\pi\mu f\sigma}}}\right)\right] \quad (1)$$

where

$S_{11} \approx 1.8411838$ = first root of the first derivative of the first-order Bessel function of the first kind
r = radius of waveguide cylinder
f = frequency in hertz
ε = dielectric constant of the medium inside the waveguide (relative, usually ≈ 1.0003 for air)

385-582 O - 70 - 15

$c \approx 2.997925 \times 10^8$ m/s = velocity of light
μ = permeability of guide wall (relative, usually taken to be 1.00000)
σ = conductivity of guide wall.

An examination of the equation shows that the most important considerations other than the radius are frequency and conductivity. Brown [21-g] shows that the quantity in square brackets contributes negligible phase shift at IF frequencies in waveguides constructed of good conducting materials. The actual phase shift is less than one degree in 100 dB of attenuation.

At lower frequencies skin effect increases the effective electrical radius of the guide and contributes large uncertainties, since machining affects wall conductivity. For this reason and the possibility of leakage through the attenuator walls, WBCO attenuators are seldom used below 1 MHz. The upper frequency limit is determined by the cutoff frequency and the necessity to operate well below cutoff to decrease the frequency dependence of the attenuator. At extremely high frequencies (above 1 GHz) guide dimensions become so small that the necessary mechanical tolerances for an accurate standard cannot be readily achieved.

Equation (1) requires dominant-mode operation which can be achieved by proper excitation and adequate mode filtering. The coupling coils used in the attenuator must be sufficiently separated at all times to prevent mutual coupling and loading effects from excessively affecting the excited mode [21-l]. A separation equivalent to 30 dB (insertion loss) will make this effect negligible. A method of reducing this minimum insertion loss, by feedback circuitry to maintain constant current in the exciting coil and a new coil design to achieve more efficient excitation of the mode, has been partially successful [21-k] but at the cost of added complexity.

The remaining important considerations for the circular piston attenuator relate to dimensional tolerances on the guide and the resolution and accuracy with which the displacement of the exciting and receiving coils can be measured. The guide must not only be perfectly circular but must remain constant in diameter. If any ellipticity exists in the guide the propagation constant will be different from the circular case, degenerate modes may be excited, and further, if the major axis has angular variations, this will produce deviations from the exponential law. Typical commercially available WBCO attenuators for use at IF frequencies have inside guide diameters ranging from 0.75 to 1.50 inches (1.905 to 3.810 cm) for corresponding attenuation rates of 40 to 20 decibels-per-inch (1.813 to 0.907 neper-per-centimeter) displacement of the piston. To keep the uncertainty to 0.005 decibel per 10-decibel increment requires tolerances of a few ten-thousandths of an inch. Tolerances of the same order are required in measuring displacement. The traditional lead screw and mechanical counters are now being replaced by precision scales and optical readout devices. Greater precision has been achieved using the techniques of interferometry and direct digital readout.

D. Rotary-Vane Attenuators [22 a–h]

In the field of microwave attenuation the rotary-vane attenuator has enjoyed wide use as a direct reading attenuator and will probably replace the piston attenuator for very high-frequency measurements. It has a calculable attenuation law, which is a requirement for a standard attenuator. The ideal attenuation is given [22-a] by the expression

$$A = 40 \log_{10} \sec \theta + C \text{ dB}$$

where θ is the angle of the rotating vane with respect to the polarization of the TE_{11} mode in the waveguide, and C is the residual attenuation when $\theta = 0$. Tables of the variable component of attenuation as a function of the vane angle θ are available [22-f].

Errors in fabricating the attenuator cause deviations in actual attenuation from that indicated on the dial. The deviation is small at low settings but can be appreciably larger at high values of attenuation. Among the factors causing these deviations are mismatch, misalignment of the rotating vane, insufficient attenuation of the rotating vane, imperfections in the drive gear, parallax in the readout device, and warping of the vane.

At present further investigation is needed concerning alignment problems and perturbation of the field by the rotating vane. Calibration of a commercial rotary-vane attenuator by the dc substitution technique showed attenuation difference measurement in agreement to within 0.004 dB for the dial settings of 3 dB and less. This is a testimonial both to the theory upon which the attenuator is based and to the design and fabrication of the commercially available attenuators. At present the rotary-vane attenuator is available in the waveguide sizes operating from 2.6 to 220 GHz. The most convenient method for calibrating the rotary-vane attenuator is with reference to IF standards, but greater resolution and accuracy are obtainable with reference to dc and audio standards.

IV. Interlaboratory Calibration Standards

This section is general in nature with the desirable features of an attenuator being listed but no attempt made to categorize existing attenuation standards. The subcommittee on attenuation standards of the IEEE Group on Instrumentation and Measurement is presently working on recommendations for an attenuator and attenuation measurement standard.

Precision calibrated attenuators used as interlaboratory standards require careful handling and should meet the following criteria.

1) Over long periods of time, i.e., several years, the electrical characteristics should be stable.
2) When carefully handled and shipped they should be mechanically stable and rugged so that no permanent changes of characteristics occur.
3) For changes in environmental conditions such as temperature, pressure, humidity, etc., the changes in electrical characteristics should be small.

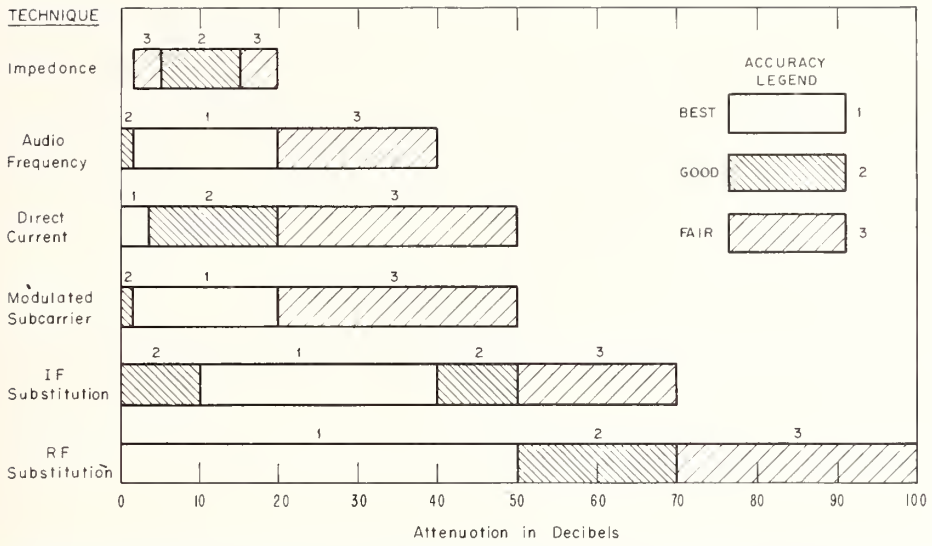


Fig. 3. Comparative ranges of accurate measurement for most frequently used attenuation calibration systems without special precautions. The "best" region represents systematic errors of 0.05 dB or less, the "good" region represents errors of less than 0.1 dB, and the "fair" region represents errors of less than 0.5 dB.

4) For power level changes, the changes in electrical characteristics should be small.
5) Reflections should be low (VSWR less than 1.2).
6) Input and output waveguide axes should be in line with each other.

A. Coaxial Standards

When the average power of 10 watts or higher is encountered in the measurement of attenuation, sections of lossy cable (RG-21/U) have been used as attenuation standards [23]. This type of standard is not usually suitable for precise work.

In the construction of attenuators in coaxial line, the T and π configurations are most often used. With precise construction these attenuators have stable characteristics and low reflections over frequency ranges from dc to 12 GHz. Recently broadband [15], [16] attenuators have been built that are useful at frequencies of dc to at least 18 GHz.

Most continuously variable attenuators with the exception of WBCO attenuators are not suitable at their present stage of development for use as interlaboratory standards. A number of different types of coaxial attenuators have been constructed using strip-line and rotating disks of resistive material [24], [25], hybrid or ring circuits in strip-line with sliding contacts, and sliding contacts on resistive cards [26]. Another type has been constructed by using a capacitively coupled metal slider, short-circuiting portions of a resistive center conductor [27]. Other partially successful attenuators have been variable directional-coupler types [28], [29]. In addition to construction in rectangular waveguide, the flap-type has been built in coaxial line. More development is needed in this area.

B. Rectangular Waveguide Standards

In rectangular waveguide, both fixed and variable attenuators are commercially available which are suitable for calibration. Usually one port is offset from the other port by a distance equal to one of the transverse dimensions of the waveguide. This displacement of one port of the attenuator makes it more difficult to perform an accurate calibration of the attenuator because the attenuation calibration system must accommodate not only the axial distance represented by the spacing between the two attenuator ports but also a small transverse displacement from the axis For fixed attenuators the in-line directional coupler [30] and related types possess the best characteristics for a calibrated standard.

There are two designs of variable attenuators that are suitable for calibration. One type has a resistive vane which moves into the waveguide field, from a side wall [25], [31]. The other type is the rotary-vane attenuator. Each type of attenuator has a region of best resolution, the rotary-vane attenuator at low values of attenuation and some models of the parallel translating-vane type at high values of attenuation. The rotary-vane attenuator is less frequency-sensitive and has less incremental phase change than the other type. In order to realize the presently available accuracy [32] of attenuation calibration corresponding to the dial resolution of commercially obtainable rotary-vane attenuators, further development in internal design is necessary.

V. Measurement Methods

This section includes well known measurement methods as well as some specialized or novel approaches. The most useful and important methods—dc substitution, audio substitution, IF substitution, RF substitution, subcarrier, and impedance—are covered in some detail and compared graphically in Fig. 3. The suggested regions for best accuracy are approximate, since they depend upon frequency, type of waveguide, type of attenuator, etc. and are based on the authors' experience. All of the systems may be used over greater ranges with degraded accuracy.

A. Dc Substitution

A converter is used in dc substitution methods to produce a change in dc power as a function of the change in RF power to be measured. It is necessary to know precisely how much the converter deviates from its ideal "law" in order to obtain accurate results. A dc substitution technique has been described [32] using bolometers (barretters or thermistors) in a balanced bridge arrangement. In this method, a change in a barretter's dc resistance from its nominal value, corresponding to a certain dc bias power applied, is caused by microwave power dissipated in the barretter. Dc power is decreased until the barretter resistance reverts to its initial value. The microwave power dissipated in the barretter is nearly equal to the amount of dc power removed, i.e., the ratio of the two powers is almost unity. Since only relative power is necessary in an attenuation measurement, the value of this ratio is not required, but it is necessary to know that it is constant during a measurement.

The constancy of this ratio has been investigated [33] by an experiment in which microwave power was fed through a power divider (3-dB directional coupler) to a barretter mount and a thermistor mount; the microwave power level was varied in steps over a 20-dB range and the ratio of the dc substituted powers in the two bolometers was noted for each step. Within the experimental error (0.1 percent) no change in the ratio was observable.[6] Over a 20-dB range an error of 0.1 percent corresponds to 0.0043 dB. Thus, one can presumably attach this uncertainty to the results of a one-step 20-dB attenuation measurement. For an attenuation increment smaller than 20 dB the uncertainty should be considerably less. A typical assigned measurement error of 0.1 percent plus 0.1 μW in power difference will create an error of 0.0001 to 0.0043 dB in attenuation as the attenuation increment measured varies from about 0.06 to 3 dB.

For measurements up to 3 dB one should measure the difference between the changes in dc power, $W_2 - W_1$, rather than the differences from initial dc bias power, $W_0 - W_1$ and $W_0 - W_2$. The attenuation A is given by

$$A = 10 \log_{10} \frac{W_0 - W_1}{W_0 - W_2} = 10 \log_{10} \frac{1}{1 - \dfrac{W_2 - W_1}{W_0 - W_1}}.$$

Analysis of equivalent errors in the two measurement approaches shows an overall error that is almost an order of magnitude less using $W_2 - W_1$ instead of $W_0 - W_2$ and $W_0 - W_1$ [32].

When measuring attenuation increments larger than 3 dB, the dc power differences $W_0 - W_2$ and $W_0 - W_1$ should be measured. In order to take full advantage of the inherent accuracy of the dc standard, other sources of error must be small. Particular attention must be paid to the reduction of system reflections and to careful evaluation of the remaining mismatch errors. This technique is ideal for calibrating rotary-vane attenuators because of their low reflections and their high resolution over the lower part of their range. A recent paper [34] describes an accurate dc substitution system, composed almost entirely of commercially available components, which is capable of 0.0001-dB precision for low attenuation values.

B. Audio Substitution

The drum-type attenuator has been available for many years as an accurate standard of attenuation at audio frequencies. This type of standard is relatively inexpensive and available in most laboratories. Accurate square-law converters such as barretters are also readily available and are inexpensive. Although the range of these converters is limited, the audio substitution technique of attenuation is extensively employed [35]–[38].

This technique of measurement involves the following steps: modulation of the signal source or its output,[7] square-law detection, attenuation of the audio frequency by a calibrated audio attenuator, amplification, and lastly detection and indication. The accuracy of this method depends on how much the converter or detector deviates from a square-law response.

A study [40] of barretters and crystal detectors describes how measurement range is related to the permissible deviations from a square-law response. For a deviation of 0.2 dB, the crystal video diode and the barretter have a range of about 38 and 53 dB, respectively. If the deviation is to be less than 0.01 dB a barretter has a range of about 20 to 30 dB: however, the use of a crystal is not recommended.

Other types of square-law detectors have been used which include thermistors [41], thermocouples, bolomisters [42], ferrite [43] and ferroelectric [44] detectors, and thermoelectric films [45]. Square-law deviations vary among these detector types, so they should be used only after their characteristics have been checked.

The effects of noise on errors in attenuation measurements have been investigated for several different detectors [46]. The increase in dc output caused by system noise and the rms fluctuation of linear detector output as a function of signal-to-noise ratio at the detector input are illustrated graphically in a handbook [47].

The range of attenuation measurements can be increased by precalibrating 20- or 30-dB fixed pads as "gage" blocks. For example, in the measurement of a 40-dB attenuator, a "gage" block of 20 dB used ahead of the detector reduces the range of level change at the converter to 20 dB. This procedure prevents the square-law range of the converter from being exceeded. Additional mismatch errors [48] arise with use of the "gage" block; thus, pads with low reflections are most desirable. A block diagram of this procedure is shown in Fig. 4.

[6] This determination error was for single-element bolometer mounts; if dual-element mounts are used, the error will be substantially more. The reader is cautioned to thoroughly review an article by Engen [41].

[7] Care should be employed with square-wave modulation because the bolometer time constants will cause distortion of the output ac waveform [39].

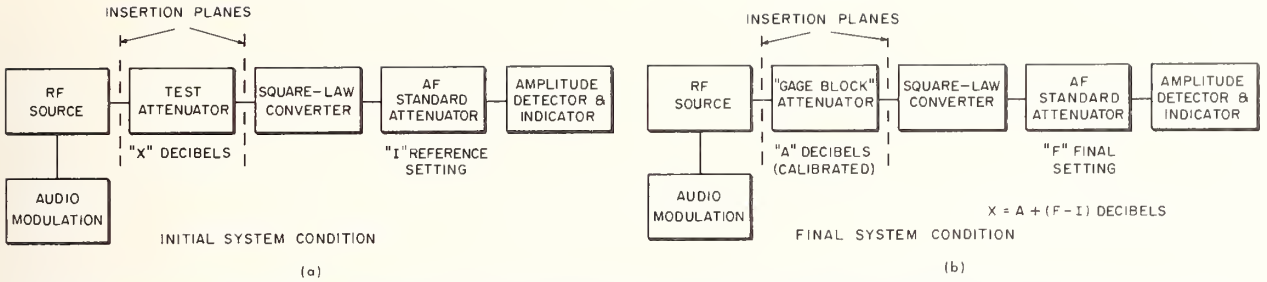


Fig. 4. Arrangement for extending range of attenuation measurement by using a "gage block" attenuator. (a) Initial system condition, standard attenuator set to minimum position, indicator at reference level. (b) Final system condition, "gage block" attenuator previously calibrated, standard attenuator set to higher value representing difference between unknown and "gage block" attenuators, indicator returned to reference level.

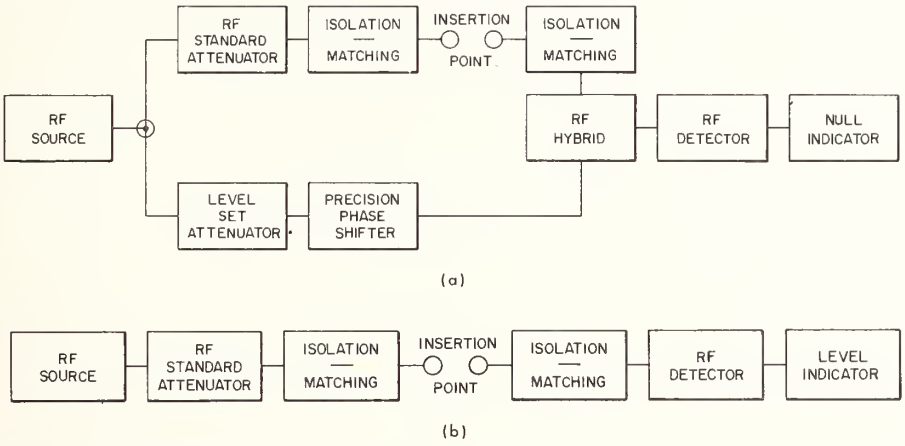


Fig. 5. RF attenuation measurement systems. (a) Dual-channel null system where null is obtained by matching channels in amplitude and adjusting for 180° phase difference. (b) Single-channel system, where a convenient level is selected on the indicator as a reference.

C. Direct Substitution

If the standard operates at the same frequency as the device under test when measuring attenuation or insertion loss the method is known as direct substitution. However, the IF, audio, and dc substitution techniques have a definite advantage because the standard operates at a fixed frequency without regard to the frequency chosen for the attenuator calibration. With the present state-of-the-art direct substitution methods offer the greatest measurement range. Two direct substitution arrangements are shown in Fig. 5(a) and (b).

The parallel substitution arrangement of Fig. 5(a) provides high resolution and range, and keeps the effects of amplitude instability in the signal source at a minimum. A phase shifter is required which should have a constant loss during adjustment. The parallel substitution arrangement is most desirable when measuring waveguide below-cutoff attenuators which have a high residual attenuation and a small change of phase.

The series substitution arrangement of Fig. 5(b) is frequently used when a signal generator-attenuator combination is to be calibrated. Although source and detector drift cause errors in the measurement, the method is convenient and requires no phase shifter. It is well suited for use with the rotary-vane attenuator which has small residual attenuation and introduces very little loss between the signal source and the detector.

Amplitude modulation of the source is frequently employed in both systems to permit the use of coherent detection to extend measurement range.

Other standards can be used in direct substitution systems in addition to the rotary-vane and waveguide below-cutoff attenuator. An example is the use of a slotted line. Methods using a slotted line are limited in range and errors have not been thoroughly investigated, but they may be of interest to those having limited or specialized facilities. The methods involve using a short-circuited attenuator after the slotted line [59-a], [74], [75], or placing the attenuator ahead of the slotted line and short-circuiting the slotted line itself [50]. The width-of-minimum method [7], [49] is generally used to obtain the necessary reflection coefficients to calculate attenuation in the short-circuited attenuator method. Attenuation in the second method is calculated from the measured probe displacement relative to the standing-wave minimum. Both methods are limited to a maximum attenuation increment of 15 dB or less. The best accuracy to be expected is about 0.05 dB and can be substantially worse if much mismatch is present.

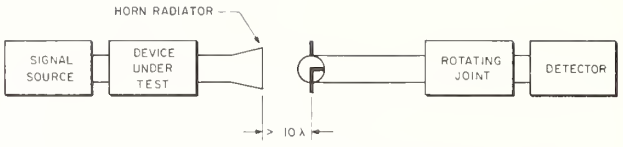


Fig. 6. Far-zone, principal-axis method of measuring attenuation with a radiator and rotating dipole pickup.

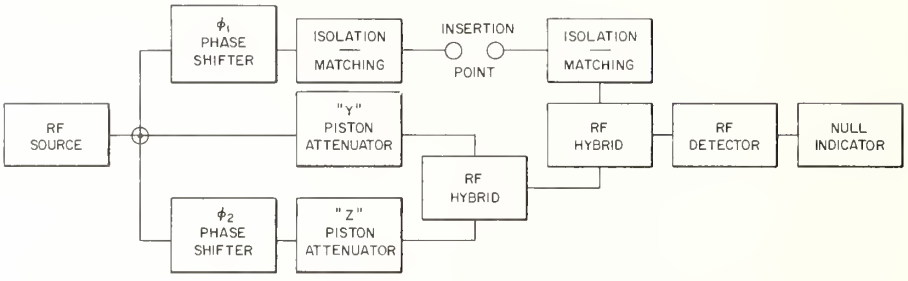


Fig. 7. Allred three-channel self-calibrating attenuation measurement system.

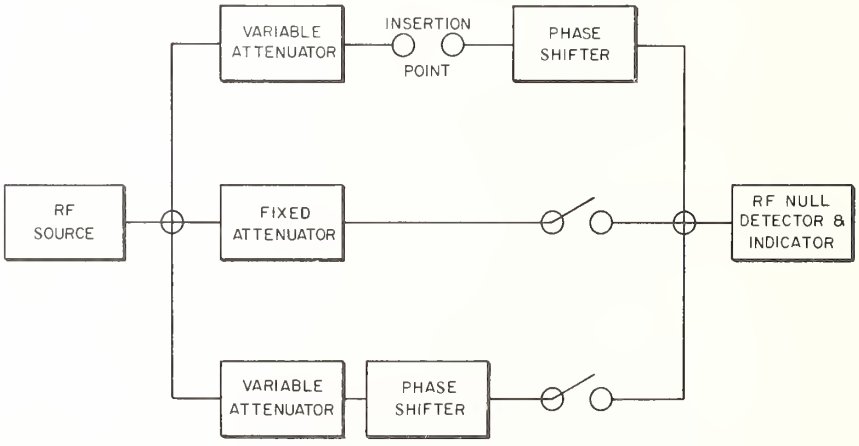


Fig. 8. Laverick self-calibrating attenuation measurement system.

A rotating dipole antenna in a linearly polarized field is another standard of attenuation for direct substitution [1]. This is shown in Fig. 6. Ideally, the probe response is proportional to the cosine function of the angle between the *E*-field and the dipole. With known angular position of the dipole relative to the *E*-field the values of the relative levels can be predicted. Many precautions must be observed when using this method. A study of antenna-gain measurement techniques is essential.

D. Methods Not Requiring Attenuation Standards

In the measurement of attenuation a standard is not actually required, as basically it is the determination of a dimensionless ratio. Methods devised for determining attenuation without a standard can be tedious; however, they are valuable when a standard is not available. Also, they provide a means of independently checking an attenuation standard. At about the same time two similar methods were developed by Allred [51-a] and Laverick [51-b]. These two methods require three-branch systems. The systems of Allred and Laverick are shown in Figs. 7 and 8. The first method obtains both phase and magnitude of the propagation constants of the two WBCO attenuators and, incidentally, of the unknown attenuation step added at the insertion point. The procedure is to balance the bridge for a null with the insertion point closed and to note the readings, y_a and z_a, of the two attenuators. Phase shifter $\phi 2$ is normally adjusted so that the output voltages of the two variable attenuators are approximately 90° apart. This provides easiest system adjustment but is not a necessary condition. The unknown is inserted and the attenuators adjusted for a new null without disturbing either phase shifter. The two new attenuator settings, y_b and z_b together with y_a and z_a are inserted in the following equation involving the insertion ratio of the unknown, K:

$$K = \frac{e^{-\gamma_y y_a} + We^{-\gamma_z z_a}}{e^{-\gamma_y y_b} + We^{-\gamma_z z_b}}$$

where γ_y and γ_z are the propagation constants of the two attenuators and W is a constant of the system expressing the relative phase angle and relative magnitude of the output voltages of the two attenuators. By changing phase shifter $\phi 1$ to a new setting and repeating the procedure, a set of simultaneous equations is obtained involving the same in-

sertion ratio, K. These equations may then be used to solve for the unknowns γ_y, γ_z, K, and W. The equation applies to WBCO attenuators operating in a single mode. With additional modes or with other types of continuously variable attenuators, a larger number of constants may be required to describe the system.

The system of Laverick uses the phase shifters to set the other two branches 180° out of phase with the "unknown" branch. The three branches are adjusted for equal output by nulling the "unknown" branch against each of the other two. The three branches are connected together and the attenuator to be calibrated in the "unknown" branch is set for a new null. This corresponds to a 6.02-dB change in attenuation. Repeating the operation by switching the fixed branch in and out yields attenuation increments of 3.52 dB, 2.50 dB, etc. (voltage ratios of 2, 3/2, 4/3). The chief advantage of this method is that no equations need be solved, while the major disadvantage is that the test attenuator must be continuously variable and the calibrated increments are 6 dB and less.

A number of sources of error are present with these systems of measurement even though those associated with the attenuation standard have been eliminated. For example, there are reflections at the insertion point of the test attenuator that cause errors, and changes in reflection that affect the power division between channels. Further, there are changes in the Laverick system caused by readjustment of phase shifters, and, in both systems, by frequency drifts that cause changes in level.

Limited resolution in these methods causes a small error each time that the signal in any channel is adjusted to produce a detector null. This error is influenced by detector sensitivity, attenuator repeatability, and noise level. These errors have a cumulative effect as each step is taken in the measurement process; however, with care these methods can achieve measurements up to 20 dB with uncertainties of less than 0.02 dB. Variable attenuators which do not produce large changes in phase shift are best suited for calibration by these techniques.

Other methods not requiring a standard attenuator have been described by Peck [51-c] and Davies [51-d]. In one of the methods described by Peck, the response of the detector must vary linearly with the magnitude of the signal amplitude. The constants in the linear relationship are eliminated by graphical means. The accuracy of this method is primarily limited by the linearity of the detector and by the precision of the graphical analysis.

The method of Davies utilizes two "identical" directional couplers and a two-channel circuit. The adjustment of phase shifters is not required. The accuracy is stated to be 1 percent of the attenuation in decibels, in contrast to the estimate of 0.1 percent by the method described by Laverick [51-b].

E. Impedance Technique

This technique determines the attenuation of a two-port over a range of 1 to 20 dB with fair accuracy by the measurement of input impedance at each port, with the other port terminated in a specified manner. Such a determination is derived from network theory. The methods and procedures used at microwave frequencies differ from those used at lower frequencies and bear description.

The method to be described [52] places no limitation on the attenuator except that it be a passive, reciprocal, linear network.

The two components of attenuation, reflective and dissipative attenuation, are obtained separately. Let the two attenuator ports be designated 1 and 2. The reflective attenuation is given by

$$A_R = 10 \log_{10} \frac{(\sigma_{1m} + 1)^2}{4\sigma_{1m}},$$

where σ_{1m} is the VSWR at port 1 with port 2 terminated in a matched load. This measurement is readily made with a good standing-wave machine. The matched load is obtained by tuning any available load at the frequency of interest.

The dissipative attenuation component is obtained by reversing the attenuator, and replacing the matched load with a lossless variable reactance normally consisting of a short-circuited section of transmission line of variable length. The reflection coefficient measurement made with this configuration is a function of the terminating reactance and has a circular locus when plotted in the reflection coefficient plane. The radius R_2 of this circle is equal to the efficiency η_m. The corresponding equation for the dissipative attenuation is

$$A_D = 10 \log_{10} \frac{1}{\eta_m} = 10 \log_{10} \frac{1}{R_2}.$$

A_R and A_D so determined are valid only when power flows from port 1 toward port 2.

The method is potentially as accurate as the normal power ratio method of measurement. In practice, difficulty is encountered in approaching a perfectly matched load and lossless variable reactance.

If the method is attempted with a tuned reflectometer instead of a standing-wave machine, some of the problems in obtaining the reflection coefficient circle are eliminated and the uncertainty of measurement correspondingly reduced. This method seems best suited to be used when measuring the attenuation of a two-port whose input and output ports are different configurations. Such a device might have a rectangular waveguide input and a coaxial output.

A more general technique [53] permits the determination of the attenuations of nonreciprocal or active devices. This method has limited application at present and the measurement is usually performed by other, more common, methods.

F. IF Substitution

This technique has become the most widely accepted and versatile method of performing attenuation measurements over a wide range with excellent accuracy. In simplest terms it involves frequency translation from a higher microwave frequency to a lower frequency, typically 30 or 60 MHz.

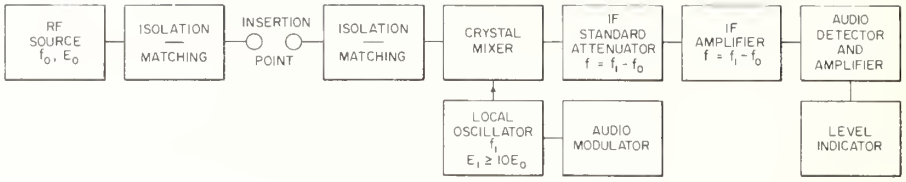


Fig. 9. Single-channel IF substitution system.

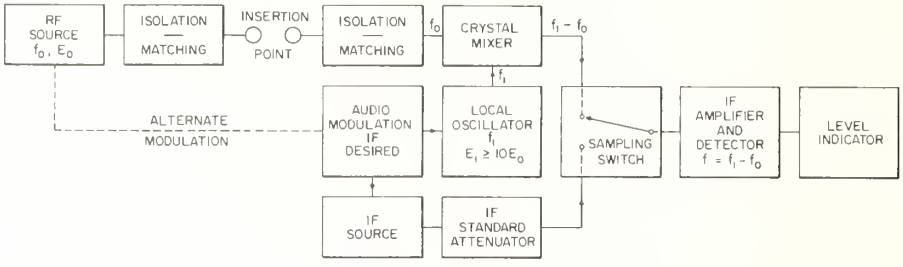


Fig. 10. Gainsborough two-channel IF system.

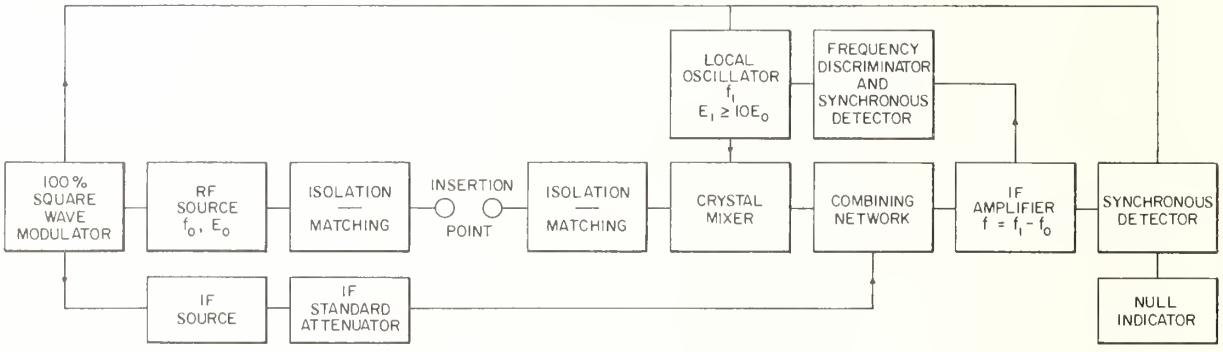


Fig. 11. Weinschel two-channel IF system.

At the lower frequency adequate standards of extreme precision are commercially available. The technique involves use of an extremely low modulation index to approach one-to-one correspondence between changes in power levels of the microwave and intermediate frequencies. The mixer is generally a point contact silicon diode, operated single-ended or balanced with the local oscillator power adjusted so that the crystal is operating in the linear power-conversion region. When the signal power is maintained 30 dB below the local oscillator level, the IF output level will track the signal level within 0.001 dB over a 55- to 60-dB range. Noise creates increasing system nonlinearity for greater ranges. If nonlinearities of the order of 0.01 dB can be tolerated, the range can be increased to 80 dB by operating at both higher and lower signal levels without calculating corrections. (See Fig. 15 and earlier articles [21-m], [54], [56].)

Several system configurations are possible with this measurement technique. Although the basic method is called IF substitution or IF plus partial RF substitution [48], the commonly used systems are of the series substitution or parallel comparison type.

The series substitution scheme involves maintaining a constant level at the detector with the accompanying problems of detector and source instability and drift. Audio modulation of the microwave signal is usually employed to narrow the effective bandwidth for low-level operation, but this technique must be used with care because of the possibility of introducing additional error.

The series substitution method is shown in Fig. 9. It is self-explanatory and represents a convenient, accurate measurement method. The major objection to this system, in spite of its simplicity, is the high minimum insertion loss of the standard WBCO attenuator. Modulation is shown applied to the local oscillator, but could be just as easily applied to the RF source if care is taken to prevent the introduction of frequency modulation. The use of more sophisticated equipment to phase-lock and level-stabilize the RF sources, noise injection, and coherent detection can extend measurement range and decrease random measurement error. Most of these schemes depend on the equipment available to the laboratory and the competence and ingenuity of the metrologist.

The parallel comparison technique is the other major IF method. Two approaches to system configurations are shown in Figs. 10 and 11. Close inspection of the two figures reveals that the two approaches are very similar. In the case of the Gainsborough system [21-c], continuous comparison

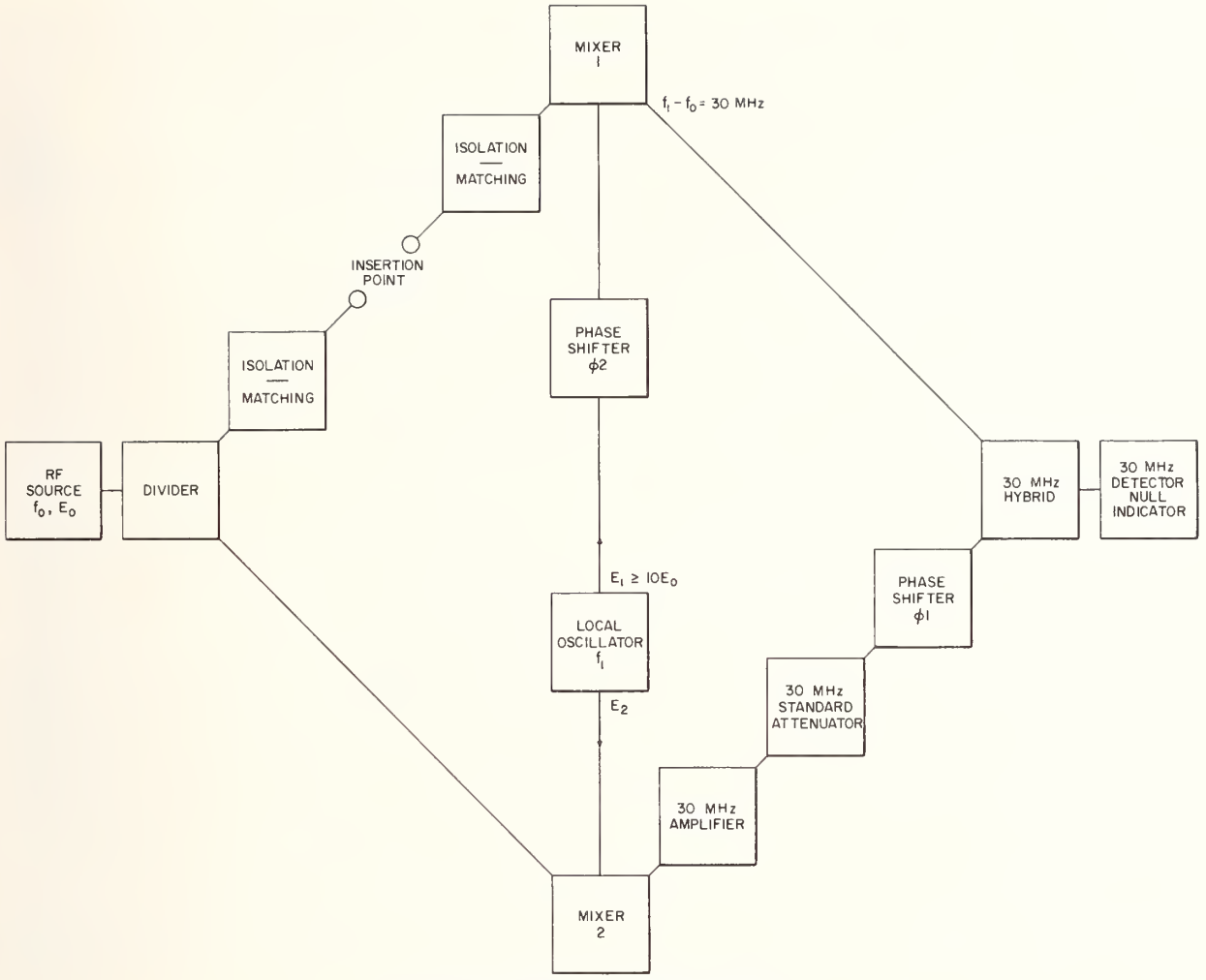


Fig. 12. Twin-channel unmodulated IF system.

of the relative levels of the two channels is accomplished by a sampling switch. The availability of semiconductor switches makes this a very attractive system to increase measurement range beyond the capability of the series configuration. The limitation of the high minimum insertion loss of the IF standard attenuator is circumvented and single-step measurements of 90 or 100 dB are possible. The linearity of the mixer is unchanged, however, and substantial error will be present for attenuation increments between 70 and 100 dB. For a given mixer, linearity curves may be plotted and corrections applied. The switch can introduce a subtle error that becomes very unsubtle at large attenuation increments. Cross-talk within the switch, which is never equal for the two switch positions, will create an offset error which may be difficult to recognize. To eliminate this problem and easily allow the use of synchronous detection to minimize random effects due to noise, the Weinschel system shown in Fig. 10 was developed [54], [55]. In this system, the two channels are square-wave-modulated in counterphase at an audio rate. This has the effect of alternately switching the two channels into one IF amplifier. The composite signal will consist of the IF frequency plus an audio modulation component. If the duty cycle and phase are properly adjusted and the IF levels are identical, the audio modulation will approach zero and a null at this frequency can be detected.

The IF source will have a lower noise level than the IF derived from the mixer. This will create an offset in detected dc level and a corresponding measurement error. This problem can be minimized by noise injection to provide balance for a no-signal condition [70]–[72].

In both systems, some form of AGC must be employed to maintain constant detection sensitivity since the input power to the IF amplifier is constantly changing. AFC should be employed to provide close agreement between the mixer IF output frequency and the frequency of the comparison IF source.

An approach which permits simultaneous phase and attenuation measurement has been reported but requires complex equipment in its realization [56]. The basic system is shown in Fig. 12. It has no modulation and indicates balance when a CW null is achieved at the IF frequency. The signal channels must be coherent and means must be provided for shifting the phase of the IF frequency. Since

phase is preserved in the mixing process, a phase shift in the measurement channel can be measured by placing a calibrated IF phase shifter in the reference channel. The attenuation is measured in the usual way with the standard in the reference channel. No range increase was reported with this system, but simultaneous attenuation and phase measurements are sometimes desirable.

The present state-of-the-art indicates that the greatest range and versatility can be obtained with IF techniques, although greater accuracies over limited ranges can be achieved with other methods. The homodyne and subcarrier systems promise equivalent performance but have not been widely implemented.

G. Modulated Subcarrier

An attenuation calibration system was developed at NBS, using a modulated subcarrier [57], [58]. This method of measuring attenuation difference utilizes commercially available components to obtain resolutions of 0.0001 dB and 0.02 dB at 0.01 dB and 40 dB, respectively, and accuracies comparable to those of the best previously available systems. In this system, a single unmodulated signal source is divided into two channels as shown in Fig. 13. The unmodulated channel contains a phase shifter, while the modulated channel includes the device under test. The modulating signal is also fed through a phase shifter to the standard audio attenuator. The RF signals are combined and fed to a phase-sensitive detector, where they are compared with a signal from the standard audio attenuator at the modulating frequency.

The modulated subcarrier technique has some advantages over the IF substitution technique which are as follows: 1) requires only one RF signal source, 2) uses easily obtainable and accurate commercial audio standards, 3) allows the determination of the phase shift of the device under test, and 4) has high resolution at low values of attenuation. Its disadvantages are as follows: 1) the measurement is not as easily made, 2) the range is less than a comparable parallel IF substitution system, 3) there are new sources of error, and 4) is somewhat more difficult to set up.

Systems such as this, employing two parallel microwave channels, are best suited for the measurement of incremental attenuation. If the system must be opened to insert an attenuator, substantial error may be introduced because the reference channel must be opened also. This limitation is most pronounced in systems using rigid transmission lines. The modulated subcarrier system is well suited for the calibration of rotary-vane attenuators because of small changes in phase of these attenuators and because of their high resolution at low values of attenuation. For the measurement of a fixed attenuator a waveguide section of equal length can be substituted to complete the waveguide circuit but the attenuation of this section also introduces an error.

There are other errors to consider in the measurement of attenuation with the modulated subcarrier technique. In making a measurement, the necessary phase shift adjustment can cause amplitude changes that are unknown. Other causes of errors are imperfect isolation between channels,

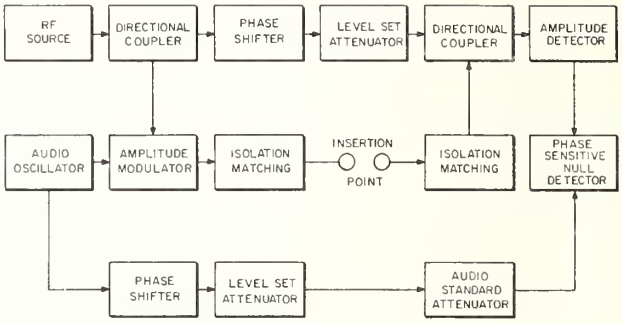


Fig. 13. Subcarrier attenuation measurement system.

modulation instability, ground loops, and pickup at the modulation frequency.

H. Measurement of Small Attenuation [59]

Attenuation measurement of the order of 0.1 dB or less is of interest in determining the attenuation of connectors and other low-loss components, typically short waveguide sections.

Several convenient techniques have been described in the literature, ranging from direct substitution to impedance standards. A very simple technique [59-c] treats a waveguide section as a transmission cavity with coupling accomplished by means of precalibrated irises. The measurement is repeated with different lengths of waveguide and differences are obtained. The loss is measured with a precision rotary-vane attenuator.

Audio substitution techniques are also effective but require more elaborate equipment [59-d, e].

Another method having high resolution and accuracy uses a two-channel configuration with a CW null being obtained with a phase shifter and an attenuator [59-f]. The frequency is selected to make the two-port a half wavelength. At this frequency, as no adjustment of the phase shifter is required to restore the null, the phase shifter is eliminated as a source of error. Using a rotary-vane attenuator as a standard, a precision of 0.001 dB is readily obtained.

It is also possible to effectively determine small losses by measuring the VSWR of a short-circuited section of cable or waveguide [58-a, b].

VI. Errors Related to Calibration Systems

Advancement in skill generally improves the precision of measurement, but the effort to provide a corresponding improvement in accuracy requires the re-examination of errors previously regarded as negligible. Confidence in a precise measurement of attenuation is justified only when all the errors are accounted for to the best of one's ability.

When an interlaboratory standard is calibrated the value assigned to the standard is reported to be correct within certain limits. Errors that determine these limits are both systematic and random. A systematic error often has a unidirectional bias while random errors present in the measurement tend to be positive or negative with equal probability

[22-c]. Some errors have both systematic and random components that are not easily separable.

In the measurement process some of the errors are random in nature and their effect is reduced by averaging the result of a number of measurements. An example of the random error is the operator's inability to reset the indicator dial exactly to the reference position after adjusting the attenuator under test. Some systematic error could also exist, depending on the habits of the metrologist, for example, the way he views the indicator dial.

An important source of error in any measurement method relates to alignment of the device under test with the rest of the system. The error due to this factor is most pronounced in systems which must be opened to accommodate the unknown attenuator and then closed. To reduce this error, "optical" alignment benches, rolling equipment on tracks [76] and special slides are all effective.

A. Mismatch Error

Both systematic and random error can be considered in the evaluation of the mismatch error at the insertion point of a calibration system. The ordinarily systematic error in a given "matched" system can be made random by suitably changing or adjusting a portion of the system between measurements. However, even without adjustment, a small random component can be introduced if there are changes in frequency between measurements.

The errors in the measurement of standard attenuation due to mismatch and to connector deficiencies have been analyzed [7], [8]. Regardless of the method of measurement the interaction of system reflections and reflections from the attenuator constitute a significant source of error. The mismatch errors for fixed pads and variable attenuators have been analyzed [48], [60]–[62] and graphs [63] are available for rapid estimate of error limits. Figure 14 is an example of such a graph.

It cannot be overemphasized that a major stumbling block in accurate attenuation measurement is the connecting mechanism. To expend great effort in reducing other system errors without a corresponding effort to improve connectors is a waste of time. Some commercially available connectors for coaxial systems contribute errors due to mismatch, leakage, and repeatability that can be equal in magnitude to all other errors combined. Coaxial connectors of high precision, suitable primarily for laboratory use, have recently become available. General precision versions of these connectors suitable for field use may eventually reduce typical mismatch error by an order of magnitude.

The same mismatch problems exist with waveguide flanges, although to a lesser degree. Precision standardized flanges will be necessary on interlaboratory standards before further accuracy improvement is possible.

To determine and correct for mismatch error requires information on both the magnitudes and phases of a number of reflection and scattering coefficients. Usually only their magnitude is readily determined. This allows calculation of the limits of error without regard to where the error may lie between these limits. Further, the accuracy to which

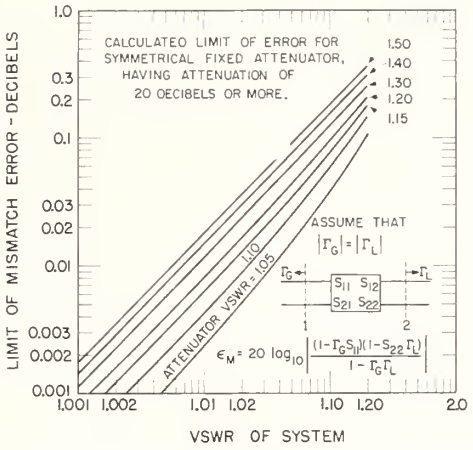


Fig. 14. Mismatch error curve for single, fixed, symmetrical attenuator.

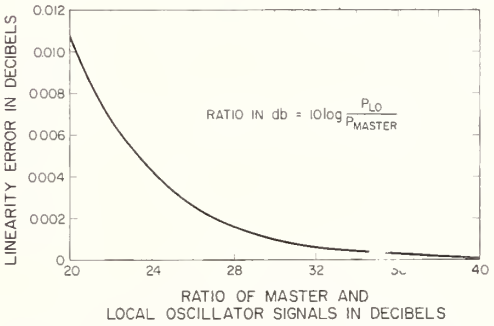


Fig. 15. Ideal crystal mixer curve for IF substitution system.

nonreflecting conditions can be recognized is limited. Thus, limit of mismatch error is really a cautious limit on the ability to determine how closely one has approached the desired nonreflecting condition. Attempts to eliminate mismatch error by special techniques have apparently not been successful [64].

B. Converter Errors

In the IF calibration system the power conversion linearity of the mixer or converter is of primary importance. If certain precautions are not observed significant linearity deviations will occur. In order to reduce these deviations from linear power conversion, the signal level incident on the mixer must be substantially lower than the incident local oscillator power. Experimentally it has been shown that the deviations remain less than 0.01 dB for a range of at least 60 dB, provided the signal power in the mixer is at least 21 dB below the local oscillator power in the mixer [21-m], [54]. Figure 15 shows the deviation from linearity computed for an ideal crystal mixer. This curve is based on an analysis of envelope detection in the ideal mixer. In practice other factors such as change in conversion efficiency, the actual curvature of the I-V characteristic, and effects as yet undefined cause deviations from this curve. Some analysis of this error has been done [54], [56], [65] but additional work will be required if all the sources of error in the IF method are to be reduced to the point where 0.001 dB is significant. The careful metrologist will run linearity curves

using a "bootstrap" technique to evaluate each mixer he uses.

Others errors normally associated with the converter can be caused by instability of the local oscillator signal level and frequency, pulling of the local oscillator by load changes, and instability of the crystal.

C. Instability of Signal Source

The magnitude of signal source instability in a calibration system is a major factor in limiting the ability of the operator to return to the previous reference level. For an incremental attenuation measurement the operator makes two settings. An output fluctuation of ± 0.005 dB could cause an uncertainty of ± 0.010 dB under the most unfavorable condition.

Frequency instability of the source will cause changes in the output level of a calibration system using frequency-sensitive components. The use of stable power supplies and water cooling [66], [67] of the oscillator permits output fluctuations of the system caused by signal source instability to be held to about ± 0.02 dB. The system stability can be improved by an order of magnitude if the oscillator is phase-locked to a harmonic of a stable source operating at a lower frequency, locked to a reference cavity, or using stable frequency injection.

D. RF Leakage

The RF energy received by the detector from paths other than through the attenuator under test can be termed leakage [68]. The magnitude of this type of error during a measurement depends on the imperfection of system components and the decibel value of the measurement. (It also depends upon the relative phases of the energies reaching the mixer by different paths.)

For values of measurement less than 40 dB the leakage effect usually is not a serious problem. When the leakage present is large, the movement of a metal object near the RF source may affect the output. Another possible way to detect leakage is to progressively increase the attenuation in the RF section of the circuit until no further decrease in the output level can be observed; then, the residual signal is due to either leakage or noise. It is possible to tell the difference by observing the character of the output signal, or by turning off the oscillator. Since the phase of leakage energy is so important, one can insert a half-wave section of transmission line in front of the test attenuator and observe the result. If leakage is present, the output will change radically. With the use of an auxiliary receiver having a cable and movable probe, the source of leakage can sometimes be located by exploring the vicinity of joints and connectors. In order to measure high values of attenuation, it is usually necessary to place the oscillators in shielded enclosures, to use gaskets at joints, to reface flanges, and to wrap leaking components with steel wool, metal mesh, or use magic.

Both systematic and random components are in the errors caused by leakage. By varying the phase of leakage in some manner from one measurement to another, the errors can be made more random. This can be accomplished by making each measurement at a different frequency or by using a line stretcher to change the phase of the RF that reaches the mixer by the desired path through the attenuator.

E. Noise

The uncertainties that are caused by noise in attenuation measurements are less than 0.005 dB when the signal levels in the system are at least 30 dB above the noise. However, the measurement of large values of attenuation requires a large amount of IF attenuation and shifts the reference level of the output signal toward the noise level. This noise may originate from such components as the IF amplifier, the mixer, the local oscillator, or the signal source.

In the single-channel IF substitution system with no IF preamplifier, the same signal level is presented to the IF amplifier with the test attenuator in or out of the system; thus, the noise generated in the IF amplifier has little effect on the measured value of attenuation other than to decrease system resolution and increase random error.

The noise originating in the mixer has a greater effect on the system output during a measurement when the IF attenuator is at its lower, or reference value, and a lesser effect when it is set to the greater value of attenuation. In measuring a large attenuation increment, the detector input signal from the IF amplifier is composed of different components, amplified signal and noise in one case, and chiefly amplified signal in the other. When noise is present, less signal is required for a given indication and the IF attenuation change is less than the RF change. Thus, the apparent measured value of RF attenuation will be too low [69].

Noise generated in the RF circuit will have a different effect depending on whether it originates ahead of or behind the RF attenuator being measured. If it originates ahead of the RF attenuator it will have very little effect since the signal-to-noise ratio will tend to remain constant as the attenuator is varied. If it originates after the RF attenuator, it will have a similar effect to noise originating in the mixer and will create a similar error.

The best way to offset noise error in measuring large attenuation increments is to inject noise into the IF channel [70]–[72]. If this is carefully done, the error due to noise offset can be minimized. The undesirable noise originating in the local oscillator can be decreased by using a balanced mixer. The random effects of noise can be reduced substantially by synchronous detection at the intermediate frequency, since the effective bandwidth is readily reduced and intermodulation noise content is removed. The nonlinearity effects due to noise can also be reduced by calibrating the system by progressively decreasing the signal level in known increments.

F. Errors Associated with the Device Under Test

Certain sources of error are related directly to the characteristics of the device under test. The most common in

attenuation measurement is the mismatch error which depends upon the reflections between the device and the insertion point. However, equations in an article by Beatty [60] show that the mismatch error will disappear regardless of the reflections from the device under test if the system itself has no reflections. Since we cannot achieve a reflectionless system, in practice, for a given system the uncertainty of attenuation measurement will have lower limits for the device having the lower reflections.

Poor construction, wear, and improper design of the connectors on an attenuator will cause deviations in the measurement of insertion loss or attenuation. Also, due to mechanical defects in a variable attenuator, one may not be capable of repeating the same value of incremental attenuation. With an inexperienced operator these defects in the device may be mistaken for system errors rather than malfunctions in the device under test.

G. Errors in the Standard Attenuator

The errors in the standard WBCO attenuator are basic and have been discussed previously in Section III-C. With proper design, construction, and operation, these attenuators approach measurement uncertainties of 0.001 dB at 20 dB [21 a–n].

The uncertainties associated with dc and audio standards are discussed in Section III-B. Detailed analyses of their errors are left to other sources [17]–[20].

Conclusion

It is essential to carefully describe the attenuator to be measured and the measurement conditions. Unless this is done it is impossible to transfer an attenuation measurement without imposing uncertainty limits that reduce the value of a calibration.

The present state-of-the-art in attenuation measurement offers many methods to measure attenuators [77]. The accuracy or range required will determine the most suitable method. It is also possible to measure attenuation without reference to a standard attenuator. Systems are available for the measurement of attenuation and phase shift, but where extreme accuracy is required these systems have not been widely implemented.

Presently, the dc substitution technique can claim the lowest uncertainty: ± 0.0001 dB at 0.01 dB. A range of 150 dB is obtainable at 30 MHz using a WBCO attenuator of high accuracy in a series-parallel substitution arrangement. The WBCO attenuator is presently the most important standard for wide-range measuring systems and promises to be so for some time to come, as it is simple, accurate, and reliable.

The rotary-vane attenuator has potential as a standard, and could even be used as an "IF" standard to minimize frequency translation ratios for millimeter attenuation measurement. Optical techniques are generally followed above this region.

References

[1] R. W. Beatty, "Microwave attenuation standards and measurements," NBS Mono. 97, April 3, 1967.

[2] ——, "Intrinsic attenuation," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-11, pp. 179–182, May 1963.

[3] "IRE Standards on Antennas and Waveguides: Definitions of Terms, 1953 (53 IRE 2. S1)," *Proc. IRE*, vol. 41, pp. 1721–1728, December 1953.

[4] "IRE Standards on Antennas and Waveguides: Waveguide and Waveguide Component Measurements, 1959 (59 IRE 2. S1)," *Proc. IRE*, vol. 47, pp. 568–582, April 1959.

[5] R. W. Beatty, "Insertion loss concepts," *Proc. IEEE*, vol. 52, pp. 663–671, June 1964.

[6] ——, "The system of electromagnetic quantities at frequencies above 1 GHz," *Metrologia*, vol. 2, pp. 46–54, January 1966.

[7] I. A. Harris, "Mode of use and assessment of precision coaxial connectors," *Proc. IEE (London)*, vol. 112, pp. 2025–2032, November 1965.

[8] R. W. Beatty, "Effects of connectors and adapters on accurate attenuation measurements at microwave frequencies," *IEEE Trans. Instrumentation*, vol. I-13, pp. 272–284, December 1964.

[9] J. S. Elliott, "Coaxial attenuation standards," Bell Telephone Labs., Inc., Record 27, pp. 221–225, June 1949.

[10] B. O. Weinschel, "High-frequency radial coaxial attenuator," U. S. Patent 2 994 049, July 25, 1961.

[11] H. J. Carlin and J. W. E. Griemsmann, "A bead supported coaxial attenuator for the frequency band 4000–10 000 Mc/Sec," *Proc. NEC*, vol. 3, pp. 79–89, November 1947.

[12] B. O. Weinschel, "Inside-out attenuator for high-frequency coaxial lines," U. S. Patent 3 002 166, September 26, 1961.

[13] B. O. Weinschel and J. F. Cottrell, "Frequency-compensated coaxial attenuator," U. S. Patent 3 005 967, October 24, 1961 (lossy-line type).

[14] H. A. Norman, "Frequency-compensated coaxial attenuator having part of resistive film reduced and bridged by capacitance," U. S. Patent 3 105 211, September 24, 1963 (lossy-line type).

[15] B. O. Weinschel, "Card attenuator for microwave frequencies," U. S. Patent 3 157 846, November 17, 1964.

[16] H. Bacher and E. R. Seitter, "Multi-layer card attenuator for microwave frequencies," U. S. Patent 3 260 971, July 12, 1966.

[17] A. F. Dunn, "Calibration of a Kelvin-Varley voltage divider," *IEEE Trans. Instrumentation and Measurement*, vol. IM-13, pp. 129–139, June–September 1964.

[18] W. J. Karplus, "Design and error analysis of high-accuracy dc voltage-measuring systems," *IEEE Trans. Instrumentation and Measurement*, vol. IM-13, pp. 139–145, June–September 1964.

[19] T. L. Zapf, "Voltage ratio measurements with a transformer capacitance bridge," *J. Res. NBS*, vol. 66C, pp. 25–32, January–March 1962.

[20] ——, "The accurate measurement of voltage ratios of inductive voltage dividers," *Acta IMEKO III*, pp. 317–331, 1964.

[21] Selected bibliography on waveguide below-cutoff attenuators:

a) D. E. Harnett and N. P. Case, "The design and testing of multirange receivers," *Proc. IRE*, vol. 23, pp. 578–593, June 1935.

b) R. N. Griesheimer, "Microwave attenuators. Cutoff attenuators," in *Technique of Microwave Measurements*, M.I.T. Radiation Lab. Ser., vol. 11. New York: McGraw-Hill, 1947, ch. 11, pp. 685–719.

c) G. F. Gainsborough, "A method of calibrating standard-signal generators and radio-frequency attenuators," *J. IEE (London)*, vol. 94, pt. III, pp. 203–210, May 1947.

d) J. J. Freeman, "Theory and design of a cavity attenuator," *J. Res. NBS*, vol. 40, pp. 235–243, March 1948.

e) R. E. Grantham and J. J. Freeman, "A standard of attenuation for microwave measurements," *Trans. AIEE*, vol. 67, pt. I, pp. 329–335, June 1948.

f) H. A. Wheeler, "The piston attenuator in a waveguide below cutoff," Wheeler Labs., Great Neck, N. Y., Mono. 8, January 1949.

g) J. Brown, "Corrections to the attenuation constants of piston attenuators," *Proc. IEE (London)*, vol. 96, pt. III, pp. 491–495, November 1949.

h) H. M. Barlow and A. L. Cullen, *Microwave Measurements*. London: Constable, 1950, pp. 242–254, 384–388.

i) C. M. Allred, "Chart for the TE_{11} mode piston attenuator,"

J. Res. NBS, vol. 48, pp. 109–110, February 1952.
j) D. M. Kerns and R. W. Hedberg, "Propagation constant in rectangular waveguide of finite conductivity," *J. Appl. Phys.*, vol. 25, pp. 1550–1551, December 1954.
k) C. M. Allred and C. C. Cook, "A precision RF attenuation calibration system," *IRE Trans. Instrumentation*, vol. I-9, pp. 268–274, September 1960.
l) L. A. Birger, "Nonlinearity of cutoff attenuators due to the effect of loading," *Ismeritel'naya Tekhnika*, no. 11, pp. 59–61, 1960.
m) D. L. Hollway and F. P. Kelly, "A standard attenuator and the precise measurement of attenuation," *IEEE Trans. Instrumentation and Measurement*, vol. IM-13, pp. 33–44, March 1964.
n) B. O. Weinschel and G. U. Sorger, "Waveguide below-cutoff attenuation standard," *Acta IMEKO III*, pp. 407–424, 1964.

[22] Selected bibliography on the rotary-vane attenuator:
a) G. C. Southworth, *Principles and Applications of Waveguide Transmission*. New York: Van Nostrand, 1950, pp. 374–376.
b) B. P. Hand, "A precision waveguide attenuator which obeys a mathematical law," *Hewlett-Packard J.*, vol. 6, pp. 1–2, January 1955.
c) W. Larson, "Analysis of rotation errors of a waveguide rotary vane attenuator," *IRE Internat'l Conv. Rec.*, pt. 3, pp. 213–219, 1962.
d) P. F. Mariner, "An absolute microwave attenuator," *Proc. IEE (London)*, vol. 109, pt. B, pp. 415–419, September 1962.
e) A. V. James, "A high-accuracy microwave-attenuation standard for use in primary calibration laboratories," *IRE Trans. Instrumentation*, vol. I-11, pp. 285–290, December 1962.
f) W. Larson, "Table of attenuation error as a function of vane-angle error for rotary-vane attenuators," NBS Tech. Note 177, May 1963.
g) ——, "Gearing errors as related to alignment techniques of the rotary-vane attenuator," *IEEE Trans. Instrumentation and Measurement*, vol. IM-14, pp. 117–123, September 1965.
h) ——, "Table of attenuation as a function of vane angle for rotary-vane attenuators ($A = 40 \log_{10} \cos \theta$)," NBS Tech. Note 229, January 7, 1965.

[23] R. W. Beatty and C. E. Huber, "Lossy cable attenuator pads," *Rev. Sci. Instr.*, vol. 24, pp. 1002–1003, October 1953.

[24] J. D. Cappucci, "Zero loss wideband coaxial line variable attenuators," *Microwave J.*, vol. 3, pp. 49–53, June 1960.

[25] C. W. Miller, M. C. Crowley-Milling, and G. Saxon, "Note on waveguide attenuators," *J. IEE (London)*, vol. 93, pt. IIIA, pp. 1477–1478, March–May 1946.

[26] B. O. Weinschel, "Precision variable coaxial attenuator," U. S. Patent 3 184 694, May 18, 1965.

[27] H. Bacher and J. E. Ebert, "Linear variable lossy line attenuator," U. S. Patent 3 237 133, February 22, 1966.

[28] J. K. Shimizu and E. M. T. Jones, "Coupled-transmission-line directional couplers," *IRE Trans. Microwave Theory and Techniques*, vol. MTT-6, pp. 403–410, October 1958.

[29] J. R. Mohr, "New coaxial variable attenuators," *Microwave J.*, vol. 8, pp. 99–102, March 1965.

[30] W. Larson, "Inline waveguide attenuator," *IEEE Trans. Microwave Theory and Techniques (Correspondence)*, vol. MTT-12, pp. 367–368, May 1964.

[31] *Technique of Microwave Measurements*, M.I.T. Radiation Lab. Ser., vol. 11, C. G. Montgomery, Ed. New York: McGraw-Hill, 1947, pp. 748–751.

[32] G. F. Engen and R. W. Beatty, "Microwave attenuation measurements with accuracies from 0.0001 to 0.06 decibel over a range of 0.01 to 50 decibels," *J. Res. NBS*, vol. 64C, pp. 139–145, April–June 1960.

[33] G. F. Engen, "Recent developments in the field of microwave power measurements at the National Bureau of Standards," *IRE Trans. Instrumentation*, vol. I-7, pp. 304–306, December 1958.

[34] C. T. Stelzried, M. S. Reid, and S. M. Petty, "A precision dc potentiometer microwave insertion loss test set," *IEEE Trans. Instrumentation and Measurement*, vol. IM-15, pp. 98–104, September 1966.

[35] J. Korewick, "Audio modulation substitution system for microwave attenuation measurements," *IRE Trans. Microwave Theory and Techniques*, vol. MTT-1, pp. 14–21, March 1953.

[36] ——, "A-M system measured microwave attenuation," *Electronics*, vol. 27, pp. 175–177, January 1954.

[37] B. O. Weinschel, "Insertion loss test sets using square law detectors," *IRE Trans. Instrumentation*, vol. PGI-4, pp. 160–164, October 1955.

[38] C. J. Finnie, D. Schuster, and T. Y. Otoshi, "A.C. ratio transformer technique for precision insertion loss measurements," Jet Propulsion Lab., California Institute of Technology, Pasadena, Rept. 32-690, November 1964.

[39] G. U. Sorger, "The thermal time constant of a bolometer," *IRE Trans. Instrumentation*, vol. PGI-4, pp. 165–170, October 1955.

[40] G. U. Sorger and B. O. Weinschel, "Comparison of deviations from square law for RF crystal diodes and barretters," *IRE Trans. Instrumentation*, vol. I-8, pp. 103–111, December 1959.

[41] G. F. Engen, "A dc-RF substitution error in dual-element bolometer mounts," *IEEE Trans. Instrumentation and Measurement*, vol. IM-13, pp. 58–64, June–September 1964.

[42] "Square-law microwave detector," *Military Sys. Design*, vol. 6, p.58, September–October 1962.

[43] G. B. Bogdanov, "Possible application of ferrites for absolute measurement of microwave power," *Radiotekhnika i Elektronika*, vol. 4, p. 663, 1961. English transl.: *Radio Engrg. and Electronic Phys.*, vol. 4, p. 587, 1961.

[44] F. L. Wentworth, J. D. Rodgers, J. W. Dozier, and M. Cohn, "Millimeter wave harmonic generators, mixers, and detector," presented at the Millimeter and Submillimeter Conf. (*Program and Digest*, pp. 103–105), 1963. (Copies available from E. H. Houseman, MP-75, Martin Co., Orlando, Fla.)

[45] S. Hopfer, N. H. Riederman, and L. A. Nadler, "The properties of thermo-electric elements as microwave power detectors," *IRE Internat'l Conv. Rec.*, pt. 3, pp. 77–84, 1962.

[46] G. U. Sorger and B. O. Weinschel, "Precise insertion loss measurements using imperfect square law detectors and accuracy limitations due to noise," *IRE Trans. Instrumentation*, vol. PGI-4, pp. 55–68, October 1955.

[47] *The Microwave Engineer's Handbook and Buyer's Guide*, pp. 318–319, 1966.

[48] R. W. Beatty, "Cascade-connected attenuators," *J. Res. NBS*, vol. 45, pp. 231–235, September 1950.

[49] *Technique of Microwave Measurements*, M.I.T. Radiation Lab. Ser., vol. 11, C. G. Montgomery, Ed. New York: McGraw-Hill, 1947, pp. 268–269.

[50] *Ibid.*, pp. 816–818.

[51] Selected bibliography on measurement of attenuation without using a standard:
a) "Self-calibrating method of measuring insertion ratio," *NBS Tech. News Bull.*, vol. 41, pp. 132–133, September 1957.
b) E. Laverick, "The calibration of microwave attenuators by an absolute method," *IRE Trans. Microwave Theory and Techniques*, vol. MTT-5, pp. 250–254, October 1957.
c) R. L. Peck, "A method for the self-calibration of attenuation-measuring systems," *J. Res. NBS*, vol. 66C, pp. 13–18, January–March 1962.
d) M. C. Davies, "An absolute method for calibrating microwave attenuators," *Proc. IEE (London)*, vol. 109, pt. B, suppl. 23, pp. 796–800, May 1962.

[52] R. W. Beatty, "Determination of attenuation from impedance measurements," *Proc. IRE*, vol. 38, pp. 895–897, August 1950.

[53] H. M. Altschuler, "The measurement of arbitrary linear microwave two ports," *Proc. IEE (London)*, vol. 109, pt. B, suppl. 23, pp. 704–712, May 1962.

[54] B. O. Weinschel, G. U. Sorger, and A. L. Hedrich, "Relative voltmeter for VHF/UHF signal generator attenuator calibration," *IRE Trans. Instrumentation*, vol. I-8, pp. 22–31, March 1959.

[55] A. L. Hedrich, B. O. Weinschel, G. U. Sorger, and S. J. Raff, "Calibration of signal generator output voltage in the range of 100 to 1000 megacycles," *IRE Trans. Instrumentation*, vol. I-7, pp. 275–279, December 1958.

[56] D. H. Russell, "An unmodulated twin-channel microwave attenuation measurement system," *ISA Trans.*, vol. 4, April 1965.

[57] G. E. Schafer and R. R. Bowman, "A modulated subcarrier technique of measuring microwave attenuation," *Proc. IEE (London)*, vol. 109, pt. B., suppl. 23, pp. 783–786, May 1962.

[58] W. E. Little, "Further analysis of the modulated subcarrier technique of attenuation measurement," *IEEE Trans. Instrumentation and Measurement*, vol. IM-13, pp. 71–76, June–September 1964.

[59] Selected bibliography on the measurement of small attenuations:
a) W. T. Blackband and D. R. Brown, "The two-point method of measuring characteristic impedance and attenuation of cables at 3000 Mc/s," *J. IEE (London)*, vol. 93, pt. IIIA, pp. 1383–1386, 1946.

b) J. H. Vogelman, "Precision measurement of waveguide attenuation," *Electronics*, vol. 26, pp. 196–199, December 1953.
c) P. D. Lacy and K. E. Miller, "An improved method for measuring losses in short waveguide lengths," *Hewlett-Packard J.*, vol. 9, November–December 1957.
d) B. O. Weinschel, "Insertion loss test sets using square law detectors," *IRE Trans. Instrumentation*, vol. PGI-4, pp. 160–164, October 1965.
e) C. T. Stelzried and S. M. Petty, "Microwave insertion loss test set," *IEEE Trans. Microwave Theory and Techniques* (*Correspondence*), vol. MTT-12, pp. 475–477, July 1964.
f) R. W. Beatty, "A two-channel nulling method for measuring attenuation constants of short sections of waveguide and the losses in waveguide joints," *Proc. IEEE* (*Correspondence*), vol. 53, pp. 642–643, June 1965.

[60] R. W. Beatty, "Mismatch errors in the measurement of ultrahigh-frequency and microwave variable attenuators," *J. Res. NBS*, vol. 52, pp. 7–9, January 1954.

[61] G. E. Schafer and A. Y. Rumfelt, "Mismatch errors in cascade-connected variable attenuators," *IRE Trans. Microwave Theory and Techniques*, vol. MTT-7, pp. 447–453, October 1959.

[62] K. Tomiyasu, "Intrinsic insertion loss of a mismatched microwave network," *IRE Trans. Microwave Theory and Techniques*, vol. MTT-3, pp. 40–44, January 1955.

[63] E. Weber, "The measurement of attenuation," in *Technique of Microwave Measurements*, M.I.T. Radiation Lab. Ser., vol. 11. New York: McGraw-Hill, 1947, sect. 13.5, ch. 13, pp. 824–832.

[64] A. N. Leber, "Method free from mismatching errors for measuring the loss of attenuators," *IEEE Trans. Microwave Theory and Techniques* (*Correspondence*), vol. MTT-12, pp. 480, July 1964.

[65] F. E. Terman, *Radio Engineering*, 2nd ed. New York: McGraw-Hill, 1937, pp. 448–449.

[66] E. Niesen, R. W. Beatty, and W. J. Anson, "Water-cooling of low-power klystrons used in the laboratory," *Rev. Sci. Instr.*, vol. 29, pp. 791–792, September 1958.

[67] W. Anson and E. Niesen, "Fast melting alloy forms water jacket for small klystrons," *Microwaves* (suppl. to *Electronic Design*, vol. 11, no. 6), p. 42, March 15, 1962.

[68] *Technique of Microwave Measurements*, M.I.T. Radiation Lab. Ser., vol. 11, C. G. Montgomery, Ed. New York: McGraw-Hill, 1947, pp. 830–831.

[69] *Ibid.*, pp. 845–846.

[70] A. L. Hedrich, B. O. Weinschel, G. U. Sorger, and S. J. Raff [55], p. 278.

[71] B. O. Weinschel and G. U. Sorger, "Standard source comparison system for high-frequency voltage ratio meter including alternately connectable load terminations," U. S. Patent 3 098 968, July 23, 1963.

[72] H. L. Kaylie, "A new technique for accurate RF attenuation measurements," *IEEE Trans. Instrumentation and Measurement*, vol. IM-15, pp. 325–332, December 1966.

[73] B. O. Weinschel, "Standardization of precision coaxial connectors," this issue.

[74] L. Essen, "The measurement of balanced and unbalanced impedances at frequencies near 500 Mc/s, and its application to the determination of the propagation constants of cables," *J. IEE* (*London*), vol. 91, pt. III, pp. 84–95, June 1944.

[75] W. Jackson, *High Frequency Transmission Lines*, Methuen's Monographs on Physical Subjects. New York: Wiley, 1951.

[76] A. Y. Rumfelt and R. J. Como, "Rapid insertion device for coaxial attenuators," *Rev. Sci. Instr.*, vol. 30, p. 687, 1959.

[77] The current state-of-the-art in attenuation standards can be inferred from manufacturers' literature and from articles in the literature such as the following:
a) R. E. Larson, "Microwave measurements in the NBS electronic calibration center," *Proc. IEE* (*London*), vol. 109, pt. B, suppl. 23, pp. 644–650, May 1962.
b) R. W. Beatty, "Standards and measurements of attenuation, impedance, and phase shift, September 1960–December 1962 (URSI U.S.A. National Committee Report, XIVth General Assembly, 1963)," *J. Res. NBS* (*Radio Science*)/*USNC-URSI*, vol. 68D, pp. 529–533, May 1964.
c) ——, "Microwave standards and measurements, A progress review 1960 to 1963," *IEEE Trans. Instrumentation and Measurement*, vol. IM-12, pp. 134–138, December 1963. Also see the book *Progress in Radio Science, 1960–1963, Vol. I* (*Radio Standards and Measurements*). Amsterdam: Elsevier, 1965. Articles in this book by D. Woods, S. Okamura and K. Sakurai, D. L. Hollway, and others discuss attenuation measurements.
d) Engineering Division, Radio Standards Lab., NBS, "Microwave and high-frequency calibration services of the National Bureau of Standards—Part II," *IEEE Trans. Microwave Theory and Techniques* (*Correspondence*), vol. MTT-12, pp. 572–576, September 1964.
e) "Accuracy in measurements and calibrations, 1965," W. A. Wildhack, R. C. Powell, and H. L. Mason, Eds., NBS Tech. Note 262, June 15, 1965.
f) J. E. Ebert and G. U. Sorger, "Survey of precision microwave attenuation measuring techniques," *ISA Trans.*, vol. 3, pp. 280–290, July 1964.
g) B. O. Weinschel, "An accurate attenuation measuring system with great dynamic range," *Microwave J.*, September 1961.

[78] The following textbooks contain valuable source material on microwave attenuation measurements and standards:
a) *Technique of Microwave Measurements*, M.I.T. Radiation Lab. Ser., vol 11, C. G. Montgomery, Ed. New York: McGraw-Hill, 1947. (See pt. IV, beginning on p. 679.)
b) D. D. King, *Measurements at Centimeter Wavelength*. Princeton, N. J.: Van Nostrand, 1952. (See sec. 3.13, beginning on p. 99.)
c) E. L. Ginzton, *Microwave Measurements*. New York: McGraw-Hill, 1957. (See ch. 11, beginning on p. 462.)
d) *Handbook of Microwave Measurements*, 3d ed., M. Sucher and J. Fox, Eds. Brooklyn, N. Y.: Polytechnic Press, 1963. (See ch. VII, beginning in vol. 1, p. 377.)

Reprinted from the PROCEEDINGS OF THE IEEE

Vol. 55, No. 6, June, 1967

pp. 942-959

The Institute of Electrical and Electronics Engineers, Inc.

UHF and Microwave Phase-Shift Measurements

DOYLE A. ELLERBRUCH

Abstract—**A phase-shift standard, a measurement system, and the techniques for determining the corresponding limit of uncertainty are all required for obtaining the phase-shift characteristics of UHF and microwave components.**

Differential phase-shift standards, measurement techniques, and measurement uncertainties are all discussed in a general sense and a comprehensive bibliography is included to supplement the general discussion.

I. Introduction

PHASE-SHIFT calibration of UHF and microwave components can be accomplished provided that the following three ingredients are available: a phase-shift standard or components that can be assembled to serve this purpose, a measurement system, and the techniques for determining the corresponding measurement uncertainty.

Many different phase-shift standards and many different measurement techniques exist today and the development of most of them has been reported in the literature. The results of careful analysis of particular standards and particular measurement systems have been reported and a number of general review papers have also been published [1]–[12]. The limits of uncertainty are discussed to some extent in practically every measurements publication; however, there has been no general discussion of the uncertainties involved in the measurement of phase shift.

The prime objective in this paper is to report on the state-of-the-art of differential phase-shift measurements at UHF and microwave frequencies. Differential phase-shift standards, measurement techniques, and measurement uncertainties are discussed. Technical criteria that should be considered when selecting a standard or a measurement system are given and techniques for minimizing the limits of uncertainty are included. All the discussion is of a general nature; a bibliography is included to supplement the general discussion.

II. Differential Phase-Shift Standards

Differential phase shift is defined in this paper as "the magnitude of the change in phase of a field quantity at the output of a two-port network which is produced by an adjustment of the characteristics of the two-port network." A differential phase-shift standard is then a network, or device, whose phase-shift characteristics are known, within a limit of uncertainty.

The uncertainty associated with a differential phase-shift standard is usually of major significance; therefore the choice of a phase shifter as a standard is governed primarily by the limit of uncertainty that is prescribed for the measurement. Other technical criteria that must be considered when selecting a standard include the phase shifter's loss characteristics, its operational frequency bandwidth, its maximum differential phase shift and its ease of operation.

Each differential phase shifter is designed to operate in either the analog (continuous) or the digital (stepped) mode. The discussion here is centered on analog phase shifters because digital phase shifters have found little application to date as standards. The design and performance of digital phase shifters have been discussed in several articles [33], [35], [38].

The class of analog phase shifters can be subdivided in several different ways. In this discussion they are subdivided into four general groups entitled reflection-type phase shifters, line-stretcher phase shifters, dielectric phase shifters, and electrically controlled phase shifters.

A brief discussion of the principles of operation for a number of devices in each group follows. Each phase shifter is examined with respect to the selection criteria outlined above and the factors which limit the uncertainty are pointed out.

A. Reflection-Type Phase Shifters

Reflection-type differential phase shifters can be quickly constructed in the laboratory from multiport components and sliding short-circuit terminations [35]. Three different arrangements are shown in Fig. 1. In all three arrangements,

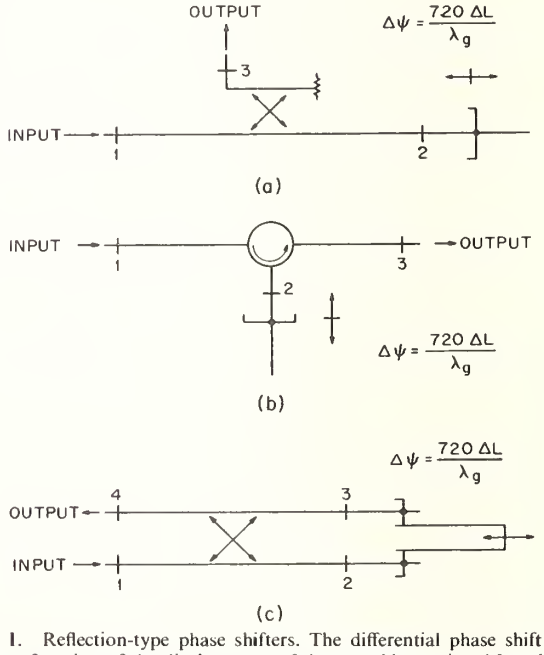


Fig. 1. Reflection-type phase shifters. The differential phase shift $\Delta\psi$ is a function of the displacement of the movable portion ΔL and the waveguide wavelength λg. (a) Directional coupler arrangement. (b) Circulator arrangement. (c) Three-dB symmetric hybrid arrangement.

Manuscript received February 9, 1967; revised March 3, 1967.
The author is with the National Bureau of Standards, Boulder, Colo.

the output signal is derived from the reflected signal. A differential phase shift is produced when the positions of the short circuits are changed. Ideally, the amount of differential phase shift is determined only by the displacement of the short circuits and the waveguide wavelength.

A directional coupler is illustrated in Fig. 1(a). Considerable mismatch exists at the input (port 1) of this arrangement because port 2 is terminated in a short circuit. Assuming that a 10-dB directional coupler is used, the theoretical return loss as seen at the input is 0.92 dB, which corresponds to an input reflection coefficient of approximately 0.9. Analysis shows that both ports must be connected to a well-matched system in order to minimize the uncertainty caused by the mismatch.

The tolerance involved in the measurement of short-circuit displacement, the dimensional tolerance of the waveguide, and the frequency instability of the UHF or the microwave signal each cause various amounts of measurement uncertainty. Analysis shows that a limit of uncertainty of ± 1.0 degree is typical at 10 GHz if commercial components are used. A limit of uncertainty on the order of ± 0.1 degree is possible if matching transformers, a precision waveguide section, and other special equipment are properly used.

The phase shifter's input reflection coefficient is reduced if a circulator is used in place of the directional coupler, as shown in Fig. 1(b). It is necessary to use this phase shifter in a well-matched system in order to maintain a minimum limit of uncertainty. The limit of uncertainty with the circulator arrangement is, in principle, the same as that for a directional coupler arrangement. However, circulators are not widely used in this application because they are not available for all frequencies; those in existence are relatively narrowband devices, and they are not found in as many measurement laboratories as are directional couplers.

A variety of hybrid junctions, such as E-H tees, magic tees [18], 3-dB directional couplers, rat races [20], and ring circuits [20], are suited for use in reflection-type phase shifters. The 3-dB symmetric hybrid, shown in Fig. 1(c), is especially convenient because it is easy to couple together the sliding short ciruits in waveguides which have parallel axes [24].

The signal incident on port 1 is divided equally between ports 2 and 3 and the characteristics of the hybrid cause the signal components emerging from those ports to be in phase quadrature. These signal components travel paths of equal length and are reflected by a pair of short circuits. The reflected signal from each short circuit is divided equally between ports 1 and 4, resulting in two signal components emerging from both ports 1 and 4. The reflected signal components emerging from port 1 cancel while the reflected signal components emerging from port 4 add. Therefore, a reflected signal emerges from port 4 only and the phase shifter is, in principle, matched at its input (port 1).

The uncertainties in this arrangement are caused by superposition of the residual reflections from the phase shifter and from the system to which it is attached, the uncertainty in measuring the displacement of the short circuits, and the frequency instability of the UHF or the microwave signal source. The limit of uncertainty for the hybrid arrangement has a range of $\pm(0.25\text{–}3.0)$ degrees, depending upon the components used and the frequency of the UHF or the microwave signal.

Typical insertion losses of the devices in Fig. 1 are approximately 6, 10.5, or 20 dB for the directional coupler (corresponding to the use of either a 3-, 10-, or 20-dB directional coupler, respectively), 1 dB for the circulator, and 0.1 dB for the hybrid junction. Typically, the change of insertion loss that occurs when the phase shifter is adjusted will be less than 0.05 dB at frequencies up to 12 GHz.

The phase shifters shown in Fig. 1 can be constructed from either coaxial or rectangular waveguide components. The maximum differential phase shift that can be produced with any of the arrangements is determined by the waveguide wavelength and the maximum displacement of the short circuits. For example, a differential phase shift of 360 degrees will be produced in a coaxial waveguide arrangement if the short-circuit position is changed 15 cm when the frequency is 1 GHz and a displacement of 1.5 cm will produce a differential phase shift of 360 degrees if the frequency is 10 GHz.

B. Line-Stretcher Phase Shifters

Line stretchers operate by changing the length of waveguide in a portion of the circuit, which results in a change in the electrical length of that circuit [17], [19], [20], [28]. This change of length is accomplished by sliding together two waveguides of slightly different sizes, as illustrated in Fig. 2.

In the telescoping type, shown in Fig. 2(a), the change in electrical length is proportional to the displacement of the

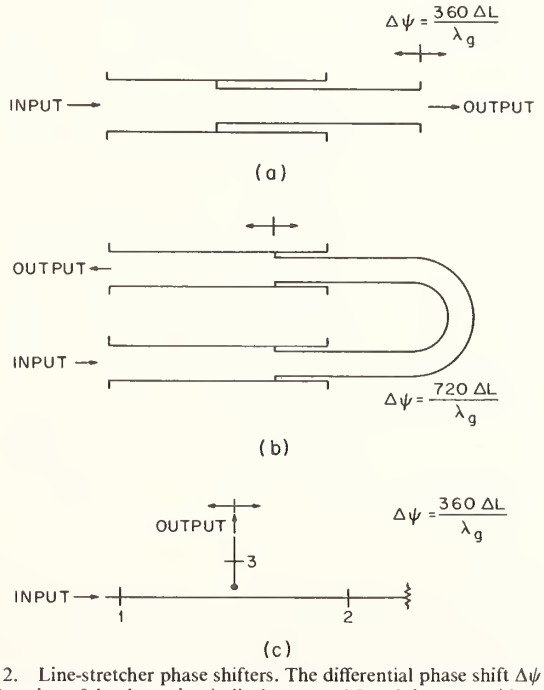


Fig. 2. Line-stretcher phase shifters. The differential phase shift $\Delta\psi$ is a function of the short-circuit displacement ΔL and the waveguide wavelength λg. (a) Telescoping-type line stretcher. (b) Trombone-type line stretcher. (c) Slotted-line-type line stretcher.

385-582 O - 70 - 16

movable portion, while in the trombone type, shown in Fig. 2(b), the change in electrical length is proportional to twice the displacement of the movable portion. This is significant in that the telescoping type must have a displacement capability of twice that for the trombone type in order to produce an identical differential phase shift, assuming that the waveguide wavelength is the same in both configurations. But for a given uncertainty in displacement measurement, the phase-shift uncertainty of the trombone type is twice that for the telescoping type. However, the trombone type is usually used because of its convenience of installation and operation.

It is ordinarily required that a line stretcher be capable of producing a differential phase shift of at least 180 degrees if it is to be useful. Thus, the maximum change of length of a line stretcher usually limits the lower frequency of operation. If a telescoping line stretcher of coaxial waveguide has a maximum length variation of 15 cm, for example, the lower frequency limit would be 1 GHz. It would still operate satisfactorily at lower frequencies, but it would produce less than 180 degrees of phase shift. For rectangular waveguide line stretchers, the practical lower frequency limit is that recommended for the waveguide, but it would still operate at frequencies down to the cutoff frequency of the waveguide.

The upper frequency limit of a line stretcher is usually determined by its deterioration of performance due to reflections whose effects increase with frequency. These reflections arise from unavoidable discontinuities which occur at the steps where a telescoping tube slides over another tube of slightly smaller dimensions. Usually attempts are made to compensate for these discontinuities, but as the frequency increases the compensation will eventually become ineffective. In the case of a rectangular waveguide line stretcher, the practical upper frequency limit may be that recommended for the waveguide so that higher modes are not excited.

In addition to the need to compensate for unavoidable discontinuities, there are several other considerations which are important in the design of line stretchers: 1) it is desirable to maintain a constant impedance as the line is lengthened, even though waveguide dimensions must be different in the movable portion than in the fixed portion, 2) it is desirable to have no "noise" or variation of impedance caused by sliding contacts or the transition from the fixed portion to the movable portion, 3) leakage from the unavoidable joint between fixed and movable portions of the line stretcher should be minimized, 4) the waveguide of which the line stretcher is made should have uniform dimensions, and 5) the drive mechanism (if used) for the movable portion should operate smoothly.

The limit of uncertainty of commercial coaxial and rectangular waveguide line stretchers ranges from $\pm(0.1\text{–}1.0)$ degree for the frequency range of 1–12.4 GHz, the uncertainty increasing with frequency. The bulk of the uncertainty is caused by interaction between the residual reflections from the line stretcher and the system where it is attached, the uncertainty in measuring the displacement of the movable section, and the frequency instability of the UHF or the microwave signal source.

The insertion loss for both the telescoping and the trombone line stretcher is low, 0.5 dB or less. The change of insertion loss caused by changes in line length is less than 0.05 dB at frequencies up to 12.4 GHz. The maximum differential phase shift that can be produced with either the telescoping or the trombone type is dependent upon the maximum displacement of the movable portion and the waveguide wavelength.

Another arrangement which is classified as a line stretcher in this discussion is shown in Fig. 2(c). The slotted section is terminated in a matched load. A small portion of the UHF or the microwave signal is coupled out of the slotted section with a probe, and a differential phase shift is produced by changing the probe position. A major uncertainty is caused by the mismatch of the terminating element. As an example, if the terminating element has a VSWR of 1.10, uncertainty in the phase shift is on the order of ± 2.5 degrees, while a VSWR of 1.50 could result in an uncertainty of ± 12.0 degrees. The maximum differential phase shift that can be produced with this arrangement is dependent upon the maximum displacement of the probe and the waveguide wavelength.

The insertion loss of the slotted-line device is typically 20 dB and the typical change of insertion loss over the range of phase shifts is less than 0.05 dB.

C. Dielectric Phase Shifters

A differential phase shift is produced when a strip of dielectric material is placed within a waveguide as shown in Fig. 3(a). The change is small if the dielectric strip is placed in the weakest portion of the electric field and a larger change is produced if the dielectric strip is placed in a

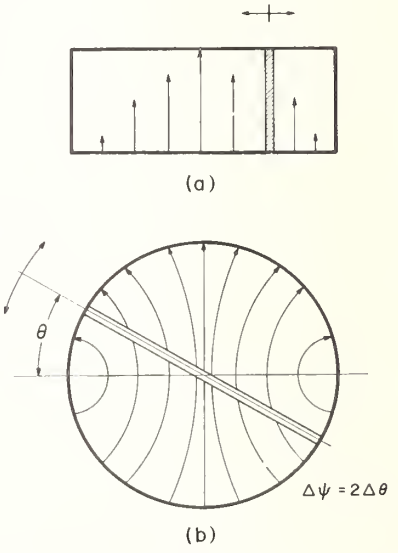


Fig. 3. Dielectric-vane phase shifters. The differential phase shift $\Delta\psi$ of the rotary-vane phase shifter is a function of the vane rotation $\Delta\theta$. (a) Dielectric vane in rectangular waveguide (TE_{10} mode). (b) Rotary-vane phase shifter in circular waveguide (TE_{11} mode). The differential phase shift $\Delta\psi$ is a function of the differential rotation angle $\Delta\theta$.

stronger portion of the electric field. Hence, a differential phase shifter can be constructed by mounting a movable dielectric vane [17], [19] in rectangular waveguide (TE_{10} mode). Vane-type attenuators can be converted into differential phase shifters by replacing the lossy vane with a dielectric vane. Dielectric materials such as polystyrene and glass have been used in that conversion.

The maximum differential phase shift that can be produced with the arrangement shown in Fig. 3(a) is a function of the vane dimensions, the vane material, and the microwave frequency. Usually a differential phase shift of at least 90 degrees can be produced with a repeatability ranging from a few tenths of a degree to ± 3.0 degrees. These phase shifters are not direct reading devices; they must be calibrated at each frequency of use because their characteristics are not accurately predictable.

In the rotary-vane phase shifter [16], [21], a differential phase shift is produced when the angular position of a dielectric vane is changed, as shown in Fig. 3(b). Because the principle of operation of the rotary-vane phase shifter cannot be briefly summarized, it is not discussed here. In practice, circular waveguide operating in the circularly polarized dominant (TE_{11}) mode is used. Fortunately, the differential phase shift of the device is nearly independent of the microwave frequency so that the phase shifter's dial can be marked in degrees and this same scale can be used for any frequency recommended for that waveguide size.

Rotary-vane phase shifters are commercially available for the frequency range 5.0–110 GHz. They are manufactured with rectangular waveguide inputs and outputs. Their specified limits of uncertainty range between $\pm$(2–5) degrees, depending upon the waveguide size and the frequency of operation. A differential phase shift of unlimited magnitude can be produced if one continues to rotate the dielectric vane. The insertion loss for these devices ranges from 1.0 to 2.0 dB for any frequency and for any dielectric vane angle.

D. Electrically Controlled Phase Shifters

A number of electrically controlled differential phase shifters have been developed in recent years and the most prominent in this category are the ferrite phase shifters [22], [32], [34], [35].

The ferrite phase shifter consists of a ferrite material placed in a homogeneous dc magnetic field within a waveguide and an arrangement for controlling the intensity of the magnetic field. The maximum differential phase shift for these devices is typically 500 degrees at 10 GHz. The greatest difficulty with the ferrite phase shifter is its sensitivity to temperature variations.

Other electrically controlled phase-shifter arrangements are possible [25], [35], such as the use of varactors with 3-dB hybrids, with directional couplers, or with circulators; however, the difficulties of these arrangements limit their use as standards. The difficulties include 1) nonlinearity of bias voltage vs. phase-shift characteristics, 2) generation of harmonics at high power drive levels due to the inherent nonlinearities, 3) self-biasing which changes the phase-shift characteristics with respect to biasing current, and 4) insertion-loss variation which is a function of the bias level and the level of the UHF or the microwave signal.

III. Differential Phase-Shift Measurement Systems

The technical criteria that govern the selection of a measurement system are nearly those given for a phase-shift standard. Still of prime importance is the limit of uncertainty that is prescribed for the measurement; however, the operational bandwidth of the system, its ease of operation, and the effect that an attenuation change has upon system performance must also be considered.

In this paper three general groups of measurement systems are considered: single-channel systems, dual-channel systems, and IF systems. The discussion for the systems in each group is based upon basic block diagrams which show only the essentials of the system; matching devices, power supplies, wavemeters, etc., are not shown.

This discussion is centered upon fixed-frequency measurement systems but it should be noted that equipment has been developed so that most of the systems are adaptable to swept-frequency and pulsed-signal operation. References to articles which treat measurement under swept-frequency or pulse-signal operation are included in the bibliography [50], [63], [74].

A. Single-Channel System

Simple arrangements for measuring the differential phase shift of low-loss, reciprocal components are shown in Fig. 4. In Fig. 4(a), a slotted section serves as a null position detector. A minimum of the standing wave pattern is used as a reference to position the probe when the device under test

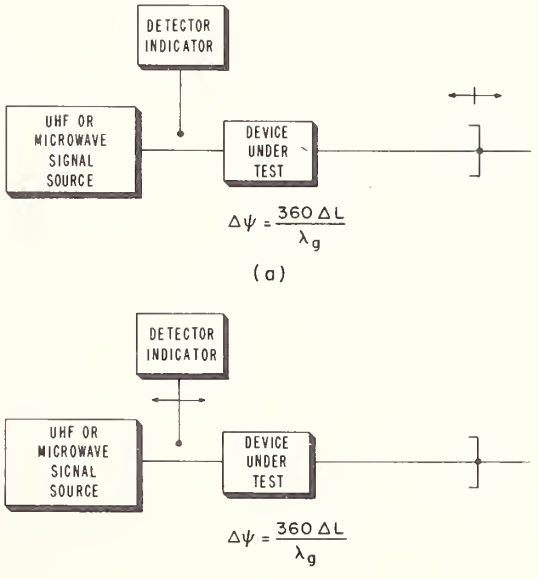


Fig. 4. Single-channel system arrangements for calibrating low-loss reciprocal components. The differential phase shift $\Delta\psi$ is a function of the short-circuit (or probe) displacement ΔL and the waveguide wavelength λg. (a) Single-channel system using a calibrated short circuit as the phase-shift standard. (b) Single-channel system using a slotted section as the phase-shift standard.

is adjusted to its initial dial setting. After the device under test is adjusted to a new dial setting, the minimum of the standing wave pattern is restored to the reference position of the probe by moving the calibrated short circuit [39], [40]. The calibrated short circuit is, in this arrangement, the differential phase-shift standard.

In Fig. 4(b), the position of the probe is adjusted to restore the minimum of the standing wave pattern to its reference position [39]. The slotted section in this arrangement serves as both the detector and the phase-shift standard.

An accurate calibration cannot be obtained with the measurement systems of Fig. 4 unless the device under test has very low reflections (mismatch). Considering the maximum mismatch for a commercial reciprocal phase shifter (VSWR = 1.35), the limit of uncertainty is on the order of ± 17.0 degrees when calibrated with either of the arrangements of Fig. 4. If the device under test has an input VSWR of 1.15, the limit of uncertainty is reduced to ± 8.0 degrees.

B. Basic Two-Channel System

If the measurement system shown in Fig. 5(a) is used, the mismatch uncertainty present in the single-channel systems can be reduced because the device under test is isolated from the phase-shift standard. For theoretical purposes, the isolation between the standard and the device under test is assumed infinite and in the practical situation it must be high (60 dB or more). A high isolation eliminates the mismatch interaction between the device under test and the standard. Furthermore, tuners can be used for minimizing the uncertainty caused by the superposition of residual reflections.

The two-channel terminology is used in this paper when the UHF or the microwave energy in both channels is derived from a common signal generator. The signal which propagates through the phase-shift standard is called the reference signal, while the signal which propagates through the device under test is called the test signal.

A null occurs at the detector of Fig. 5(a) when the reference and test signals have equal amplitudes and their phase angles differ by 180 degrees. With the device under test adjusted to its initial dial setting, a null is established in the detector by adjusting both the phase-shift standard and the variable attenuator. The device under test is then adjusted to its next dial setting and, assuming that the levels of the reference and test signals remain constant, the null is restored in the detector by readjustment of the phase-shift standard. The differential phase shift of the device under test is then equal to the differential phase shift required of the standard for restoring the null.

If the device under test or the standard should introduce amplitude variations along with the differential phase changes, a minimum signal rather than a null will be established at the output of the detector. The phase resolution of the system is seriously degraded if the amplitude of one of the signals differs from the amplitude of the other by 3 dB or more. If their amplitudes differ by 10 dB, it may be difficult to locate a minimum signal.

A null can be restored at each calibration point with the additional adjustment of a variable attenuator. If the amplitudes of the reference and test signals are equalized at each calibration point, maximum phase resolution is maintained in the measurement system. However, the attenuator itself may introduce small phase changes along with the amplitude changes. In precise measurements the phase changes caused by the attenuator should be taken into consideration in order to obtain a suitable measurement uncertainty.

Considerable effort has been spent to reduce the uncertainty of this system since it was originally developed. An uncertainty on the order of ± 0.3 degree is possible if a system is carefully constructed and a precise differential phase shift standard is used [42].

The instrumentation used to indicate the null, or minimum signal condition, in this two-channel system can have a wide variety of configurations. A simple arrangement consists of connecting a standing wave indicator to the crystal video detector, provided that the UHF or microwave signal

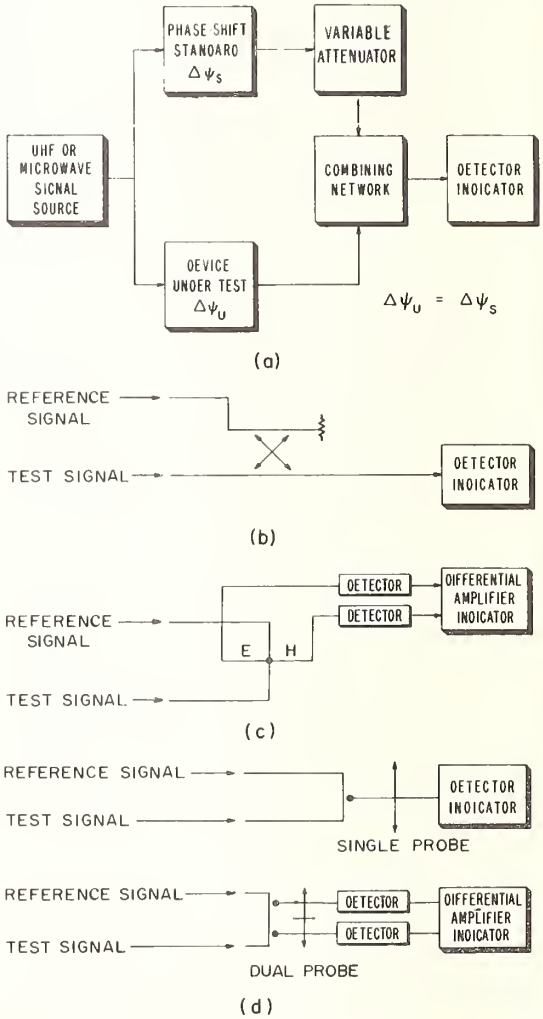


Fig. 5. Block diagram of a basic two-channel phase-shift measurement system and various signal-combining networks. (a) Block diagram of a basic two-channel system. (b) A directional coupler is the combining network. (c) A magic tee is the combining network. (d) Slotted sections are the combining networks.

is amplitude-modulated at its source. Usually the UHF or the microwave signal is amplitude-modulated at its source for instrumentation purposes; however, modulation is not essential for operation of the system.

Other instrumentation techniques are shown in Fig. 5. The signals from the two channels can be recombined with a directional coupler [Fig. 5(b)] magic tee [Fig. 5(c)], or slotted section [Fig. 5(d)]. The magic tee of Fig. 5(c) is particularly useful because it is arranged so that a null is established at each calibration point regardless of any attenuation changes that may occur [41], [53], [58]. Thus the necessity of establishing a null with an auxiliary variable attenuator is eliminated.

By injecting reference and test signals into opposite ends of the slotted section, as shown in Fig. 5(d), a standing wave pattern is set up. The probe is manually positioned so as to obtain a null at the detector output (or a minimum signal is observed at the detector output if the reference and test signals are not of equal amplitude). A phase change caused by adjustment of the device under test will cause the position of the null (or minimum) to shift from its initial position within the slotted section, and the differential phase change produced by adjustment of the device under test is proportional to the distance that the null (or minimum) has shifted. (The proportionality constant is 2β in the ideal case, where β is the phase constant of the slotted section.) A single-probe arrangement can be used to measure the distance that the null (or minimum) of the standing wave pattern has shifted; however, a dual-probe arrangement has also been used [44]. The slotted section serves as both a signal recombiner and a phase-shift standard.

C. Two-Channel Systems Which Use Amplitude Modulation in One Channel Only

The two-channel systems described here utilize amplitude modulation in only one channel, as shown in Fig. 6. The test signal is amplitude-modulated at some convenient frequency (usually between 1 and 10 kHz). The modulated test signal is combined with the reference signal and their resultant goes into the detector.

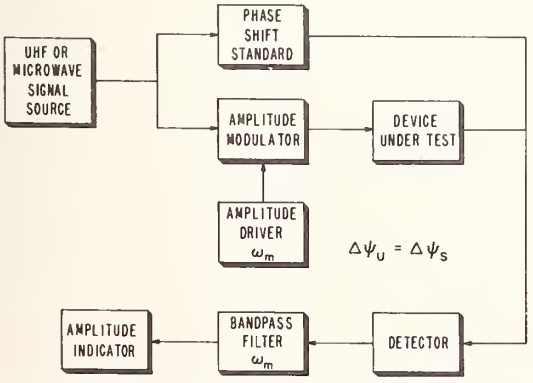


Fig. 6. Block diagram of a two-channel phase-shift measurement system which requires amplitude modulation in only one channel. The differential phase shift $\Delta\psi_u$ produced by the device under test is equal to the differential phase shift $\Delta\psi_s$ required from the standard to restore a null.

The bandpass filter is tuned to the modulation frequency. The amplitude of the detected signal is dependent upon the amplitudes of the reference and test signals and the difference between their phase angles. If the amplitudes of both signals are held constant but the phase angle of one is adjusted, the amplitude of the modulation-frequency signal from the filter will vary between a null and a maximum, provided the amplitude of the test signal is smaller than that of the reference signal. In contrast with the case of a standing wave in a slotted line, here it is not necessary to maintain reference and test signals of equal amplitude to establish a null at the indicator. A null will be established in this system for any relative-amplitude relationship.

Two different types of amplitude modulation have been used in this system. The original arrangement used a double-sideband-suppressed-carrier modulator (usually called a balanced modulator) [61]. If a balanced modulator is used, a null occurs at the indicator when the difference between the phase angles of the reference and test signals at the detector input is 90 degrees. Since the phase-angle difference at the null condition (hereafter called the null angle) is independent of the amplitudes of the reference and test signals, the balanced-modulator arrangement is very useful for measuring the differential phase-shift characteristics of lossy devices such as variable attenuators.

One difficulty with a balanced-modulator arrangement lies in trying to locate one which completely suppresses the carrier signal. Usually a residual signal at the carrier frequency is present at the output of the balanced modulator. This is undesirable because the null angle is no longer 90 degrees but is now dependent upon the amplitudes of the test signal and the residual carrier signal.

A system that uses an amplitude modulator which does not suppress the carrier has been developed [62], [64]. In this system, called the modulated-subcarrier system, a change of attenuation in one of the channels will cause the null angle to change. It is common practice to arrange for the magnitude of the test signal to be much smaller (by 40 dB or more) than the magnitude of the reference signal. Then amplitude changes of the test signal will cause the null angle to change only an insignificant amount. In addition, a technique has been developed [62] for canceling the effect of attenuation changes. With that technique, the modulated-subcarrier system can also be used to measure the differential phase-shift characteristics of lossy devices, such as variable attenuators.

Differential phase shifters are usually calibrated in the systems of Fig. 6 by means of the nulling technique described for the basic two-channel system. It is possible, however, to indicate the phase change in other ways, such as with an oscilloscope [65] or with a ratio detector [63], [66].

D. IF Systems

Differential phase-shift measurement systems whose operation depends on the generation of a suitable intermediate frequency (IF) are distinguished from two-channel systems in that two UHF or two microwave signal sources are required to generate the IF signal. Differential phase-

shift information can be transferred from a UHF or a microwave signal to a signal of lower frequency by two different methods. In both, a differential phase change in the UHF or the microwave signal appears as a differential phase change of equal magnitude in the lower frequency (IF) signal. The actual phase measurement is done at the IF.

Consider first the heterodyne arrangement shown in Fig. 7(a). Mixer 1 develops an IF signal which is used as a reference signal in the phase meter. Mixer 2 develops an IF signal whose phase angle tracks the change of phase angle of the UHF or the microwave test signal. An IF phase meter, or other devices such as a resolver, can be used to measure the difference between the phase angles of the two IF signals.

In the system which uses a single-sideband (SSB) modulator, shown in Fig. 7(b), the UHF or the microwave signal in the modulated arm is offset in frequency (the frequency can be either increased or decreased) by an amount equal to the modulation frequency. Normally the carrier is suppressed in these systems; however, the theory of measurement for the system where the carrier is unsuppressed has also been developed [69].

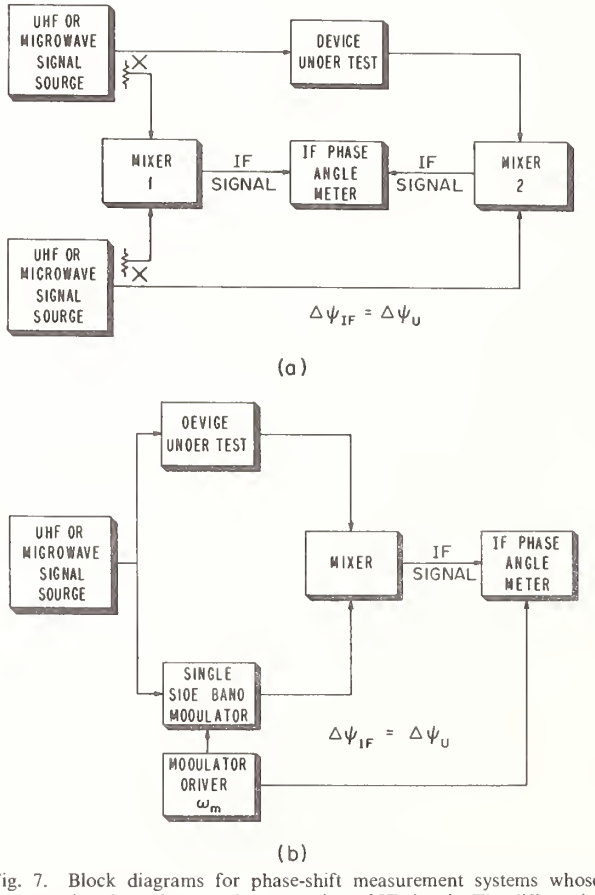


Fig. 7. Block diagrams for phase-shift measurement systems whose operation depends upon the generation of IF signals. The differential phase shift $\Delta\psi_u$ produced by the device under test is equal to the differential phase shift of the intermediate frequency signal $\Delta\psi_{IF}$. (a) Heterodyne phase-shift measurement system which uses two UHF or microwave oscillators. (b) Heterodyne phase-shift measurement system which uses one UHF or microwave oscillator and a single-sideband modulator.

The modulator in the arrangement shown in Fig. 7(b) is considered to be the second UHF or microwave signal source. The reference and test signals are combined in the mixer and the frequency of the signal out of the mixer is then the same as the modulation frequency. The differential phase shift of the device under test is determined by measuring the difference between the phase angles of the two IF signals.

The major disadvantage with the SSB arrangement is the unavailability of suitable SSB modulators. Usually SSB modulation is accomplished by applying sawtooth modulation to the helix of a traveling wave tube [71], [73]; this is often referred to as the serrodyne process.

IV. The Limits of Uncertainty

The limit of uncertainty for each measurement system is a function of the measurement technique and the quality of the components used to instrument that technique. Although each measurement system must be analyzed in relation to its own measurement uncertainties, some uncertainties are common to all systems. The purpose of this section is to point out these common uncertainties and to offer techniques for minimizing their magnitudes.

A. Measurement Uncertainty Caused by Mismatch

This uncertainty is usually the most significant of all those present in a measurement system. It has been carefully analyzed [83]–[85] and the results cannot be easily summarized. Therefore, only the results of sample calculations will be given here to demonstrate the magnitude of this uncertainty.

Basically the uncertainty is caused by the superposition of reflections from the phase-shift standard and reflections from the measurement system as seen at the insertion-point terminals. Another measurement uncertainty is caused by the superposition of reflections from the device under test and reflections from the measurement system at its insertion-point terminals [82]. Both of these uncertainties are eliminated if the system is reflectionless at the insertion-point terminals.

For an example, assume that the device under test has a maximum VSWR of 1.35 at both ports and it is connected into a system whose insertion-point VSWR's are both 1.10. The limit of uncertainty caused by the superposition of these reflections is ± 2.0 degrees. If the system VSWR's are both reduced to 1.05, the limit of uncertainty is reduced to ± 0.90 degree. If the system VSWR's are both reduced to 1.006 and the device under test has VSWR's of 1.15 at both ports, the limit of uncertainty is ± 0.05 degree. Tuners are generally used to reduce these uncertainties to a tolerable level by reducing the reflections caused by a mismatched generator or detector in the system.

B. Measurement Uncertainty Caused by Frequency Instability

Most differential phase-shift measurement systems are two-channel systems and thus contain two propagation paths between the generator and the detector. Because the electrical lengths of the two channels are seldom equal, a differential phase shift between the reference and test signals may be produced at the detector as the frequency of the UHF or the microwave signal varies. This uncertainty is related to the frequency instability of the signal source, the propagation constants of the components in those paths, and the differential path length [68]. It can be minimized by equalizing the electrical path lengths between the generator and the detector. Stabilizing the frequency of the UHF or the microwave signal source is another means of minimizing this uncertainty.

The magnitude of this uncertainty is usually negligible when compared with some of the other uncertainties. As an example, assume that the frequency in a rectangular waveguide system is 10 GHz, that the frequency stability is 1 part in 10^6, and that the electrical path-length differential is 61 cm. For these conditions, the limit of uncertainty is approximately ± 0.01 degree. If the frequency is stabilized to 1 part in 10^8, the limit of uncertainty is reduced to approximately $\pm 10^{-4}$ degrees.

C. Measurement Uncertainty Caused by Laboratory Environmental Changes

Even though the environment within many calibration laboratories is closely controlled, inhomogeneous temperature variation may still occur. These temperature variations may be of sufficient magnitude to cause changes in length and cross-sectional dimensions of the waveguide. In general, these dimensional fluctuations will cause a differential phase shift between the reference and the test signal. In practice this problem is minimized by allowing sufficient equipment warm-up time and by collecting the calibration data as rapidly as possible.

D. Measurement Uncertainty Caused by Detector Nonlinearities, IF Amplifier Characteristics, and IF Phase Meters

This measurement uncertainty is that associated with the UHF or the microwave detectors and with the circuitry following detection. In some measurement systems, such as the heterodyne system, this uncertainty can be appreciable because tuned amplifiers and tuned phase meters are usually used. In other systems a linear-detector characteristic or a pair of matched detectors are required for the correct indication of the differential phase shift.

Each component used in each system must be selected and evaluated for the specific situation. For example, an amplifier used in the basic two-channel system need not be considered as carefully as an amplifier used in the heterodyne measurement system.

As an example of the limits of the uncertainty considered here, signal-input-level variations will produce a differential phase shift of less than ± 1.0 degree in some IF amplifiers [87] if the amplifier is not saturated. Frequency instability may also produce an appreciable differential phase shift because IF amplifiers contain tuned circuits.

A variety of direct-reading phase-shift meters now exist. The limit of uncertainty for these devices ranges from a few tenths of a degree to ± 5.0 degrees. Phase-shift resolvers which operate at the IF can also be used as the phase-shift standard and their limit of uncertainty can be as low as ± 0.30 degree [81].

E. Measurement Uncertainties Caused by Leakage Signals

The RF energies that reach the UHF or the microwave detector through spurious paths are called leakage signals. Undesired or spurious IF signals are also called leakage signals. Leakage signals may occur from connector junctions, from components in the system, from interconnecting cables, or from imperfectly shielded signal sources.

An RF or an IF leakage signal of any magnitude will introduce a measurement uncertainty. This uncertainty is minimized when the magnitude of the leakage signal is minimized. Proper shielding of leaky components and the use of RF seals at connector junctions are the most effective means of reducing RF leakage signals. The IF leakage signals can be minimized by proper shielding, by proper filtering, and by assuring that all IF equipment operates from a common ground.

The limit of uncertainty caused by a leakage signal is given (in radians) by the ratio of the amplitudes of the leakage signal voltage to the desired signal voltage, where the leakage-signal amplitude is much smaller than that of the desired signal. A limit of uncertainty of ± 0.006 degree is possible if the amplitude of either an RF or an IF leakage signal is known to be 80 dB below the amplitude of the desired signal. A leakage signal whose amplitude is known to be 60 dB below the amplitude of the desired signal has a limit of uncertainty of ± 0.057 degree, while a leakage signal level that is 40 dB below the level of the desired signal has a limit of uncertainty of ± 0.57 degree.

F. Measurement Uncertainty Caused by the Phase-Shift Standard

A major uncertainty in a differential phase-shift measurement is the uncertainty associated with the phase-shift standard. In order to arrive at a conservative limit of uncertainty for the standard, one should fully understand the theory and the operation of the device. With that knowledge, one will be able to evaluate it in terms of the UHF or the microwave frequency instability, the change in laboratory environment, the mechanical limitations, etc. Some differential phase shifters have already been evaluated and their limits of uncertainty established [79], [80]–[82].

V. Conclusions

A variety of differential phase-shift standards and phase-shift measurement techniques have been described. Usually

the limit of uncertainty required by the application determines which of those standards and measurement systems are suitable for a particular measurement; however, other criteria such as the operational frequency bandwidth and the ease of operation must also be considered.

Even though the limit of uncertainty for any standard or measurement system is a function of many error sources, the major portion is usually contributed by the interactions between the reflections from the standard (or the device under test) and from the measurement system. A limit of uncertainty on the order of ± 5.0 degrees can be expected with a rather crude arrangement; however, highly refined equipment and measurement procedures are required to achieve a limit of uncertainty of ± 0.20 degree or less.

Bibliography

A. Differential Phase-Shift Measurement

[1] R. A. Sparks, "A comparative review of phase measurement methods at microwave frequencies," *Proc. 5th Nat'l Conv. on Military Electronics*, p. 65, June 1961.

[2] P. Lacy, "Analysis and measurement of phase characteristics in microwave systems," *WESCON Conv. Rec.*, August 1961.

[3] R. A. Sparks, "Microwave phase measurements," *Microwaves*, vol. 2, pp. 14–25, January 1963.

[4] R. W. Beatty, "Microwave standards and measurements, A progress review 1960 to 1963," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-12, pp. 134–138, December 1963.

[5] "URSI National Committee Report, 14th General Assembly, Tokyo, Commission I. Radio and Measurement Methods and Standards, September (1963)," *J. Res. NBS (Radio Science)/USNC-URSI*, vol. 68D, pp. 523–546, May 1964.

[6] M. Magid, "UHF measurements," *Electro-Tech.*, pp. 67–68, October 1964.

[7] T. Mukaihata. "Technique of microwave phase measurements," presented at the 1964 19th Annual ISA Conf. and Exhibit, New York, N. Y.

[8] Y. P. Yu, "Purpose dictates choice of phase meters," *Electronic Design*, pp. 128–135, May 3, 1965.

[9] R. W. Beatty, *Progress in Radio Science*, vol. 1. Amsterdam, Holland: Elsevier, p. 79, 1965.

[10] M. C. Selby, "The system of electromagnetic quantities at 30 kHz to 1 GHz," *Metrologia*, vol. 2, pp. 37–45, January 1965.

[11] R. W. Beatty, "The system of electromagnetic quantities at frequencies above 1 GHz," *Metrologia*, vol. 2, pp. 46–54, January 1966.

[12] J. D. Dyson, "The measurement of phase at UHF and microwave frequencies," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-14, pp. 410–423, September 1966.

B. Definitions of Terms

[13] A. G. McNish, "Classification and nomenclature for standards of measurement." *IRE Trans. on Instrumentation*, vol. I-7, pp. 371–378, December 1958.

[14] "IRE Standards on Antennas and Waveguides: Waveguide and Waveguide Component Measurements, 1959," *Proc. IRE*, vol. 47, pp. 568–582, April 1959.

[15] R. W. Beatty, "Some basic microwave phase shift equations," *J. Res. NBS (Radio Science)*, vol. 68D, pp. 349–353, April 1964.

C. Differential Phase Shifters

[16] A. G. Fox, "An adjustable wave-guide phase changer," *Proc. IRE*, vol. 35, pp. 1489–1498, December 1947.

[17] "A survey of phase changers used in the microwave region," Polytechnic Institute of Brooklyn, Brooklyn, N. Y., Rept. P-175-48, PIB-120, August 30, 1948.

[18] R. H. Reed, "Modified magic tee phase-shifter," *IRE Trans. on Antennas and Propagation*, vol. PGAP-1, pp. 126–134, February 1952.

[19] F. E. Terman and J. M. Pettit, *Electronic Measurements*. New York: McGraw-Hill, 1952.

[20] H. J. Reich, P. F. Ordung, H. L. Krauss, and J. G. Skalnik, *Microwave Theory and Techniques*. Princeton, N. J.: Van Nostrand, 1958.

[21] "A new precision wave guide phase shifter," *Hewlett-Packard J.*, vol. 6, January 1955.

[22] C. M. Johnson, "Ferrite phase shifter for the UHF region," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-7, pp. 27–31, January 1959.

[23] C. F. Augustine and J. Cheal, "The design and measurement of two broad-band coaxial phase shifters," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-8, pp. 398–402, July 1960.

[24] R. J. Mohr, "Some design aspects of components utilizing symmetric 3 dB hybrids," *Microwave J.*, vol. 5, pp. 90–94, June 1962.

[25] H. N. Dawirs and W. G. Swarner, "A very fast, voltage-controlled, microwave phase shifter," *Microwave J.*, vol. 5, pp. 99–107, June 1962.

[26] J. Clark and G. R. Harrison, "Miniaturized coaxial ferrite devices," *Microwave J.*, vol. 5, pp. 108–118, June 1962.

[27] R. Stobbe, "Electrostatic phase shifter," *Electronic Products*, pp. 42–43, April 1963.

[28] G. L. Ragan, *Microwave Transmission Circuits*. Lexington, Mass.: Boston Technical Publishers, Inc., 1964.

[29] J. L. Altman, *Microwave Circuits*. Princeton, N. J.: Van Nostrand, 1964.

[30] R. W. Beatty, "A differential microwave phase shifter," *IEEE Trans. on Microwave Theory and Techniques (Correspondence)*, vol. MTT-12, pp. 250–251, March 1964.

[31] D. A. Ellerbruch, "Analysis of a differential phase shifter," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-12, pp. 453–459, July 1964.

[32] H. S. Jones, Jr. and F. Reggia, "Design of three miniature ferrite phase shifters," *Microwaves*, vol. 3, pp. 28–32, July 1964.

[33] J. D. Woermbke and J. A. Myers, "Latching ferrite microwave devices," *Microwaves*, vol. 3, pp. 20–27, October 1964.

[34] J. B. Brinton, Jr., "Electrically-controlled phase shifters," *Microwaves*, vol. 3, pp. 72–79, December 1964.

[35] K. Molz, N. Pribble, and B. Dodson, "A review of microwave phase shifters," *Microwaves*, vol. 3, pp. 80–83, December 1964.

[36] D. A. Caswell, "Understanding field displacement and differential phase shift," *Microwaves*, vol. 4, pp. 94–99, March 1965.

[37] W. J. Bleackley, "A rotary helical coaxial phase shifter," *Microwave J.*, vol. 9, pp. 38–40, February 1966.

[38] L. R. Whicker and R. R. Jones, "Design guide to latching phase shifters," *Microwaves*, vol. 5, pp. 31–39, November 1966.

D. Single-Channel Systems

[39] *Handbook of Microwave Measurements*, vol. I. Brooklyn, N. Y.: Polytechnic Institute of Brooklyn Press, 1963.

[40] A. L. Lance, *Introduction to Microwave Theory and Measurements*. New York: McGraw-Hill, 1964.

E. Basic Two-Channel Systems

[41] A. Slocum and C. F. Augustine, "6 KMC phase measurement system for traveling wave tubes," *IRE Trans. on Instrumentation*, vol. PGI-4, pp. 145–149, October 1955.

[42] M. Magid, "Precision microwave phase shift measurements," *IRE Trans. on Instrumentation*, vol. I-7, pp. 321–331, December 1958.

[43] Y. P. Yu, "Phase measurement at high frequency," *Proc. Nat'l Electronics Conf.*, vol. 16, pp. 677–684, October 1960.

[44] P. Lacy, "A versatile phase measurement method for transmission-line networks," *IRE Trans. on Microwave Theory and Techniques (Correspondence)*, vol. MTT-9, pp. 568–569, November 1961.

[45] J. R. Alday, "A simple method for measuring the phase shift and attenuation through active microwave networks," *IRE Trans. on Microwave Theory and Techniques (Correspondence)*, vol. MTT-10, p. 143, March 1962.

[46] J. A. Kaiser, H. B. Smith, Jr., W. H. Pepper, and J. H. Little, "An automatic microwave phase comparator," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-10, pp. 548–550, November 1962

[47] C. G. Montgomery, *Technique of Microwave Measurements*. Lexington, Mass.: Boston Technical Lithographers, Inc., 1963.

[48] R. M. Hollis and F. M. Kudo, "A simple phase-shift measurement method for phase modulated microwave signals," *Microwave J.*, vol. 6, pp. 83–86, September 1963.
[49] G. J. Wheeler, *Introduction to Microwaves*. Englewood Cliffs, N. J.: Prentice-Hall, 1964.
[50] R. A. White, "Swept frequency measurement of phase shift and gain of a pulsed microwave amplifier," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-13, pp. 81–88, June–September 1964.
[51] G. Lisitano, "Automatic phase-measuring system for an 8-mm carrier wave and its 4-mm harmonic," *Rev. Sci. Instr.*, vol. 36, pp. 364–367, March 1965.
[52] R. J. King, "An amplitude and phase measuring system using a small modulated scatterer," *Microwave J.*, vol. 8, p. 51, March 1965.
[53] Y. L. Bobrovskii, "Analysis of a bridge circuit for studying phase characteristics of UHF amplifiers," *Izmeritel'naya Tekhnika* (*USSR*), pp. 42–43, November 1965.
[54] A. L. Gardner, "Null method of measuring microwave phase shifts," *Rev. Sci. Instr.*, vol. 37, pp. 23–28, January 1966.
[55] N. B. Hamlin, "Double probe phase meter is simple and accurate," *Microwaves*, vol. 5, p. 42, January 1966.
[56] H. D. Dickstein, "Near field phase measurements using standard microwave equipment," *Microwave J.*, vol. 9, p. 55, February 1966.
[57] V. A. Potapov," Calibration of a phase shifter by the method of splitting an interval with a known phase difference," *Izmeritel'naya Tekhnika* (*USSR*), pp. 60–63, February 1966.
[58] C. F. Augustine, "Insertion loss insensitive phase bridge," Weinschel Engrg. Co., Gaithersburg, Md., Application Note 7.
[59] "A technique for calibrating phase shifters to high accuracy," Hewlett-Packard Co., Service Note 885A-1.
[60] N. W. Harris, "Dynamic phase shift measurement," *Proc. IEEE* (*Letters*), vol. 54, p. 1979, December 1966.

F. Two-Channel Systems Which Require Modulation

[61] S. D. Robertson, "A method of measuring phase at microwave frequencies," *Bell Sys. Tech. J.*, vol. 28, pp. 94–103, 1949.
[62] G. E. Schafer, "A modulated subcarrier technique of measuring microwave phase shift," *IRE Trans. on Instrumentation*, vol. I-9, pp. 217–219, September 1960.
[63] S. B. Cohn and H. G. Oltman, "A precision microwave phase-measurement system with sweep presentation," *IRE Internat'l Conv. Rec.*, pt. 3, pp. 147–150, March 1961.
[64] "Microwave phase shift measurements using a modulated subcarrier," *Electrical Design News*, January 1962.
[65] R. A. Sparks, "A phase measuring system for short RF pulses," *IRE Trans. on Instrumentation*, vol. I-11, pp. 298–302, December 1962.
[66] S. B. Cohn and N. P. Weinhouse, "An automatic microwave phase measurement system," *Microwave J.*, vol. 7, pp. 49–56, February 1964.
[67] R. W. Burton, "A coaxial amplitude-insensitive phase-detection system," *Microwave J.*, vol. 7, pp. 51–53, April 1964.
[68] D. A. Ellerbruch, "Evaluation of a microwave phase measurement system," *J. Res. NBS*, vol. 69C, pp. 55–65, January–March 1965.

G. IF Systems

[69] R. Mittra, "An automatic phase-measuring circuit at microwaves," *IRE Trans. on Instrumentation*, vol. I-6, pp. 238–240, December 1957.
[70] H. A. Dropkin, "Direct reading microwave phase-meter," *IRE Nat'l Conv. Rec.*, pt. 1, pp. 57–63, March 1958.
[71] C. A. Finnila, L. A. Roberts, and C. Süsskind, "Measurement of relative phase shift at microwave frequencies," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-8, pp. 143–147, March 1960.
[72] R. T. Stevens, "Precision phasemeter for CW or pulsed UHF," *Electronics*, vol. 33, pp. 54–57, March 4, 1960.
[73] B. P. Israelsen and R. W. Haegele, "Technique for the dynamic measurement of differential phase shift at microwave frequencies," *Proc. IRE* (*Correspondence*), vol. 50, pp. 474–475, April 1962.
[74] F. W. Holmes and R. H. Johnson, "A microwave phase-comparison system and its application to direction finding," *Proc. IEE* (*London*), vol. 109B, suppl. 23, pp. 670–677, May 1962.
[75] R. F. Koontz, "Microwave phase measurement of dispersed pulse transmitter systems," presented at the 1962 WESCON, Los Angeles, Calif.
[76] W. E. Blore, P. E. Robillard, and R. I. Primich, "35 and 70 Gc phase-locked CW balanced-bridge model measurement radars," *Microwave J.*, vol. 7, pp. 61–65, September 1964.
[77] W. P. Ernst, "A technique for measuring phase modulation or rapid phase changes of a microwave signal," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-13, pp. 70–76, January 1965.
[78] D. H. Russell, "Unmodulated twin-channel microwave attenuation measurement system," *J. ISA*, vol. 4, pp. 162–169, April 1965.

H. The Limits of Uncertainty

[79] A. J. Simmons, "Errors in a microwave rotary phase shifter," *Proc. IRE* (*Correspondence*), vol. 40, p. 869, July 1952.
[80] R. M. Vaillancourt, "Errors in a magic-tee phase changer," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-5, pp. 204–207, July 1957.
[81] J. Schachter, "Phase error of a two-phase resolver," *Proc. IRE* (*Correspondence*), vol. 45, p. 1018, July 1957.
[82] G. E. Schafer and R. W. Beatty, "Error analysis of a standard microwave phase shifter," *J. Res. NBS*, vol. 64C, pp. 261–265, October–December 1960.
[83] G. E. Schafer, "Mismatch errors in microwave phase shift measurements," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-8, pp. 617–622, November 1960.
[84] E. N. Phillips, "The uncertainties of phase measurement," *Microwaves*, vol. 4, pp. 14–21, February 1965.
[85] F. R. Voorhaar, "Phase linearity of microwave components," *Electronic Communicator*, vol. 1, p. 13, September–October 1966.
[86] I. M. Grankin and V. A. Ishchenko, "Errors in measuring SHF phase shifts using certain interference-type systems," *Izv. Vysshikh Uchebn. Zavedenii, Radiotekh.*, vol. 8, pp. 72–77, January–February 1965.
[87] W. L. Power, "Phase and gain characteristics of a 30 MHz IF amplifier," *Isascope* (*ISA News*, *Huntsville*, *Ala.*), vol. 3, p. 4, October 21, 1966.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967
pp. 960-969

Effects of Connectors and Adapters on Accurate Attenuation Measurements at Microwave Frequencies

ROBERT W. BEATTY, SENIOR MEMBER, IEEE

Abstract—A source of error in microwave attenuation measurements, not previously evaluated, is treated in an original analysis.

The error is caused by effects of deviations of connectors or adapters from standard specifications.

For example, if the coaxial connector at the insertion point in an attenuation measuring system deviates from standard specifications, the measured attenuation of a coaxial pad inserted in this system may differ from the measured attenuation of the same pad when inserted in a system having a standard connector at the insertion point.

In the analysis, an attenuator installed in a system is represented by three cascaded two-ports, the central one representing the core or kernel of the attenuator, and the others representing connector pairs. Three general cases are considered: the waveguide component under test has 1) either sexless or mating connectors, 2) nonmating connectors of the same type and sex, or 3) nonmating connectors of different types.

The results indicate that significant errors are possible with present connectors, such as the type *N*, but that these types of errors become negligible when high-precision connectors are used. The analysis permits one to tell how good a connector must be for a specific attenuator application.

Introduction

THE MAIN motivation for this paper was to answer the following question: (If the attenuation of a stable fixed attenuator is measured at the same operating frequency in two different systems, to what extent is the difference of results attributable to differences in the waveguide[1] joints or connectors[2] used at the insertion points?[3] (Effects of connector reflections and dissipative losses are taken into account, but leakage is not considered in this paper.)

In seeking an answer to this question, the representation ordinarily used for the insertion of an attenuator into a waveguide system, and the concept of insertion loss itself were found to be inadequate, and a new analysis was developed. The essential feature of the new approach is that the waveguide joints or connectors at insertion points are not assumed to be perfect,[2] i.e., having no loss, no reflection, and effectively no characteristic phase shift, but are represented by two-ports having appropriate characteristics. In addition, the attenuator when installed in a circuit is no longer repre-

Manuscript received July 21, 1964.

The author is with the National Bureau of Standards, Boulder, Colo.

[1] The term "waveguide" is used in this paper in a broad sense to include, for example, both uniconductor waveguide having a rectangular cross section and two-conductor waveguide having a concentric-circular or coaxial cross section.

[2] The term "connector" is used in this paper in a broad sense to designate the devices designed to join together two sections of waveguide having the same cross section. A "perfect connector pair" is one which would have no leakage, no loss, no reflection, and effectively-zero-characteristic phase shift.

[3] The term "insertion point" is used to designate the place where a waveguide component such as an attenuator is inserted into a waveguide system. It is thus not a geometrical point, but may be the region where a connector pair belonging to the system is disconnected, or where an adapter belonging to the system is removed, in order to insert a waveguide component.

Reprinted from IEEE TRANSACTIONS
ON *INSTRUMENTATION AND MEASUREMENT*
Volume IM-13, Number 4, December, 1964
pp. 272-284

sented by a single two-port, but by three cascaded two-ports, the outer ones representing the connector pairs, and the inner one representing the core or kernel of the attenuator. The quantity of interest is the loss in power delivered to the load when the above three cascaded two-ports (representing the attenuator and its associated connector pairs) are substituted for the single two-port representing the connector pair at the insertion point. This loss in power, expressed in decibels, is called the substitution loss [1].

Since the result of an attenuation measurement does depend upon the characteristics of the connectors at the insertion point, one must specify these characteristics if such a measurement is to be precisely defined. This leads to a slightly modified definition of the quantity of interest in the measurement of a quantity characteristic of the attenuator, and it is called the standard attenuation.

The analytical methods developed in this paper are applicable to other situations in which a waveguide component is inserted into a waveguide system. They might be applied for example to analyze the effects of connectors or adapters[4] on accurate measurements of the characteristic phase shift of phase shifters, or on the frequency of transmission-type cavity wavemeters. However, in this paper, they will be applied only to microwave attenuation measurements.

A brief review will be given of previous analytical methods used to obtain equations for insertion loss, attenuation, and mismatch error for the case of a single fixed attenuator. Then corresponding equations will be obtained using recently developed methods and the results will be discussed. The specific question mentioned earlier will be answered and a calculated limit given for the effect. The techniques developed will then be applied to the case of variable attenuators, and to fixed attenuators having nonmating[5] connectors. Some of the cases having immediate interest will be discussed in some detail and calculated examples given to illustrate the use of the error equations. Useful formulas and graphs supplemental to the analysis will be given in the Appendix.

Review of Previous Analyses

It has been customary [2], [3], [12] to represent a waveguide component such as an attenuator by a two-arm waveguide junction (two-port), as shown in Fig. 1, where ${}^{i}P$ and ${}^{f}P$ designate the powers dissipated in the load under the initial and final conditions, respectively. The reflection coefficients of the generator and the load

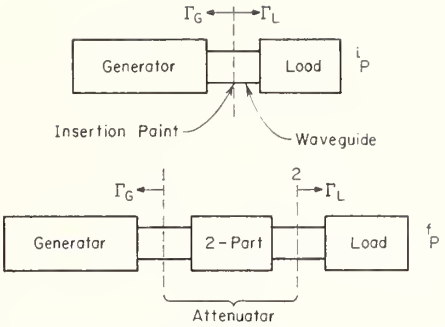


Fig. 1. Simple representation of an attenuator by a two-port inserted into a waveguide circuit.

are designated as Γ_G and Γ_L, respectively. It is usually assumed that they are the same at the times that ${}^{i}P$ and ${}^{f}P$ are observed. It is also apparent from the diagram that the connector pair or adapter at the insertion point is assumed to be perfect. Nothing is said about the properties of the connector pairs associated with the installed attenuator. It can either be assumed that they are perfect, or that the two-port labeled "attenuator" represents not only the attenuator itself, but also the associated connector pairs.

It has been customary to assume that one actually measures the insertion loss and that it can be written

$$L_I = 10 \log_{10} \frac{i_P}{f_P} = 10 \log_{10} \cdot \left[\frac{Z_{02}}{Z_{01}} \left| \frac{(1-S_{11}\Gamma_G)(1-S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L}{S_{21}(1-\Gamma_G\Gamma_L)} \right|^2 \right] \quad (1)$$

It has also been assumed that the desired quantity is the attenuation, which may be written

$$A = [L_I]_{\Gamma_G=\Gamma_L=0} = 10 \log_{10} \left[\frac{Z_{02}}{Z_{01}} \frac{1}{|S_{21}|^2} \right] \quad (2)$$

The difference between A and L_I is due mainly to system reflections or mismatch and has been called the mismatch error. It is written

$$\varepsilon_M = L_I - A = 20 \log_{10} \left| \frac{(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L}{1 - \Gamma_G\Gamma_L} \right| \quad (3)$$

In (1)–(3) the scattering coefficients of the two-port presumed to represent the attenuator are designated as S_{11}, S_{12}, S_{21}, and S_{22}, where an arbitrary choice has been made that the terminal surface toward the generator is port No. 1. The real characteristic impedances of the lossless waveguide leads are designated as Z_{01} and Z_{02}.

It is seen that the connector or adapter used at the insertion point is not shown in the diagram, and its characteristics do not appear in the equations. Hence there is no way to calculate its effect on the measurement.

[4] The term "adapter" is used to designate a device designed to join together two sections of waveguide which have already been fitted with connectors. A perfect adapter-connector combination would have no leakage, no loss, no reflection, and effectively-zero-characteristic phase shift.

[5] The term "nonmating connectors" as applied to a waveguide component such as an attenuator is meant to imply that the connectors on each end of the component are of such a type or sex that they could not be joined together even if it were possible to move them physically into a favorable position for such joining.

An Improved Representation

A two-port, or two-arm waveguide junction or transducer [4] has associated with it two waveguide leads through which energy may enter and leave. The terminal surfaces of the two-port are cross-sectional surfaces within the waveguide leads. Usually only one propagating mode is associated with each lead and each terminal surface. A waveguide component such as an attenuator evidently cannot itself be represented by a two-port unless the connectors are perfect and all connectors mate at coplanar butt joints. Since this requirement is often not very closely approximated by actual connectors, the representation of Fig. 2 is more realistic. The attenuator installed in the circuit is represented by a composite two-port composed of three cascaded two-ports, A, B, and C.

The central core or kernel of the attenuator is represented by B. The connectors B_G and B_L at each end of the attenuator mate with the connectors D_G and D_L, respectively, of the system to form connector pairs represented by two-ports A and C. When the attenuator is removed, the system may be closed as shown in Fig. 3 by joining connectors D_G and D_L.

The precise manner in which the connectors join and separate is not specified, as this would call for further more complicated analysis of specific discontinuities which is not felt to be justified at this time.

The analysis of the situation represented in Fig. 3 proceeds as follows.

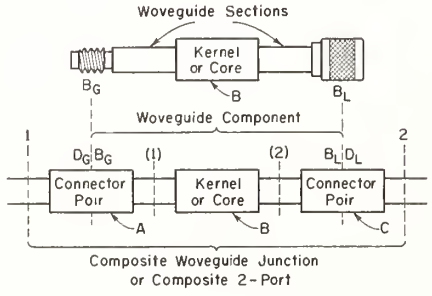


Fig. 2. Representation of a waveguide component.

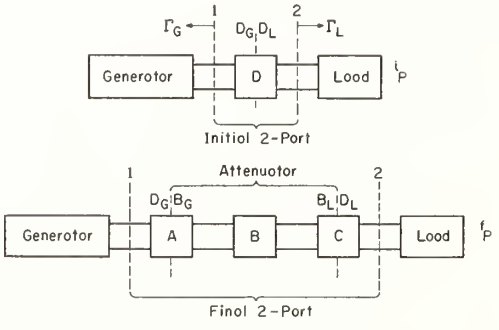


Fig. 3. Improved representation of attenuator insertion into a waveguide system (Case 1).

Substitution Loss

As before, it will be assumed that a certain quantity is desired, but actually one measures something different, and the difference is called the error.

One measures the ratio of ^iP to ^fP, the initial and final powers dissipated in the load, as shown in Fig. 3. It is similar to the ratio of (1), but the characteristics of the connectors are now implicitly involved, as they influence the scattering coefficients of the initial and final two-ports. The above ratio, expressed in decibels, is called the substitution loss L_S, and is written as follows:

$$L_S = 10 \log_{10} \frac{^iP}{^fP} = 20 \log_{10} \cdot \left| \frac{^iS_{21}}{^fS_{21}} \cdot \frac{(1 - {}^fS_{11}\Gamma_G)(1 - {}^fS_{22}\Gamma_L) - {}^fS_{12}{}^fS_{21}\Gamma_G\Gamma_L}{(1 - {}^iS_{11}\Gamma_G)(1 - {}^iS_{22}\Gamma_L) - {}^iS_{12}{}^iS_{21}\Gamma_G\Gamma_L} \right| \quad (4)$$

where the front superscripts i and f on the scattering coefficients refer to the initial and final two-ports, respectively.

Additional insight is obtained by writing (4) in two additional forms. First, it is written as the difference between the insertion losses of the initial and final two-ports as follows:

$$\begin{aligned} L_S &= {}^fL - {}^iL \\ &= 20 \log_{10} \left| \frac{(1 - {}^fS_{11}\Gamma_G)(1 - {}^fS_{22}\Gamma_L) - {}^fS_{12}{}^fS_{21}\Gamma_G\Gamma_L}{{}^fS_{21}(1 - \Gamma_G\Gamma_L)} \right| \\ &\quad - 20 \log_{10} \left| \frac{(1 - {}^iS_{11}\Gamma_G)(1 - {}^iS_{22}\Gamma_L) - {}^iS_{12}{}^iS_{21}\Gamma_G\Gamma_L}{{}^iS_{21}(1 - \Gamma_G\Gamma_L)} \right| \end{aligned} \quad (5)$$

It is apparent that, analytically, substitution loss is equivalent to a difference between two insertion losses. One could thus avoid the use of the concept of substitution loss, if desired. However the assumptions made in the definition of substitution loss are more easily realized in practice. In addition, it is more convenient to use when analyzing variable attenuators, as will be seen later.

Another form of (4) is the following

$$\begin{aligned} L_S &= ({}^fA - {}^iA) \\ &\quad + 20 \log_{10} \left| \frac{(1 - {}^fS_{11}\Gamma_G)(1 - {}^fS_{22}\Gamma_L) - {}^fS_{12}{}^fS_{21}\Gamma_G\Gamma_L}{1 - \Gamma_G\Gamma_L} \right| \\ &\quad - 20 \log_{10} \left| \frac{(1 - {}^iS_{11}\Gamma_G)(1 - {}^iS_{22}\Gamma_L) - {}^iS_{12}{}^iS_{21}\Gamma_G\Gamma_L}{1 - \Gamma_G\Gamma_L} \right| \end{aligned} \quad (6)$$

The first two terms above on the right are attenuations as defined in (2) and the last two terms are similar in form to (3), the mismatch error of the previous analysis.

More complete expressions, containing the scattering coefficients of two-ports A, B, C, and D could be obtained from (4), (5), and (6) by making the appropriate substitutions for the scattering coefficients of the composite final two-port. The equations given for cascade-

connected two-ports in the Appendix would be useful. Although it is clear how to do this, it will not be done as there is no specific use to be made of it in this paper.

The substitution loss which would occur in a non-reflecting system is of interest and can be deduced from (6) when $\Gamma_G = \Gamma_L = 0$. It is

$$(L_S)_{\Gamma_G=\Gamma_L=0} = ({}^fA - {}^iA) = 20 \log_{10} \left| \frac{{}^iS_{21}}{{}^fS_{21}} \right| \quad (7)$$

It is simply the difference in the attenuations between the initial and final two-ports.

Standard Attenuation

In practice, one is seldom interested in the absolute attenuation ${}^fA$ of the final two-port for two reasons. First, it is difficult to measure since it assumes that the initial two-port represents a perfect connector which cannot be actually realized. And second, it is not characteristic of the attenuator itself but of the attenuator kernel B plus two associated connector pairs A and C. The system connectors D_G and D_L which do not belong to the attenuator are parts of A and C.

The quantity of greatest interest is the standard attenuation which is above difference in attenuation when the initial two-port represents not a perfect connector pair, but a standard connector pair.[6] An expression for standard attenuation ${}^sA$ is obtained as follows: Equation (7) is written in terms of the scattering equations of two-ports A, B, C, and D as

$$\begin{aligned}(L_S)_{\Gamma_G=\Gamma_L=0} &= A_A + A_B + A_C - A_D \\ &+ 20 \log_{10} \left| \frac{(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}}{1 - a_{22}c_{11}} \right| \\ &- 20 \log_{10} \left| 1 - a_{22}c_{11} \right|\end{aligned}$$

or

$$\begin{aligned}(L_S)_{\Gamma_G=\Gamma_L=0} &= 20 \log_{10} \left| \frac{d_{21}}{a_{21}b_{21}c_{21}} \right. \\ &\left. \cdot [(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}] \right| \quad (8)\end{aligned}$$

If the two-port D represents a standard connector pair, and d_{21} is replaced by ${}^sS_{21}$, the standard attenuation is

$${}^sA = 20 \log_{10} \cdot \left| \frac{{}^sS_{21}}{a_{21}b_{21}c_{21}} [(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}] \right| \quad (9)$$

It is of interest to examine the form of (7) when the connector pairs are all identical and nonreflecting. One obtains

$$(L_S)_{\substack{\Gamma_G=\Gamma_L=0 \\ A \equiv C \equiv D}} = A_A + A_B \quad (10)$$

Under this assumption, (10) expresses a quantity characteristic of the attenuator itself as represented by the kernel B and connectors B_G and B_L.

To the extent that all connectors are identical and nonreflecting, (9) will also express a quantity characteristic of the attenuator itself. The standard attenuation as defined above is thus to a good approximation characteristic of the attenuator itself, and will be considered as the desired quantity in an attenuation measurement.

Connector and Mismatch Errors

The error ε_S in the measurement of standard attenuation cannot be obtained simply by subtracting (9) from (4), (5), or (6). This is evident when we consider Fig. 4 which represents the substitution loss of an attenuator measured in two different systems, M and N. Even though the attenuator is the same in both cases, it is associated with connector pairs A and C in system M and with P and Q in system N. Thus the final two-ports are different in the two systems.

In order to obtain an expression for ε_S, let system M be non-reflecting and two-port D represent a standard connector pair having an attenuation A_S. Then the error ε_S is the difference between the substitution losses in systems M and N, and is written

$$\varepsilon_S = \varepsilon_I + \varepsilon_{II} + \varepsilon_{III}$$

where

$$\varepsilon_I = (A_S - A_H) + (A_P - A_A) + (A_Q - A_C)$$

$$\begin{aligned}\varepsilon_{II} &= 20 \log_{10} \left| \frac{(1 - p_{22}b_{11})(1 - b_{22}q_{11}) - p_{22}b_{12}b_{21}q_{11}}{1 - p_{22}q_{11}} \right| \\ &- 20 \log_{10} \left| \frac{(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}}{1 - a_{22}c_{11}} \right| \\ &+ 20 \log_{10} \left| \frac{1 - p_{22}q_{11}}{1 - a_{22}c_{11}} \right|\end{aligned}$$

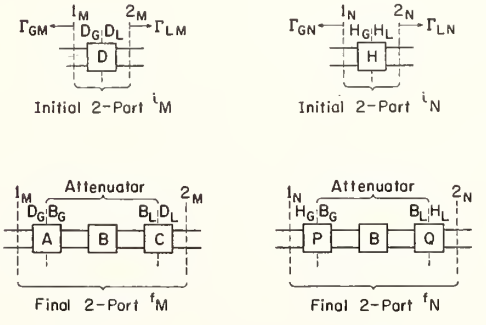


Fig. 4. Representation of same attenuator installed alternately into two different systems. (a) System "M." (b) System "N."

[6] A "standard connector" is one which is made precisely to standard specifications for the particular type of connector under consideration. Standard connector pairs usually have low but appreciable loss and reflection.

and

$$\varepsilon_{III} = 20 \log_{10} \left| \frac{(1-{}^{f}S_{11}\Gamma_G)(1-{}^{f}S_{22}\Gamma_L)-{}^{f}S_{12}{}^{f}S_{21}\Gamma_G\Gamma_L}{1-\Gamma_G\Gamma_L} \right|$$

$$- 20 \log_{10} \left| \frac{(1-h_{11}\Gamma_G)(1-h_{22}\Gamma_L)-h_{12}h_{21}\Gamma_G\Gamma_L}{1-\Gamma_G\Gamma_L} \right| \quad (11)$$

The error component ε_I will vanish if corresponding connectors at the insertion points are identical,[7] since it is seen from Fig. 4 that the resulting condition will be that $D \equiv H$, $A \equiv P$, and $C \equiv Q$. This is a sufficient, but not a necessary, condition since ε_I will vanish for any condition for which $A_S+A_P+A_Q=A_H+A_A+A_C$. Usually the corresponding connectors at the insertion points will be similar, and it is evidently worthwhile to make them as nearly identical as possible.

The error component ε_{II} will also vanish if corresponding connectors at the insertion points in the two systems are identical, and this is again a sufficient but not a necessary condition. It will also approach zero if the attenuator kernel is nonreflecting and its attenuation becomes arbitrarily large. The individual terms are of the same familiar form as (3) for mismatch error in the simpler analysis.

The error component ε_{III} is very similar in form to (3) and will vanish if the system reflection coefficients Γ_G and Γ_L vanish. Again this is a sufficient but not a necessary condition, since the relative phases of the reflection coefficients involved might possibly be such as to make the two terms vanish or cancel each other. It is clear that ε_{III} differs from the other two components in that the condition of identical corresponding connectors at the insertion points does not make it vanish. The limits of terms in ε_{II} and ε_{III} similar to (3) may be evaluated by familiar methods [2], [3], and they are typically each of several tenths of a decibel, or less. Since each such positive term is paired with a similar negative term, it is the differences which are important in these errors.

Since the various terms in (11) may not all be of the same sign, some of them may tend to cancel others, and the overall error ε_S may be lower than some of the individual terms. In careful measurements however, one cannot afford to take this for granted, but should either make a thorough investigation, or quote a conservative limit of error, assuming the most unfavorable phase relationships of the coefficients involved.

Same Fixed Attenuator in Two Systems

The motivating question for the paper can now be answered using the representation of Fig. 4, and letting two-port D represent a connector pair which is not necessarily a standard one, and letting Γ_{GM} and Γ_{LM} be representative of an actual system, and not an idealized one having no reflections.

Assume that in attempting to measure the standard attenuation of a fixed attenuator in two different systems, we actually measure the substitution loss. We are interested in the difference ΔL_S in substitution loss as measured in the two systems. As in the previous situation, a part of this difference is primarily due to differences in connectors D_G and H_G, and D_L and H_L, and the other part of the difference is due primarily to differences in system generator and load reflection coefficients.

Referring to Fig. 4 and (6), we can write

$$\begin{aligned}
\Delta L_S = L_{SN} - L_{SM} = ({}^{f}N A - {}^{f}M A) + ({}^{i}M A - {}^{i}N A) \\
+ 20 \log_{10} \left| \frac{(1 - {}^{f}S_{11}\Gamma_{GN})(1 - {}^{f}S_{22}\Gamma_{LN}) - {}^{f}S_{12}{}^{f}S_{21}\Gamma_{GN}\Gamma_{LN}}{1 - \Gamma_{GN}\Gamma_{LN}} \right| \\
- 20 \log_{10} \left| \frac{(1 - {}^{i}S_{11}\Gamma_{GN})(1 - {}^{i}S_{22}\Gamma_{LN}) - {}^{i}S_{12}{}^{i}S_{21}\Gamma_{GN}\Gamma_{LN}}{1 - \Gamma_{GN}\Gamma_{LN}} \right| \\
- 20 \log_{10} \left| \frac{(1 - {}^{f}S_{11}\Gamma_{GM})(1 - {}^{f}S_{22}\Gamma_{LM}) - {}^{f}S_{12}{}^{f}S_{21}\Gamma_{GM}\Gamma_{LN}}{1 - \Gamma_{GM}\Gamma_{LM}} \right| \\
+ 20 \log_{10} \left| \frac{(1 - {}^{i}S_{11}\Gamma_{GM})(1 - {}^{i}S_{22}\Gamma_{LM}) - {}^{i}S_{12}{}^{i}S_{21}\Gamma_{GM}\Gamma_{LM}}{1 - \Gamma_{GM}\Gamma_{LM}} \right| \quad (12)
\end{aligned}$$

The sufficient conditions under which ΔL_S will vanish are ${}^{i}M = {}^{i}N$, ${}^{f}M = {}^{f}N$, $\Gamma_{GM} = \Gamma_{GN}$, and $\Gamma_{LM} = \Gamma_{LN}$. In practice, one might try to achieve these conditions, but some uncertainty will always exist. In order to evaluate ΔL_S using (12), information would be needed on the characteristics of the two-ports ${}^{i}M$, ${}^{f}M$, ${}^{i}N$, and ${}^{f}N$, as well as on the system generator and load reflection coefficients Γ_{GM}, Γ_{LM}, Γ_{GN}, and Γ_{LN} effective at terminal surfaces 1_M, 2_M, 1_N, and 2_N.

It is possible to reduce the magnitudes of the system reflection coefficients to very low values (say 0.001) by

[7] A subtle point arises in connection with the choice of terminal surfaces 1 and 2 on either side of the connector pair at the insertion point. The observed change in loss in a given situation is obviously invariant to how we represent it, or where we choose our terminal surfaces. However, it is possible to choose them in some cases so that the connector pairs they define in the two systems are identical for one choice but not identical for another. This would be true for example if the connectors at the insertion point in one system were identical to those in the other system with the exception that the dielectric beads were of different design and material. In one case we could choose terminal surfaces to include the beads with the connector pairs, and the two pairs would be different. In the other case we could choose terminal surfaces to include the beads with the systems, not the connectors. The connector pairs would then be identical. Since we know that it does not matter where we choose our terminal surfaces (as long as they are in a region of single-mode propagation) we shall choose them in the most convenient and accessible places.

the user of tuners, in which case the last four terms of (12) would be negligible, e.g., less than 0.001 dB. Thus the case of ΔL_S for $\Gamma_{GM}=\Gamma_{LM}=\Gamma_{GN}=\Gamma_{LN}=0$ is of interest and is given by

$$(\Delta L_S)_{\substack{[\Gamma_{GM}=\Gamma_{LM}=0]\\[\Gamma_{GN}=\Gamma_{LN}=0]}} = \Delta A = (A_D - A_H) + (A_P - A_A) + (A_Q - A_C) + 20\log_{10}\left|\frac{(1-p_{22}b_{11})(1-b_{22}q_{11})-p_{22}b_{12}b_{21}q_{11}}{(1-a_{22}b_{11})(1-b_{22}c_{11})-a_{22}b_{12}b_{21}c_{11}}\right| \quad (13)$$

The above result equals $\mathcal{E}_I+\mathcal{E}_{II}$ of (11) if we replace A_D by A_S, and vanishes if connectors $D_G \equiv H_G$ and $D_L \equiv H_L$.

The use of (13) in the evaluation of errors due to differences in the system connectors is illustrated by the following calculated examples. Consider the case in which all connector pairs in system M are identical, or $A \equiv C \equiv D$, but the left-hand connector D_G at the insertion point is modified and becomes H_G, but D_L and H_L remain the same. Under this supposition, connector pairs H and P will be the same, and connector pairs A, C, D, and Q will be alike.

If connector D_G were a male type-N connector, for example, it might be modified by removing the compensating step in the outer conductor and moving the step in the male center conductor outward so as to close the gap or notch normally present. In the following examples certain values are assumed for the reflection coefficients appearing in (13) which are thought to be realistic in view of the measured results obtained at NBS for certain type-N connectors shown in Fig. 5. Since large variations among connectors of the same type are possible, these results cannot be regarded as typical of all type-N connectors.

Example 1

Given a 3 dB fixed attenuator having a kernel VSWR of 1.22, connector pairs A, C, D, and Q having VSWR's of 1.22, and connector pairs H and P having VSWR's of 1.00. Thus $|p_{11}|=|p_{22}|=|h_{11}|=|h_{22}|=0$, $|b_{11}|=|a_{22}|=|c_{11}|=|q_{11}|=0.10$, and $|b_{12}b_{21}|=0.5$.

The attenuation terms in (13) have components [1], [5] due to dissipation and reflection, and those due to reflection cancel in the above example. The dissipative components should also nearly cancel, leaving only the final term of (13). It can be written to a good approximation as

$$20\log_{10}|1+b_{11}(a_{22}-p_{22})+b_{22}(c_{11}-q_{11})+b_{12}b_{21}(a_{22}c_{11}-p_{22}q_{11})| \quad (14)$$

if the connector reflections are small, corresponding to reflection coefficients of magnitude less than 0.11. Limits of ΔA, assuming the worst phase combinations are

$$-0.131\ dB \leq \Delta A \leq 0.129\ dB$$

The limits of error in the above example are significant ones, well above the usual precision of a good attenuation measurement. However, many attenuators which are presently commercially available have better characteristics than were assumed above. Hence, another example follows.

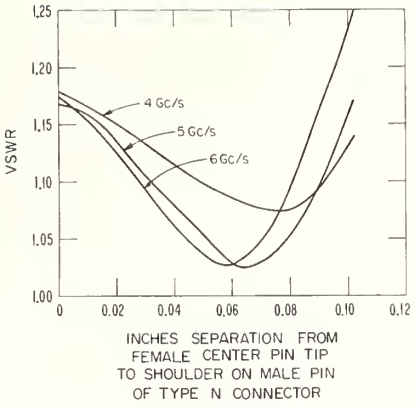


Fig. 5. Experimental data for VSWR of type-M connector pair.

Example 2

Given a 3 dB fixed attenuator having a kernel VSWR of 1.15, connector pairs A, C, D, and Q having VSWR's of 1.15, and connector pairs H and P having VSWR's of 1.00. Thus $|p_{11}|=|p_{22}|=|h_{11}|=|h_{22}|=0$, $|b_{11}|=|a_{22}|=|c_{11}|=|q_{11}|=0.07$, and $|b_{12}b_{21}|=0.5$.

Proceeding as in Example 1, the limits of ΔA are

$$-0.064\ \text{dB} \leqq \Delta A \leqq 0.064\ \text{dB}$$

The limits of error in this example are smaller, but still significant, since a precision of 0.03 dB is often obtained [6] in a measurement of 3 dB at frequencies of 4 GHz and above.

If the fixed attenuators in Examples 1 and 2 above have attenuations of 20 dB or more (instead of 3 dB), the calculated limits of ΔA will be reduced to approximately ± 0.09 dB and ± 0.04 dB, respectively.

The above examples illustrate the importance of using identical connectors at the insertion points of attenuation measurement systems, especially when comparison measurements are to be made on the same attenuator by each of the systems.

Adjusting Systems for Zero Reflections

In adjusting an attenuation measurement system in an effort to make $\Gamma_G=\Gamma_L=0$, the arrangement represented by Fig. 6 is often used. A slotted line is connected in turn to system connectors D_G and D_L, and tuners are adjusted until the reflection coefficients Γ_G' and Γ'_L observed in the slotted section effectively vanish. This does not necessarily make the reflection coefficients Γ_G and Γ_L of the system vanish because, in general, the two-ports U and V may have reflections. The two-ports U and V are composite two-ports representing a connector pair and the taper or transition section at the end of the slotted line. In addition, a tuning stub may be included, if this is used to adjust [7] for the condition $u_{22}=v_{11}=0$. This condition, together with the condition $\Gamma_G'=\Gamma_L'$

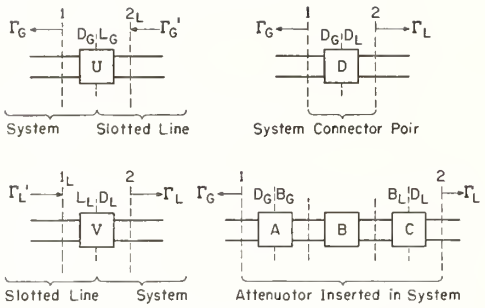


Fig. 6. Representation of arrangement often used to adjust system reflections prior to attenuation measurement.

$=0$, will make the reflection coefficients Γ_G and Γ_L of the measurement system vanish. The measured substitution loss will then equal the standard attenuation, provided that connector pair D is standard.

It is worthwhile to tune for the condition $u_{22}=v_{11}=0$ (tuning out of the residual VSWR) because one can then adjust for system VSWR's of 1.01 or less instead of having to "add" to this the residual VSWR of 1.04 to 1.10 that may be present.

If the condition $u_{22}=v_{11}=0$ is not obtained and Γ_G' and Γ_L' are not made to vanish, the measured substitution loss then is given by (4) where

$$L_G = \frac{\Gamma_G' - u_{22}}{(u_{12}u_{21} - u_{11}u_{22}) + u_{11}\Gamma_G'}$$

$$\Gamma_L = \frac{\Gamma_L' - v_{11}}{(v_{12}v_{21} - v_{11}v_{22}) + v_{22}\Gamma_L'}$$

$$^iS_{11} = d_{11},\ ^iS_{12} = d_{12},\ ^iS_{21} = d_{21},\ ^iS_{22} = d_{22}$$

$$S_{11} = a_{11} + a_{12}a_{21}\frac{b_{11} + (b_{12}b_{21} - b_{11}b_{22})c_{11}}{(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}}$$

$$^fS_{12} = \frac{a_{12}b_{12}c_{12}}{(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}}$$

$$^fS_{21} = \frac{a_{21}b_{21}c_{21}}{(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}}$$

and

$$^fS_{22} = c_{22} + c_{12}c_{21}\frac{b_{22} + (b_{12}b_{21} - b_{11}b_{22})a_{22}}{(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}} \quad (15)$$

Same Variable Attenuator in Two Systems

Both continuously variable and step attenuators can be analyzed by the same method, which is an extension of the previous analysis. A continuously variable attenuator can be regarded as though one removed an initial attenuator corresponding to the initial setting, and substituted in its place a final attenuator corresponding to the final setting. This point of view is valid, even though the continuously variable attenuator remains in the circuit at all times.

A representation which applies to the measurement of a given change in a variable attenuator first in one system and then in a different system is shown in Fig. 7. The system connectors D_G and D_L, H_G and H_L are not shown as joined together because in some cases it is not possible to do so. For example, if the step attenuators have female connectors on both ends, the system connectors will all be male and will not mate together.

Although the initial attenuator, represented by J_G-J-J_L and the final attenuator, represented by B_G-B-B_L are the same in the two systems, the initial and final composite two-ports are not.

The difference in the substitution loss measured in the two systems M and N is written down directly from inspection of Fig. 7 and a knowledge of (4).

$$\begin{aligned} L_{SN} - L_{SM} = {} & (^fA_N - {}^fA_M) + (^iA_M - {}^iA_N) \\ & + 10\log_{10}\left|\frac{(1 - {}^im_{11}\Gamma_{GM})(1 - {}^im_{22}\Gamma_{LM}) - {}^im_{12}{}^im_{21}\Gamma_{GM}\Gamma_{LM}}{1 - \Gamma_{GM}\Gamma_{LM}}\right| \\ & - 20\log_{10}\left|\frac{(1 - {}^fm_{11}\Gamma_{GM})(1 - {}^fm_{22}\Gamma_{LM}) - {}^fm_{12}{}^fm_{21}\Gamma_{GM}\Gamma_{LM}}{1 - \Gamma_{GM}\Gamma_{LM}}\right| \\ & + 20\log_{10}\left|\frac{(1 - {}^fn_{11}\Gamma_{GN})(1 - {}^fn_{22}\Gamma_{LN}) - {}^fn_{12}{}^fn_{21}\Gamma_{GN}\Gamma_{LN}}{1 - \Gamma_{GN}\Gamma_{LN}}\right| \\ & - 20\log_{10}\left|\frac{(1 - {}^in_{11}\Gamma_{GN})(1 - {}^in_{22}\Gamma_{LN}) - {}^in_{12}{}^in_{21}\Gamma_{GN}\Gamma_{LN}}{1 - \Gamma_{GN}\Gamma_{LN}}\right| \end{aligned} \quad (16)$$

In order for (16) to vanish we not only would need identical connectors at the insertion points ($D_G=H_G$ and $D_L=H_L$) but corresponding system reflection coefficients would need to be equal ($\Gamma_{GM}=\Gamma_{GN}$ and $\Gamma_{LM}=\Gamma_{LN}$). Even if these (sufficient) conditions are not obtained, it is possible for (16) to vanish under other less easily described conditions, although it is not very probable. In general, there will be a difference in the substitution losses measured in the two systems.

If both systems are nonreflecting, (16) reduces to

$$(L_{SN} - L_{SM})\left[\begin{smallmatrix}\Gamma_{GM}=\Gamma_{LM}=0\\ \Gamma_{GN}=\Gamma_{LN}=0\end{smallmatrix}\right] = (^fA_N - {}^fA_M) + (^iA_M - {}^iA_N)$$

$$= 20\log_{10}\left|\frac{{}^fm_{21}}{{}^im_{11}}\cdot\frac{{}^in_{21}}{{}^fn_{21}}\right| \quad (17)$$

The above equation is written in terms of the scattering coefficients of the individual two-ports which make up the initial and final composite two-ports as follows:

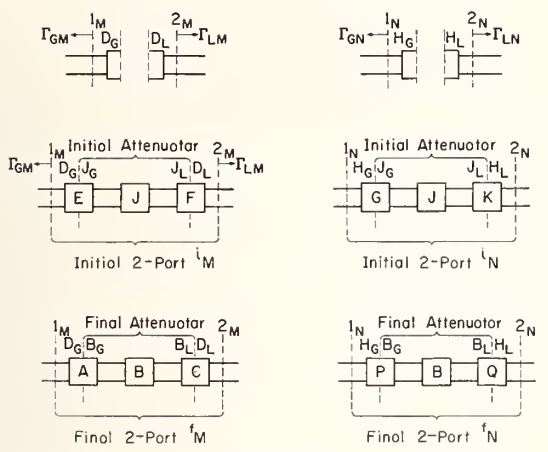


Fig. 7. Representation of substitution of same final and initial attenuators in two different systems.

$$(L_{SN} - L_{SM}) \left[\begin{matrix}\Gamma_{GM}=\Gamma_{LM}=0\\ \Gamma_{GN}=\Gamma_{LN}=0\end{matrix}\right]$$

$$= (A_A - A_P) + (A_C - A_Q) - (A_E - A_G) - (A_F - A_K)$$

$$+ 20\log_{10}\left|\frac{(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}}{1 - a_{22}c_{11}}\right|$$

$$- 20\log_{10}\left|\frac{(1 - e_{22}j_{11})(1 - j_{22}f_{11}) - e_{22}j_{12}j_{21}f_{11}}{1 - e_{22}f_{11}}\right|$$

$$+ 20\log_{10}\left|\frac{(1 - g_{22}j_{11})(1 - j_{22}k_{11}) - g_{22}j_{12}j_{21}k_{11}}{1 - g_{22}k_{11}}\right|$$

$$- 20\log_{10}\left|\frac{(1 - p_{22}b_{11})(1 - b_{22}q_{11}) - p_{22}b_{12}b_{21}q_{11}}{1 - p_{22}q_{11}}\right|$$

$$+ 20\log_{10}\left|\frac{(1 - a_{22}c_{11})(1 - g_{22}k_{11})}{(1 - e_{22}f_{11})(1 - p_{22}q_{11})}\right| \qquad (18)$$

If corresponding connectors at the insertion points in the two systems are identical ($D_G \equiv H_G$ and $D_L \equiv H_L$) (18) will vanish. Even if this condition does not hold, it could vanish if the positive and negative terms canceled each other.

It is interesting to consider the case in which corresponding connectors at the insertion points in the two systems are not identical, but corresponding connectors on the two attenuators are. ($J_G = B_G$ and $J_L = B_L$.) In this case we have $A \equiv E$, $F \equiv C$, $G \equiv P$, and $K \equiv Q$. Many terms cancel and (18) reduces to

$$(L_{SN} - L_{SM}) \left[\begin{matrix}\Gamma_{GM}=\Gamma_{LM}=0\\ \Gamma_{GN}=\Gamma_{LN}=0\\ J_G=B_G\\ J_L=B_L\end{matrix}\right]$$

$$= 20\log_{10}\left|\frac{(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}}{1 - a_{22}c_{11}}\right|$$

$$- 20\log_{10}\left|\frac{(1 - a_{22}j_{11})(1 - j_{22}c_{11}) - a_{22}j_{12}j_{21}c_{11}}{1 - a_{22}c_{11}}\right|$$

$$+ 20\log_{10}\left|\frac{(1 - g_{22}j_{11})(1 - j_{22}k_{11}) - g_{22}j_{12}j_{21}k_{11}}{1 - g_{22}k_{11}}\right|$$

$$- 20\log_{10}\left|\frac{(1 - g_{22}b_{11})(1 - b_{22}k_{11}) - g_{22}b_{12}b_{21}k_{11}}{1 - g_{22}k_{11}}\right| \qquad (19)$$

The conditions $J_G = B_G$ and $J_L = B_L$ would apply in the case of a continuously variable attenuator that was not physically removed from the circuit. They would also apply to step attenuators provided their connectors were sufficiently identical.

The additional conditions $J_G \equiv J_L \equiv B_G \equiv B_L$, $D_G \equiv D_L$, and $H_G \equiv H_L$ would not reduce the size of (19), but would reduce the uncertainty in the measurements of the two systems.

It is clear from the foregoing that in order to insure the same change in loss in two systems from the same change in settings of a variable attenuator, 1) the corresponding system reflection coefficients must be the same and 2) the corresponding system connectors at the insertion points must be the same. Even when these conditions are not realized, some reduction in the difference is obtainable if all other connectors are as nearly alike as possible. It is also clear that reduction of connector reflections and dissipative losses will also reduce the difference. A fortuitous relationship of the phases of the reflection coefficients may also make the difference vanish, although the probability of this happening is low.

Standard Incremental Attenuation

The desired quantity in a measurement of a variable attenuator is the change in attenation (incremental attenuation [8]) from an initial setting to a final setting. This equals the substitution loss when the system is nonreflecting and is given by (7). However it must be recognized that the characteristics of the connectors D_G and D_L at the insertion point are implicitly involved, even if only to small degree. This is evident from inspection of Fig. 7, where it is seen that connectors D_G and D_L are involved in a slightly different way in the initial and final two-ports.

In order to define a precisely-repeatable incremental attenuation, the connectors at the insertion point should always be the same, and should be standardized. With standard connectors at the insertion point, we then can measure "standard incremental attenuation."

This quantity may be expressed by reference to system M in Fig. 7 as

$$\Delta^s A = 20\log_{10}\left|\frac{{}^{i}m_{21}}{{}^{f}m_{21}}\right|$$

or

$$\Delta^s A = A_A + A_B + A_C - A_E - A_J - A_F$$

$$+ 20\log_{10}\left|(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}\right|$$

$$- 20\log_{10}\left|(1 - e_{22}j_{11})(1 - j_{22}f_{11}) - e_{22}j_{12}j_{21}f_{11}\right| \qquad (20)$$

where it is understood that D_G and D_L, which form parts of two-ports A, E, C, and F, are standard connectors.

The error in measuring standard incremental attenuation (due to system reflections and differences from standard conditions in the actual system) can be evaluated by comparing (20) with (16), letting M represent the idealized system with standard conditions (Γ_{GM}

385-582 O - 70 - 17

$= \Gamma_{LM} = 0$, D_G and D_L standard connectors) and letting N represent the actual system. The procedure in writing equations and calculating examples follows along lines of the analysis pertaining to fixed attenuators. The error is written

$$\mathcal{E}_{SI} = L_{SN} - \Delta^s A = 20 \log_{10} \left| \frac{{}^f m_{21}}{{}^i m_{21}} \cdot \frac{{}^i n_{21}}{{}^f n_{21}} \right|$$

$$+ 20 \log_{10} \left| \frac{(1 - {}^f n_{11}\Gamma_{GN})(1 - {}^f n_{22}\Gamma_{LN}) - {}^f n_{12} {}^f n_{12} \Gamma_{GN}\Gamma_{LN}}{(1 - {}^i n_{11}\Gamma_{GN})(1 - {}^i n_{22}\Gamma_{LN}) - {}^i n_{12} {}^i n_{21} \Gamma_{GN}\Gamma_{LN}} \right| \quad (21)$$

where the m's and n's are the scattering coefficients of the composite two-ports in Fig. 7 and it is understood that system M represents the idealized, and system N the actual, system. The first term in (21) vanishes if connectors H_G and H_L are standard, for then ${}^i M \equiv {}^i N$, and ${}^f M = {}^f N$. The second term vanishes if the reflection coefficients Γ_{GN} and Γ_{LN} of the actual system vanish.

In order to calculate a simple example, let the second term above vanish and assume that only the connector H_G is nonstandard. Then two-ports A, C, E, F, K, and Q are alike and represent standard connector pairs and two-ports G and P are alike and represent nonstandard connector pairs. The first term in (21) can be written

$$20 \log_{10} \left| \frac{{}^f m_{21}}{{}^i m_{21}} \cdot \frac{{}^i n_{21}}{{}^f n_{21}} \right|$$

$$= A_P + A_Q - A_G - A_K + A_E + A_F - A_A - A_C$$

$$+ 20 \log_{10} \left| \frac{(1 - e_{22}j_{11})(1 - j_{22}f_{11}) - e_{22}j_{12}j_{21}f_{11}}{(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}} \right|$$

$$+ 20 \log_{10} \left| \frac{(1 - p_{22}b_{11})(1 - b_{22}q_{11}) - p_{22}b_{12}b_{21}q_{11}}{(1 - g_{22}j_{11})(1 - j_{22}k_{11}) - g_{22}j_{12}j_{21}k_{11}} \right| \quad (22)$$

Applying the above assumptions, the right side of (22) reduces to just its last two terms. Assuming that reflections are small and neglecting the smaller terms, as in obtaining (14), the last two terms of (22) become

$$20 \log_{10} \left| 1 + j_{11}(g_{22} - e_{22}) + j_{22}(k_{11} - f_{11}) + j_{12}j_{21}(g_{22}k_{11} - e_{22}f_{11}) \right|$$

$$+ 20 \log_{10} \left| 1 + b_{11}(a_{22} - p_{22}) + b_{22}(c_{11} - q_{11}) + b_{12}b_{21}(a_{22}c_{11} - p_{22}q_{11}) \right| \quad (23)$$

Example 3

Suppose that the initial attenuator is a 3-dB pad having a kernel VSWR = 1.222, the final attenuator is a 10 dB pad having a kernel VSWR of 1.15, the standard connector pair VSWR's are 1.15, and the other connector pair VSWR's are 1.00. Thus $|j_{12}j_{21}| = 0.5$, $|j_{11}| = |j_{22}| = 0.1$, $|b_{12}b_{21}| = 0.1$, $|b_{11}| = |b_{22}| = 0.07$, $|a_{22}| = |c_{11}| = |e_{22}| = |f_{11}| = |k_{11}| = |q_{11}| = 0.07$, and $|g_{22}| = |p_{22}| = 0$. The resulting error limits are

$$-0.13 \text{ dB} \leqq \mathcal{E}_{SI} \leqq 0.13 \text{ dB}$$

The error limits in this example are appreciable and it is possible to make such an error in actual measurements. However, it is easy to reduce this error by careful control of the connectors used in the measuring systems.

Basic Insertion Arrangements

The representation schemes and analyses presented can be extended to other insertion arrangements for waveguide components such as attenuators.

The basic insertion arragnements considered in this paper are classified into three groups as follows, depending upon whether the waveguide component has 1) sexless or mating connectors, 2) nonmating connectors of the same type, and 3) nonmating connectors of different types.

An example of the first case was shown in Fig. 2, and examples of the other cases are shown in Figs. 8 to 10. In the last two cases, adapters are employed. Adapters for Case 2 are simpler than for Case 3, since a transition between waveguide of different cross sections is not employed.

Case 1 has already been considered, and some of the analysis applying to Fig. 7 will find applications in Cases 2 and 3. Insertion arrangements other than those shown are possible and may also be of some interest, but will not be specifically considered in this paper. For example, the case of two or more cascade-connected variable attenuators is of interest and has been analyzed [9] by simpler techniques, but will not be presently considered.

Cases 2A and 3A—Combining the Component with an Adapter

In the arrangement of Fig. 8, an adapter is connected to the waveguide component having nonmating connectors to form a composite waveguide component having mating connectors. The insertion arrangement for the composite component is then the same as for Case 1, and the previous analysis applies.

However, if we are primarily interested in the attenuation of the waveguide component itself without the adapter attached to it, this insertion arrangement will not give a direct answer.

Neglecting reflections, it is seen that the attenuation of the cascaded two-ports A, B, C, and P would be obtained, provided that $D \equiv Q$. The attenuation of A and B together would be nearly characteristic of the waveguide component itself, so that the attenuation of C and P together would need to be determined and subtracted. If reflections were taken into account, it would be even more difficult to obtain the desired characteristic loss from the measured substitution loss.

It is concluded that unless the adapter is to be permanently attached to the waveguide component, this insertion arrangement does not directly yield the desired information.

Cases 2B and 3B—Substituting the Component for an Adapter

The same basic representation shown in Fig. 9 applies analytically to either Case 2B or 3B as one can see from the examples given. The example of Case 2B is especially familiar as the drum or turret-type step attenuator. The attenuation steps are usually referred to a "zero dB attenuator" which is an adapter designed to have nominally no loss. This example has already been analyzed in the section on the same variable attenuator in two systems as represented by Fig. 7. It is apparent that calibrations of such an attenuator in different systems may not agree if the connectors at the insertion points are different. To avoid such a possibility, it is advisable to make certain that they are identical. This requires standardizing the design of various types of connectors and then adhering to that standard in their construction. (In addition, tolerances of construction should be extremely small.)

If one is interested in the change of attenuation of such an attenuator relative to an adapter, then the design of the adapter should also be standardized and the adapters should be constructed according to this standard design.

If one is interested in the attenuation characteristic of the attenuator itself, this insertion arrangement will not give direct information. Instead, additional calculations would be necessary after having first determined the characteristics of the adapter and connector pairs used. It will be found that Case 2C is a better arrangement for the above purpose.

Case 3B is similar to Case 2B except that the waveguide component such as an attenuator has nonmating connectors of different types, and the adapter for which it is substituted must also have corresponding types of connectors. This case occurs in practice for example if one desires to measure the coupling of a directional coupler by measuring the attenuation between two arms, the other arms terminated. The side arm in some cases may have a different waveguide than the main arm so that the connectors are of different types.

In using the substitution arrangement of Fig. 9, one can determine the attenuation relative to a given adapter, which must be standardized if the measurement is to be repeatable and significant. No direct information is obtained concerning the waveguide component such as the directional coupler itself, and additional calculations would be necessary, given the characteristics of the adapter. It will be found that the arrangement of Case 3C gives better direction information, but still may not be completely satisfactory.

Cases 2C and 3C—Combining the Component with an Adapter and Substituting for Another Adapter

These cases are of interest because the quantity directly measured is to a good approximation characteristic of the waveguide component alone, and no additional calculations are required to take into account the adapter. This is true at least for Case 2C, if not for 3C, for which additional measurements are required.

In the examples of Case 2C, it is seen that if the kernels J and P of the two adapters are the same, and all of the connector pairs are standard, the measured substitution loss to a good approximation can equal the attenuation of the kernel B plus one connector pair. This is characteristic of the waveguide component which consists of the kernel B plus two connectors. The argument is similar to that following (10), and is subject to the additional assumption here that the connector loss splits equally between the male and female connectors. A detailed analysis will not be given, but would follow along lines of those already presented.

In the examples of Case 3C, shown in Fig. 10, it is seen that the measured quantity is not likely to be a good approximation to a characteristic of the waveguide component itself as represented by the kernel B and connectors B_G and B_L. We would have to assume not only that connector pairs $E \equiv A$ and $F \equiv Q$ (which is quite reasonable), but also that adapter kernels $J \equiv P$, and that the losses in connector pair C equal those in connectors B_G and B_L. The latter two assumptions could be quite unrealistic and not correct to a good approximation.

A combination of Case 3A and Case 3B substitution measurements as shown in Fig. 11 could be used to obtain more or less directly a quantity characteristic of the waveguide component itself. It is evident that waveguide components such as are shown in these examples are troublesome, and require extra effort in their evaluation. In case that the waveguide component under consideration is itself an adapter, this technique is of particular interest and deserves further study. A detailed analysis is outside the scope of this paper, but would follow along the lines already presented.

Conclusions

A more rigorous representation and analysis has been presented to enable calculation of the effects of connectors and adapters on accurate attenuation measurements. The measured substitution loss replaces the insertion loss, and the former mismatch error is replaced by an error having three components. One condition under which the error vanishes is that the system is nonreflecting and has standard connectors at the insertion point.

A method of obtaining the nonreflecting condition using a tuning stub and slotted line is discussed.

The need for use of standard connectors is emphasized by some calculated examples in which error limits up to 0.13 decibel are obtained when such a connector is nonstandard. Examples are calculated for both fixed and variable attenuators and are based upon measured data on type N connectors.

Situations in which adapters are used in different insertion arrangements are discussed, and it is concluded

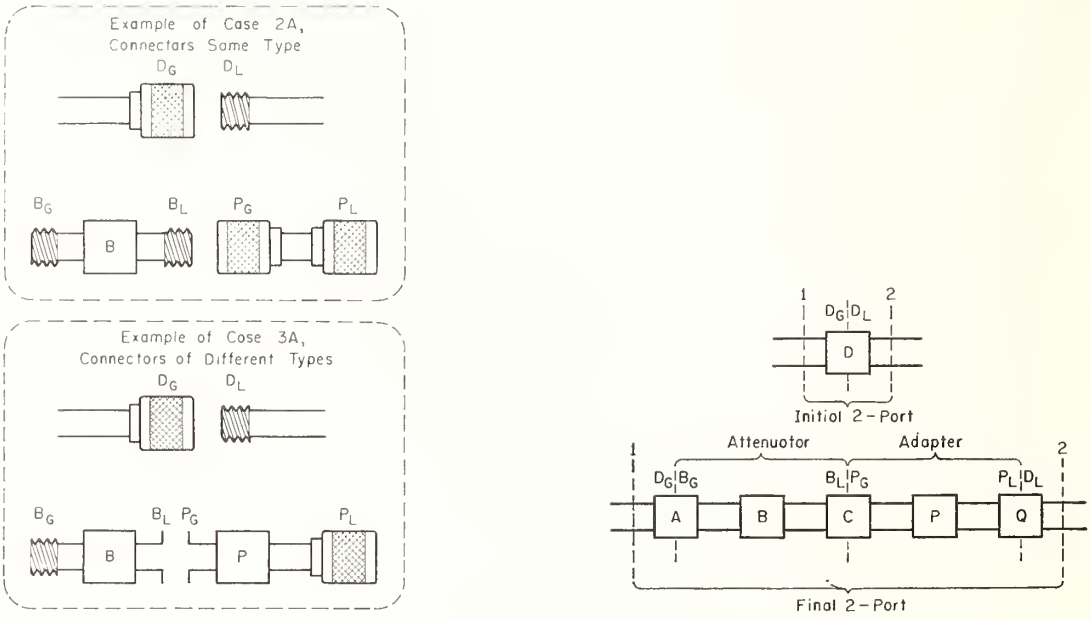


Fig. 8. Representation of insertion of a waveguide component (such as an attenuator) having nonmating connectors by connecting it to an adapter (Cases 2A and 3A).

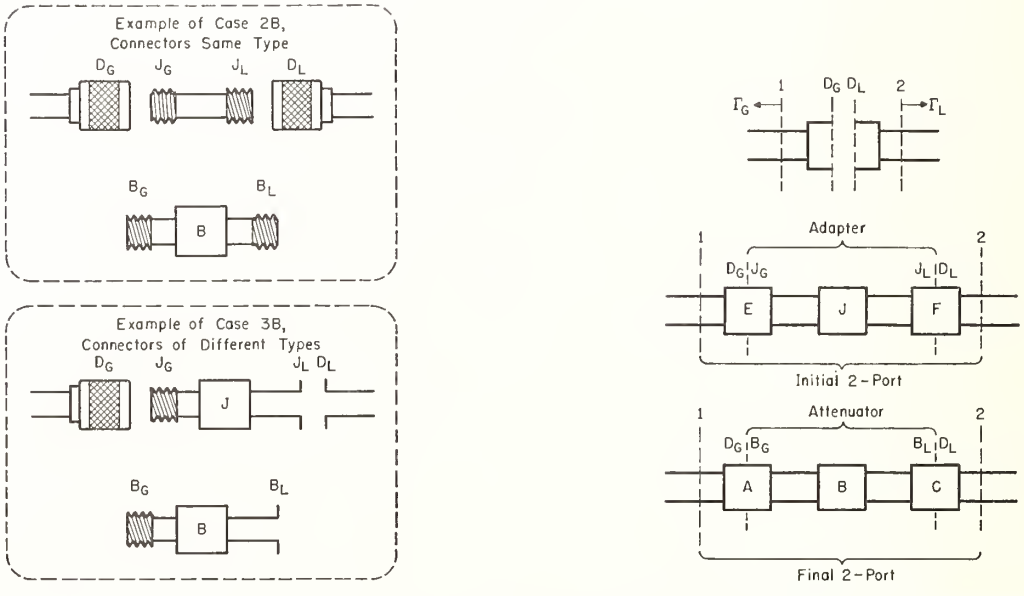


Fig. 9. Representation of insertion of a waveguide component (such as an attenuator) having nonmating connectors by substituting it for an adapter (Cases 2B and 3B).

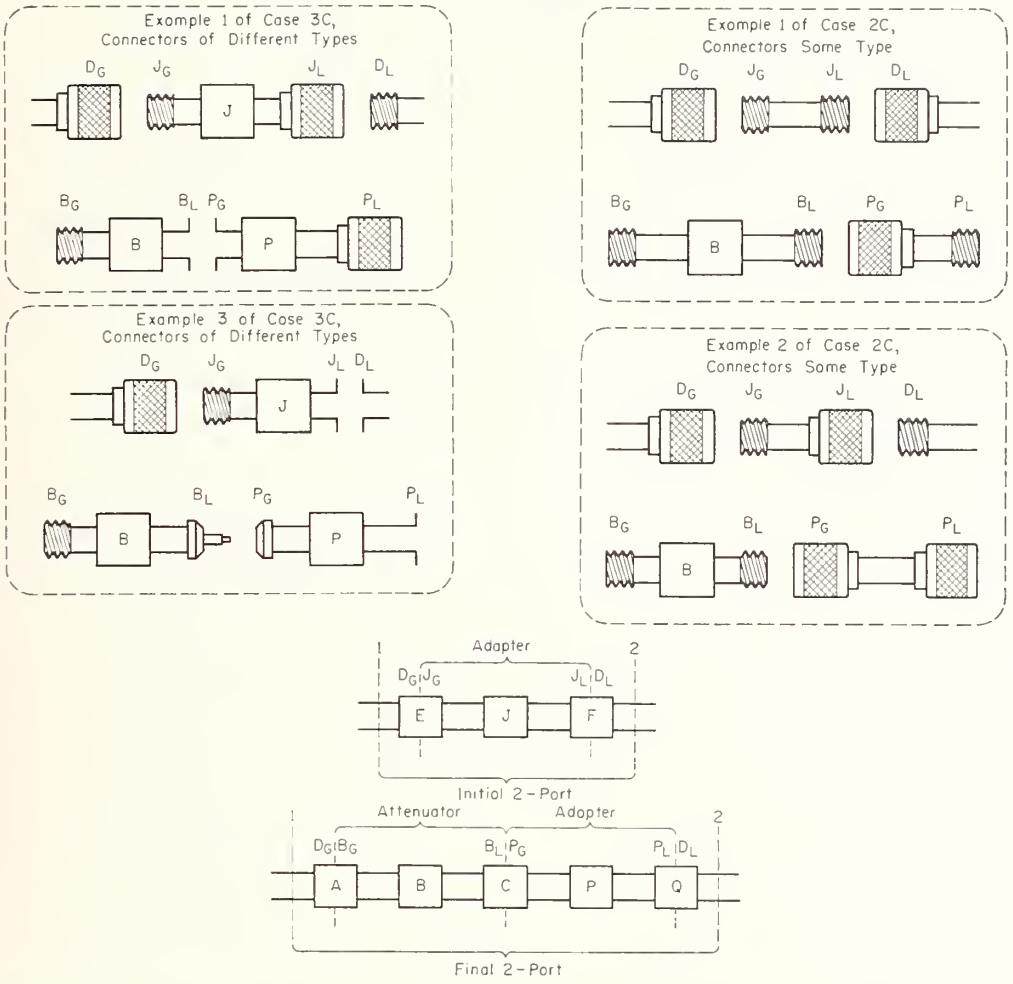
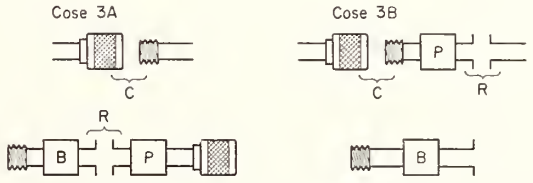


Fig. 10. Representation of insertion of a waveguide component (such as an attenuator) having nonmating connectors by connecting it to an adapter, and substituting the composite component thus formed for another adapter (Cases 2C and 3C).



Neglecting Reflections:

$L_{S1} = A_B + A_R + A_P + A_C$, AND $L_{S2} = A_B - A_P$.

$$\tfrac{1}{2}(L_{S1} + L_{S2}) = A_B + \frac{A_R}{2} + \frac{A_C}{2}$$

Fig. 11. Example of arrangement for measuring attenuation approximately characteristic of waveguide component of Case 3C-

that adapters also need to be standardized when used in precision measurement techniques.

Some basic insertion arrangements are described for waveguide components having nonmating connectors. The ones giving a measured loss most nearly characteristic of the waveguide component are singled out for special mention, although a complete analysis is not presented.

Appendix

It is convenient to have available for reference the expressions for the scattering coefficients of composite two-ports in terms of the scattering coefficients of their individual component two-ports. In the following, the individual two-ports have been designated A, B, and C reading from left to right, with terminal surface No. 1 on the left. The scattering coefficients of each two-port are designated by the corresponding small letter with appropriate subscripts. Other letter designations and numbering can be used simply by the proper substitution of letters and interchange of subscripts in a straightforward manner.

For *two cascaded two-ports A and B*, the scattering coefficients of the composite two-port are

$$\begin{cases} S_{11} = a_{11} + b_{11}\dfrac{a_{12}a_{21}}{1 - a_{22}b_{11}} & S_{22} = b_{22} + a_{22}\dfrac{b_{12}b_{21}}{1 - a_{22}b_{11}} \\ S_{12} = a_{12}b_{12}\dfrac{1}{1 - a_{22}b_{11}} & S_{21} = a_{21}b_{21}\dfrac{1}{1 - a_{22}b_{11}} \end{cases} \quad (24)$$

For *three cascaded two-ports A, B, and C*,

$$\begin{cases} S_{11} = a_{11} + b_{11}\dfrac{a_{12}a_{21}}{1 - a_{22}b_{11}} + c_{11}\dfrac{a_{12}a_{21}b_{12}b_{21}}{(1 - a_{22}b_{11})\left[(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}\right.} \\ S_{12} = \dfrac{a_{12}b_{12}c_{12}}{(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}} \\ S_{21} = \dfrac{a_{12}b_{21}c_{21}}{(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}} \\ S_{22} = c_{22} + b_{22}\dfrac{c_{12}c_{21}}{1 - b_{22}c_{11}} + a_{22}\dfrac{b_{12}b_{21}c_{12}c_{21}}{(1 - b_{22}c_{11})\left[(1 - a_{22}b_{11})(1 - b_{22}c_{11}) - a_{22}b_{12}b_{21}c_{11}\right]} \end{cases} \quad (25)$$

The form of (3) is frequently observed in the error equations. In order to facilitate calculations, the nomogram of Fig. 12 may be used. The nomogram [10], [11] gives only error limits assuming that there is some possibility that the phases of the reflections might combine to produce the greatest effect. In some cases, the error limits given by this procedure are overly conservative because the variation of the phases is limited by realizability conditions [12].

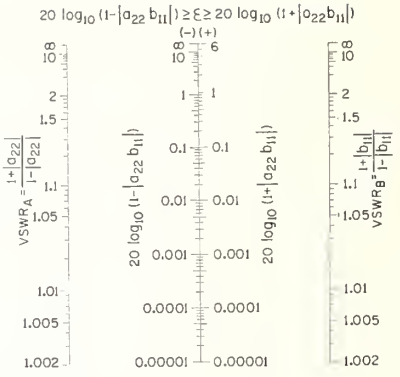


Fig. 12. Nomogram of "Error Limits."

Acknowledgment

The cooperation and constructive criticisms of D. H. Russell and R. C. Powell of NBS are gratefully acknowledged.

References

[1] Beatty, R. W., Insertion loss concepts, *Proc. IEEE*, vol 52, Jun 1964, pp 663–671.

[2] Weber, E., The measurement of attenuation, in *Technique of Microwave Measurements*, vol 11, M.I.T. Rad. Lab. Ser., New York: McGraw-Hill, Ch 12, 1947.

[3] Beatty, R. W., Mismatch errors in the measurement of ultra-high-frequency and microwave variable attenuators, *J. Res. NBS*, vol 52, Jan 1954, pp 7–9.

[4] Kerns, D. M., Basis of the application of network equations to waveguide problems, *J. Res. NBS*, vol 42, May 1949, pp 515–540.

[5] Weill, V., Insertion VSWR and insertion loss measurements of connector pairs in coaxial cables, *Microwave J.*, vol 7, Jan 1964, pp 68–73.

[6] Russell, D. H., NBS, Boulder, Colo., private communication.

[7] Mathis, H. F., Measurements of reflection coefficients through a lossless network, *IRE Trans. on Microwave Theory and Techniques*, vol 3, Oct 1955, p 58.

[8] Weinschel, B. O., "Development of attenuation measurements and standards, *J. Res. NBS*, vol 64D, Nov–Dec 1960, p 599.

[9] Schafer, G. E., and A. Y. Rumfelt, Mismatch errors in cascade-connected variable attenuators, *IRE Trans. on Microwave Theory and Techniques*, vol 7, Oct 1959, pp 447–453.

[10] Beatty, R. W., Cascade-connected attenuators, *Proc. IRE*, vol 38, Oct 1950, p 1190.

[11] Feist, H. J., Nomogram of mismatch error limits, *The Microwave Engineer's Handbook and Buyer's Guide*. Dedham, Mass.: Horizon House, 1963, p T-15.

[12] Youla, D. C., and P. M. Paterno, Realizable limits of error for dissipationless attenuators in mismatched system, *IEEE Trans. on Microwave Theory and Techniques*, vol 12, May 1964, p 289.

Mismatch Errors in Microwave Phase Shift Measurements*

G. E. SCHAFER†, SENIOR MEMBER, IRE

Summary—The phase difference between the incident and transmitted waves at the input and output ports, respectively, of a two-arm waveguide junction in a reflection free system is a characteristic of the waveguide junction and is defined as the "phase shift." The difference between the phase shift in a reflection free system and the "change of phase" observed in a system which is not reflection free will be termed mismatch error. The mismatch error depends not only on the reflections present in the system but also on the choice of the wave used as the reference wave in a phase measurement. Similar considerations hold for the measurements of variation of phase shift and the observed change of phase in adjustable components.

A formal scattering matrix analysis is used to derive expressions for phase relationships of the wave amplitudes for a two-arm waveguide junction in a system with reflections. The results of this analysis are used to evaluate mismatch error for different choices of reference waves. Two techniques of variation of phase shift measurements are analyzed. Graphs of the limits of mismatch error in a commonly used method of measurement are presented.

Introduction

"THE phase shift through a waveguide component at a ingle frequency is the phase difference under matched conditions between corresponding incident and transmitted field quantities at the input and output ports, respectively, ignoring multiples of 2π radians."[1] From this definition, it is seen that the phase shift through a waveguide component is a characteristic of the component. However, if the component is inserted in a system which has reflections, two interactions take place which cause errors in measurements of phase shift. It will be shown that the phase difference between the emergent wave from the output port (transmitted wave) and the wave incident at the input port (incident wave) depends only on the reflection coefficient of the equivalent load attached to the junction and the characteristics of the junction. However, the phase of the incident wave with respect to some independent reference such as the component of the incident wave supplied by the generator depends on the reflection coefficients of both the load and the generator, and the characteristics of the junction. Consequently, the phase of the emergent wave with respect to an independent reference depends on the reflection coefficients of the load and generator and characteristics of the junction. The difference between the phase shift and the phase change observed will be termed a mismatch error. Care must be exercised to determine which wave is being used as a reference in evaluating these mismatch errors. Similar considerations hold for measurements of variation of phase shift and the observed change of phase in adjustable components such as microwave phase shifters or attenuators.

A scattering matrix analysis is used to derive the phase relationship among various wave amplitudes in a two-arm waveguide junction inserted in a system with reflections. Mismatch errors are evaluated for two choices of reference waves. Two commonly used methods of measuring variations of phase shift in adjustable components are analyzed for mismatch error. Limits of mismatch error are calculated for the first method, a two channel arrangement, and presented in two graphs. One graph presents limits of error for lossless components and is valid for low loss phase shifters. The limits of mismatch error for a lossless phase shifter are slightly larger than those which would be encountered in a low loss component such as a commercial phase shifter, or in an attenuator when one or both of the settings is less than 20 db. The other graph is for components which have at least 20 db loss at both settings. This graph is presented since such measurements have smaller limits of mismatch error.

The second method which is treated uses a short circuit and slotted line to measure the phase shift or variation of phase shift of low-loss components. The error is evaluated and it is found to depend to the first order only on the mismatches of the component and not on the mismatches of the generator.

Theory

A two-arm waveguide junction may be represented as in Fig. 1. The phase of the emergent wave from arm 2

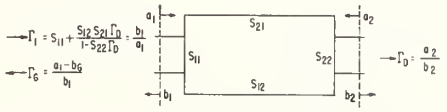


Fig. 1—A two-arm waveguide junction representation.

(the transmitted wave) with respect to the other waves associated with the junction may be derived by the use of the scattering matrix, S. In terms of this matrix,

$$b = Sa \tag{1}$$

where b is a column matrix of the emergent wave ampli-

* Received by the PGMTT, May 10, 1960; revised manuscript received, July 13, 1960.

† Radio Standards Lab., National Bureau of Standards, Boulder, Colo.

[1] "IRE standards on antennas and waveguides: waveguide and waveguide component measurements, 1959," PROC. IRE, vol. 47, pp. 568–582; April, 1959.

tudes, a is a column matrix of the incident wave amplitudes, and S is the scattering matrix of the junction. It can readily be shown from (1) and Fig. 1 that the output wave is related to the component of the input wave supplied by the generator b_G by

$$\frac{b_2}{b_G} = \frac{S_{21}}{(1 - \Gamma_1\Gamma_G)(1 - S_{22}\Gamma_D)} \tag{2}$$

where the S's are elements of the scattering matrix associated with the two arm junction, and Γ_G and Γ_D are the equivalent generator and detector reflection coefficients, respectively, and Γ_1 is the reflection coefficient of the equivalent load attached to the generator. Γ_1 may be expressed as

$$\Gamma_1 = S_{11} + \frac{S_{12}S_{21}\Gamma_D}{1 - S_{22}\Gamma_D} . \tag{3}$$

The argument of (2) is the phase difference between the emergent wave, b_2, and b_G, which is the wave that would be delivered to a reflectionless load. b_G is independent of the reflections of the system and therefore is termed the independent wave.

Eq. (2) may be written in the form

$$\frac{b_2}{b_G} = \frac{b_2}{a_1} \cdot \frac{a_1}{b_G} \tag{4}$$

where

$$\frac{b_2}{a_1} = \frac{S_{21}}{1 - S_{22}\Gamma_D} \tag{5}$$

and

$$\frac{a_1}{b_G} = \frac{1}{(1 - \Gamma_1\Gamma_G)} . \tag{6}$$

The argument of b_2/a_1 is the phase difference between the transmitted wave and the incident wave. The argument of a_1/b_G is the phase difference between the incident wave and the independent generator wave, b_G. When $\Gamma_D = \Gamma_G = 0$, the phase difference of (6) reduces to zero and both (2) and (5) reduce to

$$\frac{b_2}{a_1} = S_{21} = |S_{21}| e^{j\phi_{21}} \tag{7}$$

where ϕ_{21} is, by definition, the phase shift through the waveguide component.

Evaluation of Mismatch Error

Case I. The Reference Wave is the Independent Wave, b_G

In techniques where the independent wave is used as the reference wave, the mismatch error for a phase shift measurement may be obtained by rewriting (2) in the form

$$\frac{b_2}{b_G} = |S_{21}| e^{j\phi_{21}} |E_a| e^{j\epsilon_a}, \tag{8}$$

where

$$|E_a| e^{j\epsilon_a} = \frac{1}{(1 - \Gamma_1\Gamma_G)(1 - S_{22}\Gamma_D)} . \tag{9}$$

The difference between the measured change of phase and the phase shift of the component is just ϵ_a, the argument of (9). For differential phase shifters or attenuators, using front superscripts i and f to denote initial and final settings, respectively, the change of phase of b_2 with respect to the independent wave may be obtained from an expression derived from (2), which is

$$\frac{{}^f b_2}{{}^i b_2} = \frac{{}^f S_{21}}{{}^i S_{21}} \quad \frac{(1 - {}^i\Gamma_1\Gamma_G)(1 - {}^i S_{22}\Gamma_D)}{(1 - {}^f\Gamma_1\Gamma_G)(1 - {}^f S_{22}\Gamma_D)} , \tag{10}$$

where the argument of (10) is the change in phase of the emergent wave, b_2, with respect to the independent wave when the setting of the junction is changed. For $\Gamma_D = \Gamma_G = 0$, this reduces to

$$\frac{{}^f b_2}{{}^i b_2} = \frac{{}^f S_{21}}{{}^i S_{22}} = \frac{|{}^f S_{21}|}{|{}^i S_{21}|} e^{j({}^f\phi_{21} - {}^i\phi_{21})}, \tag{11}$$

where ${}^f\phi_{21} - {}^i\phi_{21}$ is the variation of phase shift when the setting of the junction is changed. For Γ_G and Γ_D not zero, (10) may be written in the form

$$\frac{|{}^f b_2|}{|{}^i b_2|} e^{j({}^f\psi_2 - {}^i\psi_2)} = \frac{|{}^f S_{21}|}{|{}^i S_{21}|} e^{j({}^f\phi_{21} - {}^i\phi_{21})} |E_b| e^{j\epsilon_b}, \tag{12}$$

where ${}^f\psi_2 - {}^i\psi_2$ is the change in phase of the emergent wave for Γ_D and Γ_G not zero, ${}^f\phi_{21} - {}^i\phi_{21}$ is the variation of phase shift of the component, and

$$E_b = |E_b| e^{j\epsilon_b} = \frac{(1 - {}^i\Gamma_1\Gamma_G)(1 - {}^i S_{22}\Gamma_D)}{(1 - {}^f\Gamma_1\Gamma_G)(1 - {}^f S_{22}\Gamma_D)} . \tag{13}$$

From (12) it can be seen that the mismatch error in this case is the argument of (13).

Case II. The Reference Wave is the Incident Wave at the Input, a_1

In techniques where the incident wave is used as the reference wave, (5) may be written in the form

$$\frac{b_2}{a_1} = |S_{21}| e^{j\phi_{21}} |E_c| e^{j\epsilon_c} \tag{14}$$

where ${}^f\phi_{21}$ is the phase shift of the component and

$$|E_c| e^{j\epsilon_c} = \frac{1}{1 - S_{22}\Gamma_D} . \tag{15}$$

It can be seen that the argument of (15) is the mismatch error when the incident wave is used as a reference. For adjustable components, the change in phase of the emergent wave may be written, when the incident input wave is used as a reference wave, as

$$\frac{{}^f b_2}{{}^i b_2} = \frac{{}^f S_{21}}{{}^i S_{21}} \frac{1 - {}^i S_{22}\Gamma_D}{1 - {}^f S_{22}\Gamma_D} \tag{16}$$

or as

$$\frac{{}^{f}b_2}{{}^{i}b_2} = \frac{|\,{}^{f}S_{21}\,|}{|\,{}^{i}S_{21}\,|}\, e^{j({}^{f}\phi_{21} - {}^{i}\phi_{21})}\,|\,E_d\,|\, e^{j\epsilon_d} \qquad (17)$$

where

$$|\,E_d\,|\, e^{j\epsilon_d} = \frac{1 - {}^{i}S_{22}\Gamma_D}{1 - {}^{f}S_{22}\Gamma_D} \qquad (18)$$

and ϵ_d is the mismatch error in this case.

Application I

A two-channel method of measuring variation of phase shift is illustrated in Fig. 2. Usually there is considerable isolation between the component under test and power dividing network. Under these circumstances, a portion of the energy of the oscillator traverses a separate isolated path and behaves as an independent reference wave for the change of phase measurements.

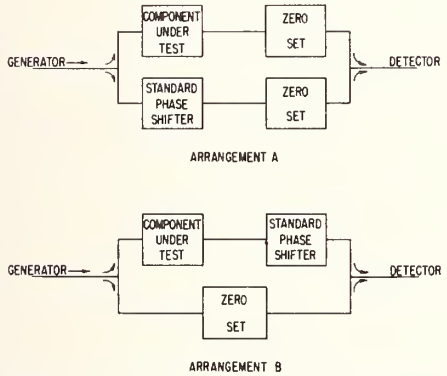


Fig. 2—Arrangement of equipment for a two-path method of measuring phase shift.

Without this isolation, the equipment may be adjusted to use the incident wave as a reference. The present discussion assumes this isolation to be infinite. Since the independent wave is used as a reference, the mismatch error is given by (13). If the magnitudes of the terms other than $|E|$ and unity in (13) are small compared to unity then to a good approximation,

$$|\,E\,|\, e^{j\epsilon} = 1 - {}^{i}\Gamma_1\Gamma_G - {}^{i}S_{22}\Gamma_D + {}^{f}\Gamma_1\Gamma_G + {}^{f}S_{22}\Gamma_D, \qquad (19)$$

and the mismatch error, ϵ, may be written approximately as

$$\epsilon = \text{argument of } [1 - {}^{i}\Gamma_1\Gamma_G - {}^{i}S_{22}\Gamma_D + {}^{f}\Gamma_1\Gamma_G + {}^{f}S_{22}\Gamma_D]. \qquad (20)$$

However, it is inconvenient or sometimes difficult to evaluate the phases of the scattering and reflection coefficients, while limits of their magnitudes are more readily determined from estimates of maximum VSWR. Therefore, limits of error (maximum error) for arbitrary phases of these coefficients are evaluated here. One may represent (19) in graphical form as shown in Fig. 3. Allowing the phases of the coefficients to take on appropriate values, the maximum error, lim ϵ, will occur when the resultant is 90° out of phase with the variables, as shown in Fig. 4. Under these conditions

$$\sin(\lim \epsilon) = |\,{}^{i}\Gamma_1\Gamma_G\,| + |\,{}^{i}S_{22}\Gamma_D\,| + |\,{}^{f}\Gamma_1\Gamma_D\,| + |\,{}^{f}S_{22}\Gamma_D\,|, \qquad (21)$$

which for small angles may be written,

$$\lim \epsilon = |\,{}^{i}\Gamma_1\Gamma_G\,| + |\,{}^{i}S_{22}\Gamma_G\,| + |\,{}^{f}\Gamma_1\Gamma_G\,| + |\,{}^{f}S_{22}\Gamma_D\,|. \qquad (22)$$

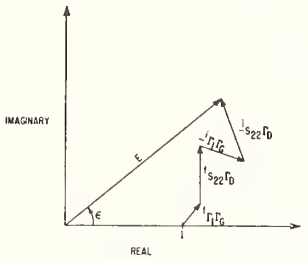


Fig. 3—Representation of (19).

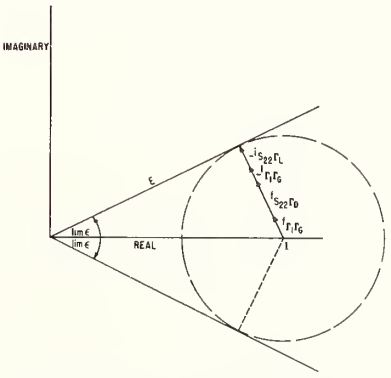


Fig. 4—Representation of (19) for maximum ϵ.

A conservative estimate of the limits of mismatch error may be quickly obtained by using the manufacturers' specifications for the magnitudes of the scattering coefficients. It should be noted, however, that for a specific measurement, determining the value of the magnitude of the scattering coefficients will usually result in smaller limits of error since manufacturers generally specify only the maximum value over the entire operating range. It should be emphasized that the limits of error calculated from (22) are maximum errors based on the assumption that the phases of the scattering coefficients change an arbitrary amount. Limits to these phase changes can frequently be determined, and for precise measurements, it is then desirable to use these limits and determine the smaller limits of error by use of (13).

Graphical Presentation of Results

The graphs are constructed to present a limit of mismatch error in variation of phase shift measurements as a function of the mismatches of the generator, detector, and phase shifter. The following assumptions were made to simplify the presentation and they introduce only a small loss of generality. It is assumed that: 1) the equivalent generator and detector reflection coefficients are of equal magnitude, $|\Gamma_G| = |\Gamma_D|$; 2) the input and output voltage standing-wave ratio (VSWR) of the phase shifter are equal, and therefore $|S_{11}| = |S_{22}|$; and 3) the detector reflection coefficient and S_{11} combine to give maximum magnitude of Γ_1.

The results for a lossless phase shifter are presented in Fig. 5, and for a lossy one in Fig. 6. The limits of error are plotted against the input or output VSWR (σ_{11} or σ_{22}) of the phase shifter or attenuator. Fig. 5 gives limits of error for a lossless phase shifter and conservative limits of error for phase shifters with less than 20 db loss or attenuators with one or both settings at less than 20 db loss. Fig. 6 gives limits of error for attenuators when both settings are at least 20 db or phase shifters with 20 db or more of insertion loss.

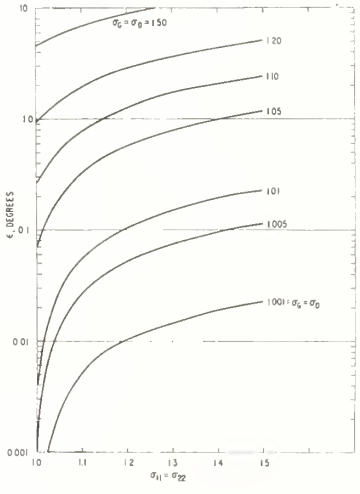


Fig. 5—Limit of mismatch error for lossless phase shifters.

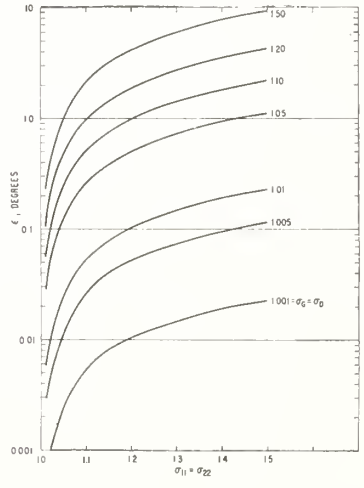


Fig. 6—Limit of mismatch error for attenuators with initial and final settings greater than 20 db.

As an example, consider a phase shifter with 0.5 db insertion loss and maximum input and output VSWR of 1.35 placed in a waveguide system with maximum VSWR looking towards the generator and detector of 1.05. The conservative limit of error as given by Fig. 5 is 0.90°. However, if this component has 20 db or more loss at both settings, the limit of error as given by Fig. 6 is 0.84°. The difference between the limits of error for the lossless and high loss cases becomes more pronounced as the ratio of $|\Gamma_D|$ to $|S_{11}|$ becomes larger and therefore both graphs are presented.

The graphs may also be used to estimate the maximum permissible VSWR of the equivalent generator and detector to attain a given accuracy of variation of phase shift with a calibrated phase shifter. One case of interest is a microwave phase shifter of maximum VSWR of 1.35 which is calibrated to 2° accuracy. To utilize this accuracy, an estimate from Fig. 5 indicates that it should be used in a system where the VSWR looking towards the generator and detector are 1.10, or less. Another case of interest is the comparison of the variation of phase shift of two components within 0.1° in a two channel method. This would be satisfied if the limit of mismatch error for each component was 0.05°. If one of the components is a microwave phase shifter with maximum VSWR of 1.35, the maximum VSWR of the equivalent generator and detector for 0.05° limit is 1.004. If the other component is an attenuator with maximum VSWR of 1.15, the maximum VSWR of the equivalent generator and detector for 0.05° limit is 1.006.

It may be useful here to emphasize the meaning of the limits of error presented in the graphs. These are maximum errors due to mismatches, since it was assumed that the phase changes of all coefficients were arbitrary. If one has knowledge of the limits of phase changes of the coefficients, or actual values, it is desirable for critical work to turn to (11) and evaluate closer limits of mismatch error, or actual mismatch error.

Application II

A method which has been used to measure the variation of phase shift of a low-loss reciprocal waveguide component by terminating it with a calibrated sliding short circuit and using a slotted section as a detector is illustrated in Fig. 7. A minimum of the input standing wave pattern is used as a reference to position the probe. When the component under test is adjusted to a

new setting, the minimum of the input standing wave pattern is restored to the reference plane of the probe by moving the calibrated short circuit. Since the wave travels through the component in both directions to form this pattern, twice the phase shift of the component is assumed to be equal to the change of phase of the reflection coefficient of the attached short circuit.

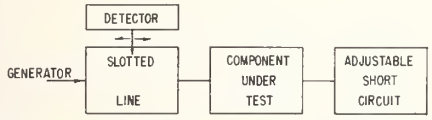


Fig. 7—Arrangement of equipment for a short circuit method of measuring phase shift of low-loss components.

Neglecting any errors caused by probe loading, the minimum of the input standing wave pattern occurs when

$$\text{argument of } \frac{b_1}{a_1} = (2n+1)\pi, \tag{23}$$

where n is an integer. The minimum of the pattern is restored to the initial position by adjustment of the short circuit and this condition may be expressed by

$$\arg \frac{{}^{f}b_1}{{}^{f}a_1} = \arg \frac{{}^{i}b_1}{{}^{i}a_1}, \tag{24}$$

or

$$\arg {}^{f}\Gamma_1 = \arg {}^{i}\Gamma_1 \tag{25}$$

where Γ_1 is the input reflection coefficient of the component when arm 2 is terminated with a sliding short circuit with reflection coefficient Γ_s. Substitution of an expression for Γ_1 in terms of the scattering coefficients of the component and the short circuit allows the adjustment conditions to be written as

$$\arg\left({}^{f}S_{11} + \frac{{}^{f}S_{21}{}^{2}\,{}^{f}\Gamma_s}{1 - {}^{f}S_{22}{}^{f}\Gamma_s}\right) = \arg\left({}^{i}S_{11} + \frac{{}^{i}S_{21}{}^{2}\,{}^{i}\Gamma_s}{1 - {}^{i}S_{22}\,{}^{f}\Gamma_s}\right), \tag{26}$$

when reciprocity in the form, $S_{21} = S_{12}$, has been assumed. For $|S_{22}\Gamma_s| \ll 1$, this may be written approximately as

$$\arg ({}^{f}S_{11} + {}^{f}S_{21}{}^{2}\,{}^{f}\Gamma_s + {}^{f}S_{21}{}^{2}\,{}^{f}S_{22}\,{}^{f}\Gamma_s{}^{2}) = \arg ({}^{i}S_{11} + {}^{i}S_{21}{}^{2}\,{}^{i}\Gamma_s + {}^{i}S_{21}{}^{2}\,{}^{i}S_{22}\,{}^{i}\Gamma_s{}^{2}). \tag{27}$$

The measured variation of phase shift is based on the assumption that $S_{11} = S_{22} = 0$ and that the arguments of ${}^{f}S_{21}{}^{2}\,{}^{f}\Gamma_s$ and ${}^{i}S_{21}{}^{2}\,{}^{i}\Gamma_s$ are equal, which leads to

$${}^{f}\phi_{21} - {}^{i}\phi_{21} = \tfrac{1}{2}({}^{i}\psi_s - {}^{f}\psi_s) \tag{28}$$

where ${}^{i}\psi_s$ and ${}^{f}\psi_s$ are the initial and final phases, respectively, of the reflection coefficient of the sliding short circuit, Γ_s.

It is apparent, however, that the actual change of phase can differ from this ideal when S_{11} and S_{22} are not zero, or

$$({}^{f}\phi_{21} - {}^{i}\phi_{21}) - \tfrac{1}{2}({}^{i}\psi_s - {}^{f}\psi_s) = \epsilon_e, \tag{29}$$

where ϵ_e is the mismatch error in this method. Eq. (29) can be shown to be equivalent to

$$2\epsilon_e = \arg \frac{{}^{f}S_{21}{}^{2}\,{}^{f}\Gamma_s}{{}^{i}S_{21}{}^{2}\,{}^{i}\Gamma_s} = \arg {}^{f}S_{21}{}^{2}\,{}^{f}\Gamma_s - \arg {}^{i}S_{21}{}^{2}\,{}^{i}\Gamma_s. \tag{30}$$

The difference between these two arguments, $2\epsilon_e$, can be seen from Fig. 8, a graphical representation of (27) which describes the actual adjustment of conditions for the general case. The limits of this difference, assuming all phases of the reflection coefficients are possible, may be obtained from

$$\begin{aligned}\sin(\lim \epsilon_e) &= \frac{1}{2}\left[\sin(\lim\epsilon_1) + \sin(\lim\epsilon_2)\right] \\ &= \frac{1}{2}\,\frac{|{}^{f}S_{11}| + |{}^{f}S_{21}{}^{2}\,{}^{f}S_{22}\,{}^{f}\Gamma_s{}^{2}|}{|{}^{f}S_{21}{}^{2}\,{}^{f}\Gamma_s|} \\ &\quad + \frac{1}{2}\,\frac{|{}^{i}S_{11}| + |{}^{i}S_{21}{}^{2}\,{}^{i}S_{22}\,{}^{i}\Gamma_s{}^{2}|}{|{}^{i}S_{21}{}^{2}\,{}^{i}\Gamma_s|}\end{aligned} \tag{31}$$

where $\lim \epsilon_e$ is the limit of error, and $\lim \epsilon_1$ and $\lim \epsilon_2$ are limits of ϵ_1 and ϵ_2 as shown in Fig. 8. A readily calculated approximation for the limit of error may be found by assuming $|S_{21}| = 1$, $|\Gamma_s| = 1$, and $|S_{11}|$ and $|S_{22}|$ do not change with adjustment. This approximate limit of error may be obtained from

$$\sin(\lim \epsilon_e) \approx |S_{22}| + |S_{11}|. \tag{32}$$

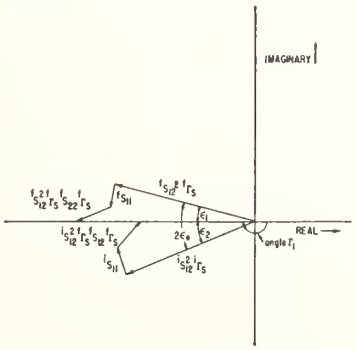


Fig. 8—Representation of (27).

If the input and output VSWR of the component are equal and the error is small, the limit of mismatch error may be written as

$$\lim \epsilon_e = 2 \left| S_{11} \right| \text{ radians} \tag{33}$$

which can be readily shown to be equivalent to a result for lossless components quoted by Magid.[2] It may be useful here to emphasize the meaning of the limits of error obtained by (33). These are maximum errors due to mismatches, since it was assumed that the phase changes of all coefficients were arbitrary.

[2] M. Magid, "Precision microwave phase shift measurements," IRE TRANS. ON INSTRUMENTATION, vol. I-7, pp. 321–331; December, 1958.

For the same phase shifter considered in Application I (VSWR <1.35), the limits of error are $\pm 17°$. It is of interest to note that the mismatch error in a variation of phase shift measurement in this method is independent of the reflection coefficient of the generator.

Additional errors in this method such as those caused by probe loading in the slotted line are not within the scope of this analysis, but should be taken into account, if they are appreciable compared to the mismatch error.

ACKNOWLEDGMENT

The author extends thanks to R. W. Beatty for constructive criticism of this analysis, and to O. L. Patty and W. A. Downing for performing the necessary calculations to evaluate the limits of error.

Reprinted from

IRE TRANSACTIONS ON MICROWAVE THEORY

AND TECHNIQUES

1960

A MODULATED SUB-CARRIER TECHNIQUE OF MEASURING MICROWAVE ATTENUATION

By G. E. SCHAFER Ph.D., and R. R. BOWMAN.

(*The paper was received 5th June*, 1961. *It was presented at the* CONFERENCE ON MICROWAVE MEASUREMENT TECHNIQUES *7th September*, 1961.)

SUMMARY

A method of measuring microwave attenuation is proposed which has the advantages of an i.f. substitution method with single-sideband operation. However, ordinary amplitude modulation is used, and neither the carrier nor one of the sidebands needs to be suppressed.

Two versions of this method are described and some operational hints are given. One of these versions is capable of high accuracy with commercially available equipment. The proposed method was tested by measuring the relative attenuation of a microwave variable attenuator at 9·389 7 Gc/s, attaining a precision of 0·000 1 dB at 0·01 dB and 0·2 dB at 50 dB. The measurements are compared with measurements performed by other methods. A special comparison with values obtained from d.c. substitution techniques was made in which environmental effects were largely eliminated.

Factors affecting the accuracy of measurements made by this technique are discussed. Some of the precautions necessary to attain high accuracy are given.

(1) INTRODUCTION

Several methods of measuring microwave attenuation[1] are currently in use. Some of these are often characterized by the terms r.f., i.f., a.f. and d.c. substitution. A particular method may fall into one of these categories, depending on the frequency at which the standard or calibrated attenuator is operated.

At present, i.f. (power linear frequency conversion) and a.f. (square-law detection) systems are most commonly used. One advantage of the i.f. substitution system over the a.f. one is the wide dynamic range over which accurate measurements can be made: one advantage of the a.f. substitution system over the i.f. one is the need for only one microwave signal source. Single-sideband i.f. substitution systems can, in principle, combine these two advantages. However, there is additional difficulty and expense involved in reducing the unwanted sideband and carrier energies to satisfactorily low levels.

The method here described combines these two advantages in a way which does not require the suppression of either the carrier or one of the sidebands. Two different versions of this method are described, and some operational hints on obtaining high accuracy are given.

Attenuation measurements are compared with those obtained from a modified d.c. substitution[2] technique and a conventional i.f. substitution technique. A special comparison in which environmental effects are largely eliminated is also reported.

The system is also capable of simultaneously measuring the phase shift of the attenuator under test. Methods of doing this are described.

(2) THEORY

Referring to Fig. 1, microwave attenuation is measured by the modulated sub-carrier method by use of a 2-channel system operating as follows. One channel, containing controls for the phase and magnitude of the wave amplitude, transmits a portion (approximately 1 mW) of the unmodulated energy to the microwave detector and is somewhat analogous to the local oscillator in an i.f. system. This is termed the 'carrier channel'. The

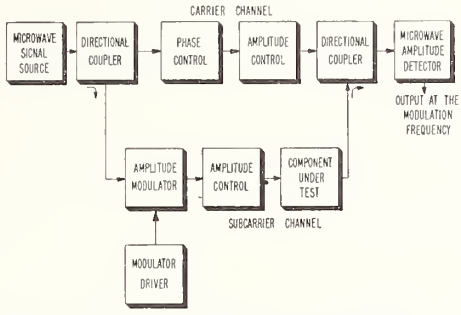


Fig. 1.—Basic microwave equipment for modulated sub-carrier method of measuring attenuation.

second channel, containing the attenuator under test, transmits a smaller fraction of the microwave energy to the same microwave detector, after this energy has been amplitude modulated. This is termed the 'sub-carrier channel'.

The general relationships amongst the wave amplitudes at the microwave detector in such a 2-channel system are shown in Fig. 2. If the carrier and amplitude-modulated sub-carrier are

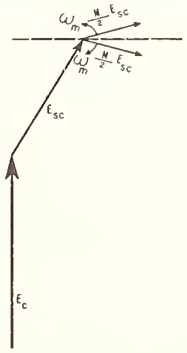


Fig. 2.—Wave amplitudes at the microwave detector.

in phase, the recombined signal delivered to the microwave detector is amplitude modulated in the same way as the sub-carrier. Therefore, in the absence of phase modulation, the magnitude of the resultant amplitude delivered to the detector, E_r, may be written

$$E_r = E_c + E_{sc}(1 + m\cos\omega_m t), \quad . \quad . \quad . \quad (1)$$

where E_c and E_{sc} are the carrier and sub-carrier amplitudes, m is the modulation index and ω_m the angular frequency of the amplitude modulation. Eqn. (1) may also be written in the form

$$E_r = (E_c + E_{sc})\left[1 + \frac{mE_{sc}}{(E_c + E_{sc})}\cos\omega_m t\right] \quad . \quad (2)$$

This is equivalent to an amplitude-modulated wave with a magnitude of $E_c + E_{sc}$ and a modulation index of $mE_{sc}/(E_c + E_{sc})$. The amplitudes of the sidebands are proportional to mE_{sc}.

The authors are at the National Bureau of Standards, Boulder Laboratories, Colorado, U.S.A.

The Institution of Electrical Engineers
Paper No. 3710 E May 1962

Therefore, if m is kept constant the amplitude of the output at the modulation frequency from a power linear detection of this combined signal will be proportional to E_{sc}, the signal passing through the attenuator under test. The term 'power linear detection' is here used to designate the condition $P_m = GP_s$, where P_m is the power at the modulation frequency delivered to the load attached to the detector, P_s is the sum of the powers in the appropriate sidebands of the microwave signal, and G is the conversion efficiency. The detection of a crystal diode is power linear over a large range of values of E_{sc} if the effective carrier, $E_c + E_{sc}$, is kept constant, a necessary condition for G to be constant. Thus, the attenuation of the modulation-frequency power may be substituted for the microwave attenuation in the sub-carrier channel.

The attenuation of the modulation-frequency power delivered from the output of the microwave detector is measured by a standard attenuator such as a standard ratio transformer. The range of attenuation measurement, as in i.f. substitution systems, is limited at the low end by noise and at the high end by departures from power linearity in excess of the mimimum error to be tolerated. In the versions shown in Figs. 3 and 4, the audio detector serves only to indicate a null or to determine when a previously used signal level is duplicated. This does not impose any requirement for linearity on the audio amplifier, and the law of the audio detector need not be known.

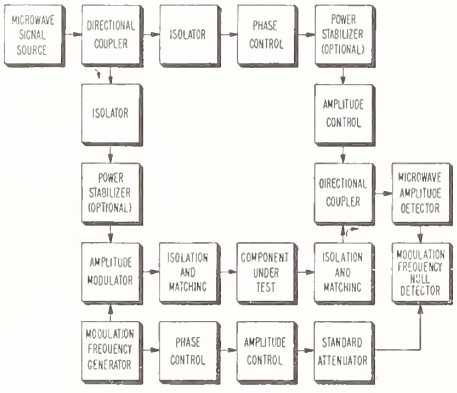


Fig. 3.—Arrangement of equipment for system with null detector.

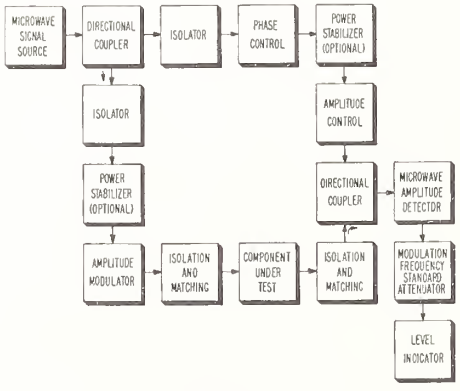


Fig. 4.—Arrangement of equipment for system with constant-level indicator.

(3) THE MECHANICS OF MAKING MEASUREMENTS

Two versions of making relative attenuation measurements by this technique are described in the following paragraphs. For the purpose of these descriptions it is assumed that the relative attenuation* of the variable attenuator is to be determined for a selected dial setting. Minor modifications of the versions permit determining the dial setting of the attenuator which corresponds to a selected value of relative attenuation.

As in any measurement of microwave attenuation by substitution methods, certain precautions are necessary at the ports of the system to which the microwave and standard attenuators are attached. Ideally, the ports to which the microwave attenuator is attached should have zero reflection coefficients. The standard attenuator should also be properly terminated.

The unmodulated carrier power delivered to the microwave detector (usually a crystal diode) is adjusted to approximately 1 mW by means of the level-set attenuator in the carrier channel. The modulated sub-carrier power delivered to the microwave detector (with the attenuator under test set at the minimum attenuation to be used) is adjusted to be sufficiently less than the carrier power to attain the desired power linearity. For less than 0·01 dB departure from power linearity, it is adequate to have the sub-carrier 23 dB (as determined experimentally) below the carrier, as in an i.f. substitution system. The amplitude modulator is adjusted to the centre frequency of the modulation-frequency detector and should produce a large percentage of modulation of the sub-carrier (at least 30%). The output signal from the microwave detector is directly proportional to this modulation index, and therefore this index should be as high as practical. Both 100% and over-modulation have been successfully used. Some modulators introduce phase as well as amplitude modulation. By following the instructions given below, the same component of amplitude modulation will appear along the resultant. Therefore the accuracy of the measurement should not be impaired if the phase modulation introduced remains constant. This has been verified experimentally. As will be discussed later, it was found that holding only the carrier constant (instead of $E_c + E_{sc}$) results in accurate measurements.

(3.1) Version I: A Null Detection Arrangement

The equipment is arranged as in Fig. 3. The microwave attenuator is set to its 'zero' or reference position. The carrier is adjusted to be in phase with the sub-carrier. Maximum output at the modulation frequency from the microwave detector may be used to determine this condition. Another maximum output occurs when the carrier and sub-carrier are 180° out of phase, which is equivalent to reversing the sign of E_{sc} in eqns. (1) and (2). The technique is still usable under these conditions. The standard attenuator (a standard ratio transformer in the various models tested) is set to its 'zero' or reference position. The phase and amplitude of the modulation-frequency signal delivered to the standard attenuator are then adjusted to produce a null at the modulation-frequency detector. The signals from the microwave detector and the standard attenuator output are then of equal amplitudes. (It may be useful to note here that by temporarily adjusting the signal levels to be slightly different at the modulation-frequency detector, greater precision is attained in equalizing the phases of the carrier and sub-carrier. This is especially useful if phase-shift measurements are to be made simultaneously.)

After a null is obtained with both attenuators at their reference positions, the microwave attenuator is adjusted to the dial setting where the relative attenuation is to be measured. The standard attenuator is adjusted so that it nearly produces a null at the modulation-frequency detector, and, if necessary, the carrier is again synchronized with the sub-carrier. Final adjustment of the standard attenuator is then made to again produce a null

* 'Relative attenuation' is here used to mean the difference in attenuation between the minimum ('zero') setting and a selected dial setting.

at the modulation-frequency detector. The microwave relative attenuation is then assumed equal to the change of attenuation of the standard.

The difference between the two settings of the phase shifter in the carrier channel is a measure of the variation of phase shift of the microwave attenuator between its two settings.

(3.2) Version II: A Constant Level Indicator Arrangement

In this simplified version the equipment is arranged as in Fig. 4. The modulation-frequency generator does not deliver a reference voltage to the standard attenuator. The principal difference in arrangement from version I is that the standard attenuator is placed between the microwave detector and the modulation-frequency level indicator. The principal difference in operation is the function of the modulation-frequency amplifier and detector. In this version the detector is used to indicate when a reference level of modulation-frequency signal is reproduced. Therefore, the precision of measurement is directly proportional to the resolution of the modulation-frequency level indicator. High resolution may be attained with differential voltmeters, but this imposes an additional requirement over version I of high stability in the modulation-frequency amplifier portion of the circuit.

With the test-attenuator dial set at the position where the relative attenuation is to be measured and the standard attenuator set to its 'zero' or reference position, the carrier and sub-carrier are synchronized as in version I. The gain of the modulation-frequency amplifier is adjusted to produce a convenient reference level at the level indicator. The microwave attenuator is then adjusted to its 'zero' or reference position, and, if necessary, the carrier is again synchronized with the sub-carrier. The standard attenuator is then adjusted to reproduce the reference level. The microwave relative attenuation is then assumed to be equal to the change of attenuation of the standard.

The difference between the phase-shifter settings in the carrier channel may again be used as a measure of the variation of phase shift of the test attenuator between its two settings. The precision of the phase adjustments is usually better in version I.

(4) EVALUATION OF THE METHOD

Equipment was set up as in version I (Fig. 3), the null detection version. Commercially available components were used throughout, except for the microwave detector used in the power stabilizer. Here two commercially available bolometer mounts were modified to form a temperature-compensated bridge arrangement in order to reduce variations of power due to ambient temperature changes.

A ferrite modulator and audio oscillator were used to produce the amplitude modulation of the microwave energy. The modulation frequency was chosen to be 1 kc/s, primarily for two reasons: one was the commercial availability of components at this frequency; the other was that 1 kc/s dispersive effects on relative attenuation should be small and the modulator could precede the test attenuator. A standard ratio transformer was used for the standard attenuator. A standing-wave amplifier with a bandwidth of 40 c/s was used for the null indicator. This was connected to form a null detector by means of an unshielded transformer. The secondary was connected to the standing-wave amplifier. One side of the primary was connected via a blocking capacitor to the microwave detector, and the other side, to the output of the standard attenuator.

The noise level at the output of the microwave detector was at least 80 dB below the signal produced when the sub-carrier was 30 dB below the 1 mW carrier in the detector crystal. Approximately 10^{-15} W of noise power was delivered to the 1 000 Ω load attached to the microwave detector. The major source of noise is the crystal, and crystal selection is sometimes necessary to attain a lower noise level. Long-term drift of the system was observed to be approximately 0·002 dB per 10 min. About ten measurements can be made during such a period. The above characteristics are all indicative of the behaviour of the initial systems built and can be improved by relatively inexpensive refinements.

The voltage standing-wave ratios at the ports leading to the generator and detector were adjusted to be 1·005 or less. The microwave frequency was 9·389 7 Gc/s. This was selected because of some previous work on the test attenuator.

With systems having the above characteristics, the relative attenuation of the microwave attenuator was measured with the attenuator in a horizontal and in a vertical position. The position had some effect on the higher relative attenuation values.

Table 1

Comparison of Measurements Obtained by Modulated Sub-Carrier and D.C. Substitution Methods

Attenuator dial reading	Measured attenuation (with attenuator vertical)	
	Modulated sub-carrier method	D.C. method
dB	dB	dB
0·01	0·011 1	0·010 7
0·02	0·021 8	0·021 4
0·03	0·030 8	0·030 3
0·04	0·041 3	0·040 7
0·05	0·053 2	0·052 1
0·06	0·062 6	0·060 9
0·07	0·071 8	0·070 0
0·08	0·081 4	0·080 2
0·09	0·091 8	0·090 9
0·10	0·103 7	0·102 1
0·12	0·120 6	0·119 1
0·14	0·138 8	0·137 5
0·16	0·159 0	0·157 3
0·18	0·179 6	0·178 3
0·20	0·202 3	0·200 7
0·25	0·248 2	0·247 1
0·50	0·499 4	0·497 9
1	1·004	1·004
2	1·999	1·996
3	3·000	2·998
5	4·991	4·990
10	9·971	9·965
15	15·01	14·99
20	19·97	19·95
25	25·00	25·01
30	30·10	30·07
40	40·44	40·33
50	52·39	52·24

Table 1 presents a comparison of the modulated sub-carrier measurements with those made by the modified d.c. substitution method. The errors in the d.c. technique had been rather carefully evaluated, while in this test neither the audio standard nor the power linearity had been satisfactorily evaluated. The attenuator was in a vertical position. The significant figures in the Table are not necessarily the limiting accuracies of the different techniques: they simply indicate the precision attained for the particular environment used. Differences between measurements may be explained by the inability to repeat settings of the attenuator exactly. Other sources of error such as temperature differences in the laboratory may also be important.

Table 2 presents a comparison of the modulated sub-carrier measurements with those made by a conventional series i.f. substitution technique. The intermediate frequency was 30 Mc/s

Table 2

Comparison of Measurements Obtained by Modulated Sub-Carrier and I.F. Substitution Methods

Attenuator dial reading	Measured attenuation (with attenuator horizontal)	
	Modulated sub-carrier method	I.F. substitution method
dB	dB	dB
0·1	0·102 6	0·101
0·2	0·201 0	0·199
0·5	0·497 8	0·498
1	1·002	0·998
2	1·997	1·999
3	2·998	2·997
5	4·989	4·980
10	9·958	9·965
15	14·98	14·99
20	19·94	19·95
25	24·95	24·95
30	30·02	30·00
40	40·24	40·15
50	51.70	51·66

and a waveguide-below-cut-off attenuator was used as the standard. In this case the microwave attenuator was in a horizontal position. Again the significant figures indicate the precision attained; differences between measurements may be due to the same causes.

Table 3

Differences Between Measurements Made Simultaneously by Modulated Sub-Carrier and D.C. Substitution Methods

Attenuator dial reading	Difference (with attenuator vertical)
dB	dB
0·01	+0·000 2
0·02	−0·000 2
0·05	−0·000 1
0·10	+0·000 5
0·50	+0·000 1
1·0	−0·001 2
2·0	−0·000 3
5·0	−0·001 1
10·0	+0·001 2
20·0	+0·006 6

Table 3 shows the differences in measurements of the relative attenuation of the microwave attenuator made simultaneously by the d.c. and sub-carrier methods. Each of these is a 2-channel method, and simultaneous measurements could be made with a 3-channel system in which the common channel contained the test attenuator in a vertical position.[3] In this comparison, differences between measured relative attenuations due to inability to reset the attenuator, mismatch errors, ageing of the attenuator and environmental conditions in general are eliminated. The remaining differences are probably due to inaccuracies in the standard and in reading the settings, and power non-linearity of the crystal detector.

(5) SOME FACTORS AFFECTING ACCURACY

The power linearity of the crystal-diode microwave detector was checked for a number of crystals, with the amplitude of the carrier held constant. It was realized that the analysis shows that the sum of the amplitudes of the carrier and sub-carrier, $E_c + E_{sc}$, must be held constant instead of only the carrier, E_c. However, holding E_c constant is much more simple, and since E_{sc} is small compared with E_c in most cases, the difference in behaviour from the analytically proper operation was expected to be small. Less than 0·01 dB power non-linearity was observed at ratios of carrier to sub-carrier of 23 dB or more. It is of interest to note that this departure from power linearity is very similar to the behaviour of a crystal diode operating as a mixer in an i.f. system.[4] The maximum departures observed for ratios down to 0 dB were less than 0·5 dB.

In proper operation ($E_c + E_{sc}$ held constant), it was found that the magnitude of the power non-linearity was very nearly the same as for E_c only held constant, for ratios down to 20 dB. Proper operation could be accomplished automatically by leaving the control element installed in the carrier channel, and by moving the sensing device somewhere between the place where the signals were recombined and the microwave detector.

The effects of different loads attached to the detector crystal were investigated. A broad maximum in the observed accuracy was found with approximately a 1 000 Ω load. There was very little deterioration in accuracy for load impedances as low as 100 Ω or as high as 5 000 Ω. Outside these ranges rather large departures from linearity (greater than 0·03 dB) were observed.

(6) CONCLUSIONS

A new method of measuring microwave attenuation has been developed which may be built with commercially available components to attain 0·000 1 dB resolution and accuracies comparable to those of the best systems now known. This method offers a distinct saving in operating costs over conventional i.f. substitution systems, since only a single microwave signal source is required.

In principle, amplitude modulation at high frequencies and waveguide-below-cut-off or resistive standard attenuators could be used. It is at present more convenient to modulate at an audio frequency and use standard ratio transformers which are commercially available with 0·000 1 dB accuracies at small ratios of attenuation and 0·01 dB accuracies at 60 dB ratios. The most accurate models of these standard ratio transformers are slightly more expensive, but also more accurate, than commercially available waveguide-below-cut-off attenuators. They may also be used with essentially zero minimum insertion loss.

The null-detection version is preferable to the constant-level version because (*a*) high precision is attained without high-resolution differential detectors, and (*b*) if a standard ratio transformer is used it will be loaded in the proper way with a null detector and bucking voltage to the secondary.

(7) ACKNOWLEDGMENTS

The authors extend thanks to O. L. Patty, G. H. Fentress, H. L. Binkly, D. R. Belsher and W. A. Downing for constructing and testing several models of the proposed apparatus. Thanks are also extended to the Microwave Impedance Standards Project for matching the microwave ports, and to Microwave Low Power Standards Project for measurements of relative attenuation.

(8) REFERENCES

(1) Ginzton, E. L.: 'Microwave Measurements' (McGraw-Hill, 1957).

(2) Engen, G. F., and Beatty, R. W.: 'Microwave Attenuation Measurements with Accuracies from 0·000 1 to 0·06 Decibel over a Range of 0·01 to 50 Decibels', *NBS Journal of Research*, Section C, 1960, **64** C, No. 2, p. 139.

(3) 'Comparison of Two Techniques for Measuring Microwave Attenuation', *NBS Technical News Bulletin*, November, 1960, p. 192.

(4) Gainsborough, G. F.: 'A Method of Calibrating Standard Signal Generators and Radio-Frequency Attenuators', *Journal I.E.E.*, 1947, **94**, Part III, p. 203.

UNWIN BROTHERS LIMITED, WOKING AND LONDON

JOURNAL OF RESEARCH of the Notional Bureou of Standards—C. Engineering and Instrumentotion
Vol. 71C, No. 2, April–June 1967

A 2:1 Ratio Inductive Voltage Divider With Less Than 0.1 PPM Error to 1 MHz

Cletus A. Hoer* and Walter L. Smith*

Institute for Basic Standards, National Bureau of Standards, Boulder, Colo. 80302

(October 25, 1966)

A simple ratio transformer having a 2:1 ratio of input voltage to output voltage can be made with a ratio error less than 0.1 ppm from 1 kHz to above 1 MHz. Applications and sources of error are discussed. Experimental results leading to an optimum transformer design are given. A bridge to measure the ratio error to 0.025 ppm from 1 kHz to 1 MHz is also described.

Key Words: Attenuator, bridge, high frequency, inductive voltage divider, ratio transformer.

1. Introduction

Techniques for winding inductive voltage dividers with adjustable ratios have been improved during the last few years to where dividers having errors less than 0.1 ppm at frequencies around 1 kHz are not too difficult to make. Unfortunately, as the frequency increases, the ratio error increases rapidly so that this accuracy cannot be maintained above about 10 kHz. At frequencies above 10 kHz the most accurate ratio device is an inductive voltage divider which has a fixed voltage ratio of 2:1: that is, one dividing an input voltage into two equal parts such as is shown in figure 1. Such dividers have recently been made with errors less than 0.1 ppm from 1 kHz to above 1 MHz. This paper summarizes the research that was done to develop this accurate high frequency divider.

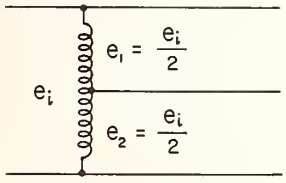


FIGURE 1. *A 2:1 ratio inductive voltage divider.*

2. Applications

A few of the many possible applications for this type of divider are shown in figure 2. Figure 2a is a bridge for comparing two impedances, z_1 and z_2.

Figure 2b shows n basic 2:1 ratio dividers connected to give a voltage ratio

$$\frac{V_{out}}{V_{in}} = \frac{1}{2^n}.$$

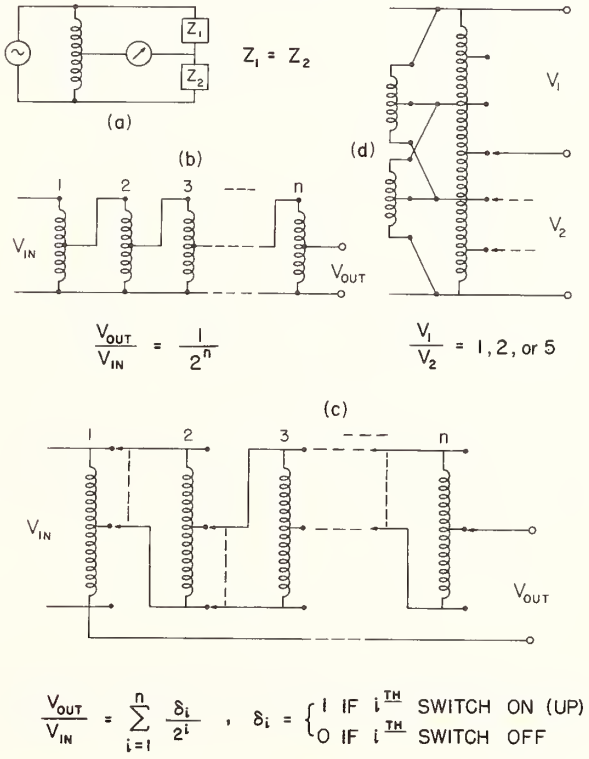


FIGURE 2. *Some possible applications for 2:1 ratio dividers.*

*Radio Standards Engineering Division, NBS Laboratories, Boulder, Colo. 80302.

385-582 O - 70 - 18

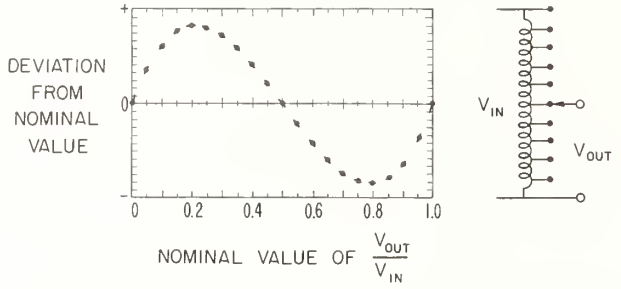


FIGURE 3. *Typical error curve for one decade of a low-frequency decade voltage divider.*

Each section of this divider accurately divides its input voltage in half to give an attenuation of 6 dB.[1] An attenuator made with n sections will have approximately $6n$dB.

Figure 2c shows an adjustable ratio "binary" inductive voltage divider. The principle of operation is the same as for low frequency decade voltage dividers except that each section divides its input voltage into only two equal parts instead of 10 equal parts. T. L. Zapf et al., have shown that the error curve for any one isolated decade of a low frequency decade voltage divider has an "S" shape such as shown in figure 3 [1].[2] As the frequency increases, the magnitude of the error increases but the curve still tends to go through zero at a ratio of 0.5. It is this error curve that suggested that 2 : 1 ratio dividers would be accurate at high frequencies as well as low frequencies. If n of these 2 : 1 ratio dividers are connected as shown in figure 2c the voltage ratio is

$$\frac{V_{out}}{V_{in}}=\sum_{i=1}^{n}\frac{\delta_i}{2^i},\quad \delta_i=\begin{cases}1 \text{ if the } i\text{th switch is up}\\ 0 \text{ if the } i\text{th switch is down.}\end{cases}$$

The input of each section is connected either across the top half or across the bottom half of the preceeding section. To obtain the same resolution with a binary divider as with a decade divider requires that the binary divider have approximately 3 times as many sections as the decade divider has.

Figure 2d shows three basic 2 : 1 ratio dividers connected in series with two more 2 : 1 ratio dividers connected so as to assure that the same voltage is applied to each of the three in series. Any number could be connected in series, however the combination shown gives the very useful ratios

$$\frac{V_1}{V_2}=1,\ 2,\ \text{and}\ 5.$$

This type of divider could be used in a bridge such as shown in figure 2a to compare standards of impedance which are often built to have nominal values in a 1-2-5 sequence.

[1] More accurately 20 log 2 = 6.020600 dB.
[2] Figures in brackets indicate the literature references at the end of this paper.

3. The 2 : 1 Ratio Divider

The 2 : 1 ratio divider is a center-tapped transformer that divides a voltage into two equal parts. It is convenient to define the complex ratio error, $\rho=\rho_r+j\rho_i$, as the error in the ratio of the output voltage to the input voltage. The output voltages e_1 and e_2 shown in figure 1 then become

$$e_1=\frac{e_i}{2}(1+\rho),\qquad e_2=\frac{e_i}{2}(1-\rho).$$

If $|\rho|\ll 1$, the ratio e_1/e_2 is approximately

$$\frac{e_1}{e_2}=1+2\rho. \tag{1}$$

To get an expression for the ratio error, consider the divider as two impedances in series which differ by a small amount ΔZ as shown in figure 4. If no current is drawn from the center tap, the voltage ratio e_1/e_2 is

$$\frac{e_1}{e_2}=\frac{Z+\Delta Z}{Z}=1+\frac{\Delta Z}{Z}. \tag{2}$$

From (1) and (2) we get

$$\rho=\frac{\Delta Z}{2Z}=\frac{\Delta R+j\omega\Delta L}{R+j\omega L} \tag{3}$$

where $R+j\omega L$ is the open circuit input impedance, Z_{in}, of the transformer. The real and imaginary components of ρ obtained from (3) are

$$\rho_r=\frac{R\Delta R+\omega^2L\Delta L}{R^2+\omega^2L^2} \tag{4}$$

$$\rho_i=\frac{\omega R\Delta L-\omega L\Delta R}{R^2+\omega^2L^2}. \tag{5}$$

If $Q^2\gg 1$ as is often the case, where $Q\equiv\omega L/R$, (4) and (5) reduce to

$$\rho_r=\frac{\Delta L}{L}+\frac{\Delta R}{\omega LQ} \tag{6}$$

$$\rho_i=-\frac{\Delta R}{\omega L}+\frac{\Delta L}{LQ}. \tag{7}$$

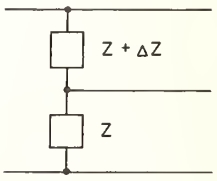


FIGURE 4. *Simplified circuit of a 2:1 ratio divider.*

These equations show the importance of winding the transformer so that it has a large inductance. The value of $\Delta L/L$ is made small by winding the transformer on a high permeability core to increase L, and by twisting the wires to minimize ΔL. A twisted cable is made by folding one wire in half and twisting the pair fairly tightly with a hand drill as shown in figure 5a. This cable is wound on the magnetic core and connected as shown in figure 5b. The wires are connected so that the current path from the upper tap to the center tap is the same as the current path from the center tap to the bottom tap. Thus, both windings link very nearly the same magnetic flux and have a small difference in inductance ΔL. Twisting the wires helps assure that both windings link the same flux.

Figure 6 shows the variation of ρ with frequency for a transformer having 42 turns of twisted No. 26 magnet wire wound on a ferrite torodial core. Note that ρ_r is essentially constant with frequency up to 2 MHz which is the upper frequency limit of our bridge. To see how high in frequency ρ_r will remain constant, we need to look at the equivalent circuit of the divider.

Figure 7 shows one of two exact equivalent circuits for an autotransformer [2] which is what the 2:1 ratio divider is. The Y_{oc} is the open-circuit admittance measured across the input terminals and Z_{sc} is the short-circuit "leakage" impedance measured across either output with the input terminals short-circuited. It can be seen from figure 7 that the voltage ratio V_1/V_2 at the output terminals will be equal to e_1/e_2 only if no current flows through Z_{sc}. If within the transformer housing the admittance from A to B differs from that from B to D by an amount ΔY_t, and if $|Z_{sc}\Delta Y_t| \ll 1$, then

$$\frac{V_1}{V_2} = \frac{e_1}{e_2} - 2Z_{sc}\Delta Y_t. \tag{8}$$

If ΔY_t is mainly a difference in capacitance, ΔC_t,

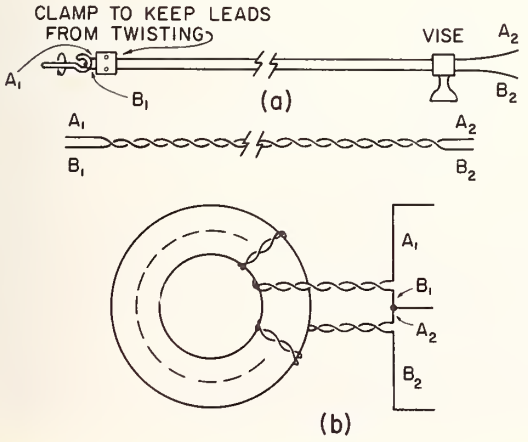


FIGURE 5. *The wires are twisted, wound on the core and connected as shown here to make a 2:1 ratio divider.*

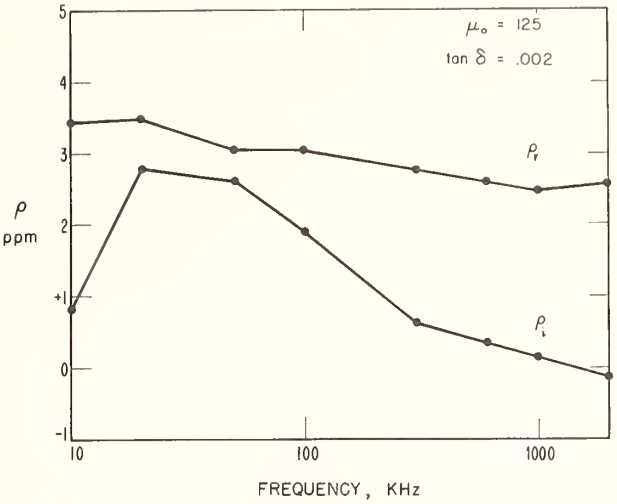


FIGURE 6. *Error curve for one of the first experimental dividers showing how ρ_r and ρ_t are relatively constant with frequency.*

$$\frac{V_1}{V_2} = 1 + 2\rho + 2\omega^2 L_{sc}\Delta C_t - j2R_{sc}\omega\Delta C_t \tag{9}$$

where L_{sc} is the leakage inductance defined by $Z_{sc} \equiv R_{sc} + j\omega L_{sc}$. A typical value of L_{sc}, such as for the transformer whose ratio error is shown in figure 6, is $L_{sc} = 0.1$ μH. By using trimmer capacitors from A to B and B to D, ΔC_t can be made less than 0.01 pF giving, in ppm,

$$2\omega^2 L_{sc}\Delta C_t < 0.1\, f^2_{\mathrm{MHz}}$$

and

$$2R_{sc}\omega\Delta C_t < 0.01\, f_{\mathrm{MHz}}$$

where f_{MHz} is the frequency in MHz. Since this loading error is proportional to the frequency squared, it will be the predominant error term at frequencies much above 1 MHz. The transformer whose characteristic is shown in figure 6 has trimmer capacitors across the output leads to adjust ΔC_t to a minimum. Without adjusting ΔC_t to a minimum the ρ_r curve increases to a value of about 5 ppm at 2 MHz.

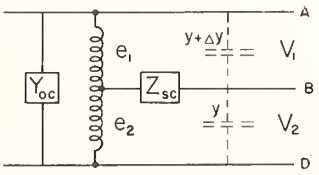


FIGURE 7. *An "exact" equivalent circuit for an autotransformer with lead admittances* Y *and* Y + ΔY *added.*

4. A Bridge to Measure ρ

A simplified diagram of the bridge used to measure ρ is shown in figure 8. The two arms of the transformer are balanced against two capacitors C', each having a nominal value of 1000 pF. The bridge is balanced with the capacitance C and the conductance G. The transformer is then removed from the bridge, rotated so that D is at the top and A is at the bottom, and then replaced on the bridge. The bridge is again balanced as before, changing C an amount ΔC and G an amount ΔG. Solving the bridge equations gives

$$\rho_r = \frac{\Delta C}{4C'} \tag{10}$$

$$\rho_i = \frac{-\Delta G}{4\omega C'} \tag{11}$$

where it has been assumed that $|\rho| \ll 1$, $C'' \ll C'$, and $C''^2 \ll \omega^2 C'^2$.

A more detailed diagram of the bridge is shown in figure 9a. The capacitors C and C' are three-terminal capacitors having admittances to ground shown as dotted lines. If these admittances are unequal they will load the transformer causing currents to flow through Z_{sc} and, thereby, change the apparent voltage ratio of the transformer. To measure the true voltage ratio e_1/e_2 of the transformer, a second detector, D_2, is placed between the center tap of the transformer and ground in order to balance the admittances to ground with Y. With both detectors nulled for both positions of the transformer, the bridge equations yield (10) and (11) as before.

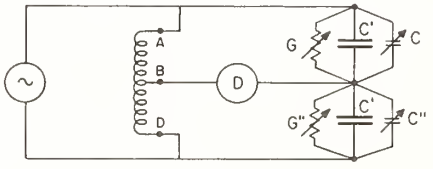


FIGURE 8. *Simplified circuit of a bridge to measure the ratio error of a 2:1 ratio divider.*

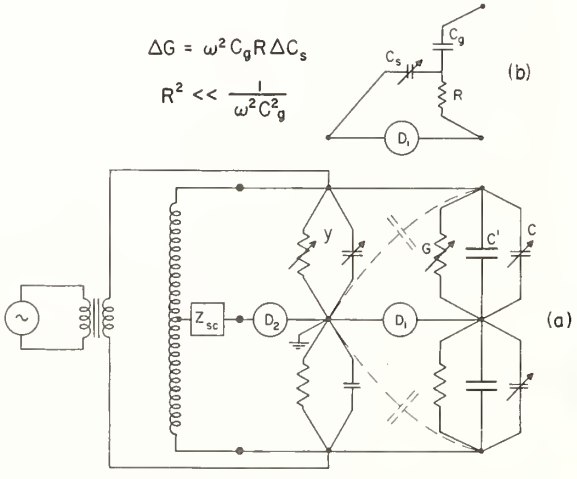


FIGURE 9. *A more detailed drawing of the circuit shown in figure 8.*

To measure the small changes in conductance, ΔG, the T-network shown in figure 9b was used for G. Changing C_s an amount ΔC_s gives a change in conductance

$$\Delta G = \omega^2 C_g R \Delta C_s \tag{12}$$

and a change in capacitance which is negligible compared to ΔC. Putting (12) into (11) gives

$$\rho_i = -\frac{RC_g}{4C'}\,\omega \Delta C_s. \tag{13}$$

The resolution for ρ_r and ρ_i is 0.025 ppm from 1 kHz to 1 MHz and 0.25 ppm to about 3 MHz. A photograph of the bridge is shown in figure 10. Bridges of similar design are described in references [3] and [4].

5. Optimum Transformer Design

Using the bridge shown in figure 10, a number of experiments were made to determine the optimum coil geometry, core material, type of wire, wire sizes, and amount of twist of the wire.

5.1. Coil Geometry

Theoretically the best coil geometry would be that which gives a maximum inductance for a minimum

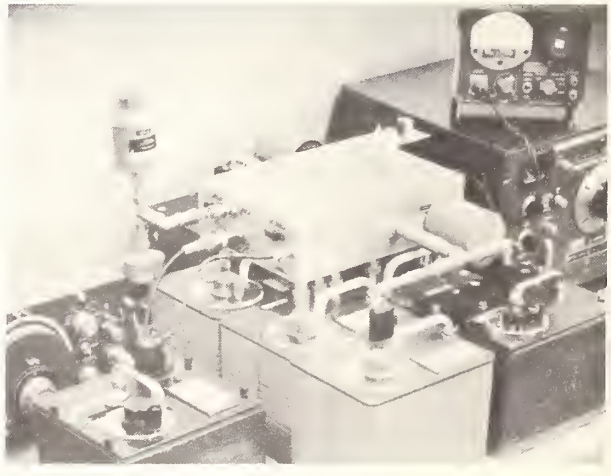


FIGURE 10. *Photograph of bridge components, showing a divider, with its shield removed, connected at the top of the bridge.*

length of cable. Large inductance is needed to decrease ρ_r, and minimum cable length is needed to make L_{sc} small. The leakage inductance per unit length of cable is given approximately by

$$\frac{L_{sc}}{l}=0.2\,\frac{P}{d}\tan^{-1}\frac{d}{P}\ \mu\text{H/m} \tag{14}$$

where

$P=2t+\delta$

$t=$ wire insulation thickness

$d=$ wire diameter

$\delta=$ skin depth $\left(\leq\frac{d}{2}\right)$

$l=$ cable length.

If $\delta > d/2$, use $d/2$ in place of δ. Equation (14) is actually the inductance per unit length of two long parallel thin tapes of width d spaced a distance P apart. By defining P and d as stated under (14), the results from (14) agree with measured data to within 10 percent which is sufficiently accurate for most applications. Note that L_{sc} is independent of the core permeability. To decrease L_{sc} it is necessary to decrease the cable length. Changing the wire diameter or insulation thickness has only a secondary affect on L_{sc}. Measured values of L_{sc} and R_{sc} from 100 kHz to 2 MHz are shown in figure 11 for No. 24 wire. The leakage inductance decreases slightly due to skin effect. The resistance R_{sc} increases slightly faster than the square root of frequency primarily due to skin effect.

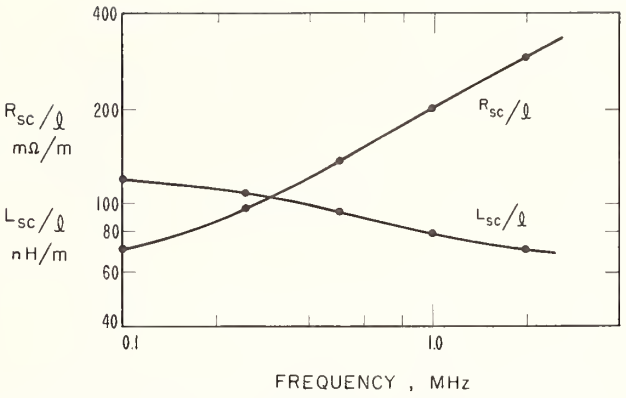


FIGURE 11. *Variation of the leakage inductance and resistance with frequency.*

5.2. Core Material

Figure 12 shows how ρ_r is decreased by increasing the permeability μ. These four transformers were identical except for the type of core material used. One long twisted cable was made and cut in half. One half was used to wind No. 10 and the other half to wind No. 11. Transformers Nos. 12 and 13 were, likewise, made from one cable cut in half. The results show that ρ_r is inversely proportional to μ. The cores were all very low-loss ferrite toroids with loss tangents of about 0.002. Cores which are very lossy do not give the same results. The Q for transformers with lossy cores is so low that the second term on the right of (6) may become larger than the first. The second term is a function of the core loss as well as the permeability. The core material, therefore, should have a high μ but not be too lossy. The core material can be fairly lossy if ΔR is small enough to make the second term on the right of (6) negligible.

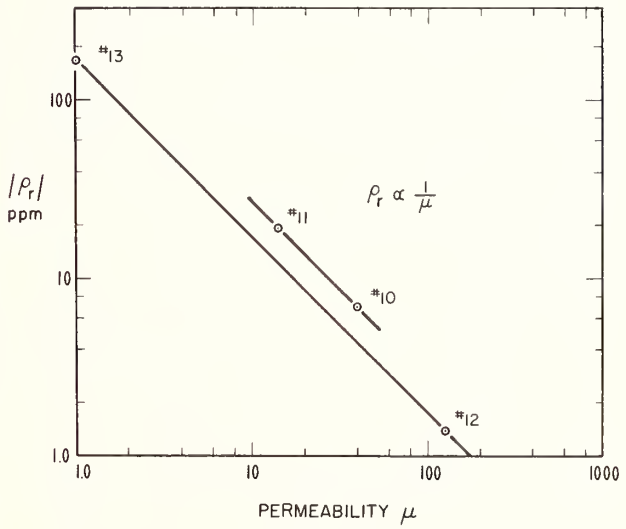


FIGURE 12. *Ratio error versus core permeability for four transformers made identical except for core permeability.*

5.3. Type of Wire

A number of experiments with different types of wire led to the conclusion that wire which has a very uniform resistivity and diameter gives the lowest ΔR values. Figure 13 shows the errors for four transformers which are identical except for the type of wire used. These transformers have air cores to eliminate the effect of magnetic core properties. Another reason for using air core transformers is that the errors are larger and easier to measure. The shape of the single layer solenoid type winding is shown in figure 14. The different types of wire used were formvar coated magnet wire (C), TEFLON® coated magnet wire (A1 & A3), and copper thermocouple wire (T). Note that the thermocouple wire ρ_r curve is constant with frequency and quite small for an air core. This means that the second term in (6) is negligible even though the Q is fairly low. Using thermocouple wire we were consistently able to wind transformers with low errors which are constant with frequency.

The thermocouple wire transformers are true "inductive" voltage dividers. The other transformers are more nearly "resistive" dividers than "inductive" dividers. The inductance of each of these four transformers is about 15 μH. Since $\rho_r \doteq 10$ ppm for the thermocouple wire transformer, (6) gives $\Delta L \doteq 150$ pH. Typical values of ΔL are 100 to 500 pH. They are constant with frequency and independent of μ. Expressing L as $L = \mu N^2 L_a$, (6) becomes

$$\rho_r \doteq \frac{\Delta L}{\mu N^2 L_a}$$

where N = number of turns on the core and L_a is the geometric inductance of the core. Typical values of ΔR are 50 to 500 $\mu\Omega$ for thermocouple wire and 10 times that for the other wire we tried. These figures are all for No. 24 wire.

It should be noted that any wire, including thermocouple wire, generally gives larger errors if the wire is stretched. A number of transformers were made with wire that was stretched approximately 1 percent to remove slight kinks in the wire. The errors were significantly larger than for transformers made from the same spool of wire with the slight kinks left in the wire. It is best to purchase wire wound on large spools so that it is relatively free of kinks when removed from the spool.

5.4. Wire Size

The experiment described above was repeated with No. 20 wire and also with No. 30 wire. The larger wire transformers gave slightly lower ρ_i errors than the corresponding No. 24 wire transformers gave. There was no significant decrease in the ρ_r errors. The results are shown in figure 15. Transformers made with No. 30 wire had considerably higher errors than those made with No. 24 or No. 20 wire.

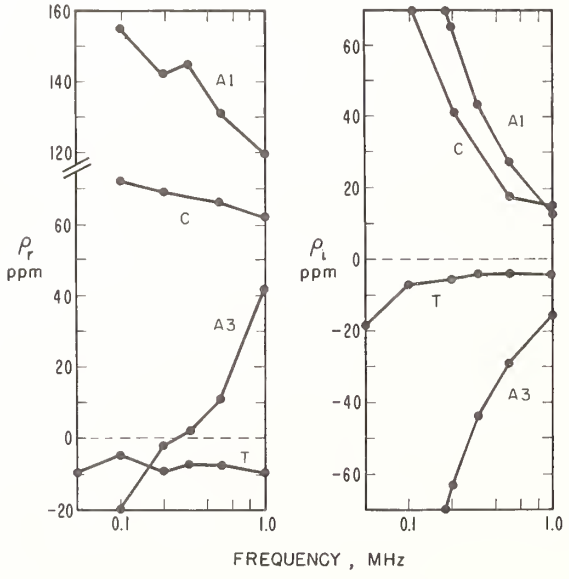


FIGURE 13. *Comparing results from four different types of No. 24 wire, where A1 and A3 are TEFLON ® coated magnet wire with single (0.0006 in) and triple (0.0015 in) insulation thickness; C is formvar coated magnet wire, and T is enamel coated copper thermocouple wire.*

5.5. Cable Twist

To determine how much the wires should be twisted, a number of transformers were made identical except for the amount the wires were twisted. Figure 16a shows ρ_r and ρ_i plotted against the number of turns per unit length of the cable. The length, l_T of one turn is defined in figure 16b. These are all air core transformers wound with No. 24 thermocouple wire. The dots are single layer solenoid windings whereas the X's are multilayer coils wound on nylon bobbins. The inductance, L, of each transformer is approximately 15 μH. As can be seen from figure 16, any twist from four to ten turns/inch (160 to 400 turns/meter) gives about the same results. Below four turns/inch the error increases. At about 12 turns/inch (500 turns/meter) the wire breaks. In terms of the twist length, l_T, and wire diameter, d, satisfactory results were obtained in the range

$$0.08 < \frac{d}{l_T} < 0.2.$$

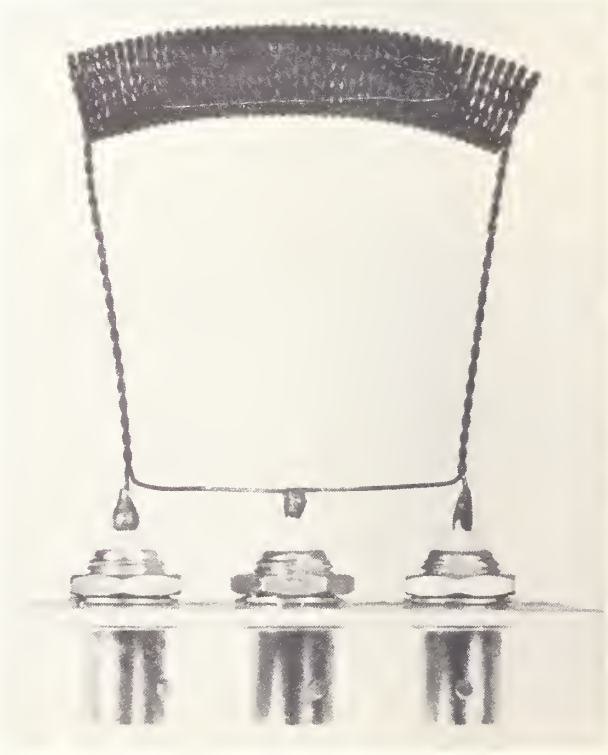


FIGURE 14. *An air core, solenoid type winding.*

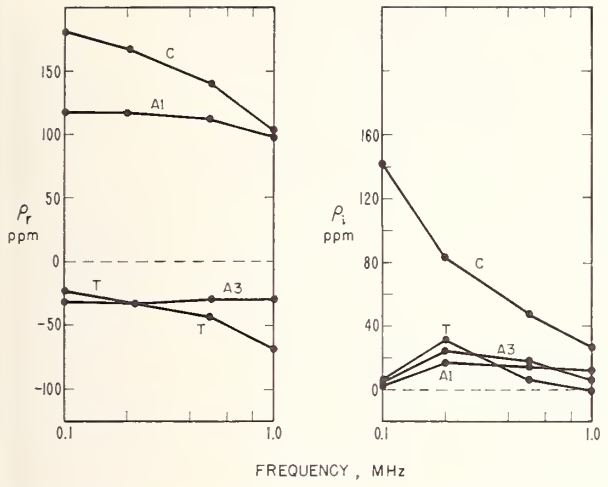


FIGURE 15. *Comparing results from four different types of No. 20 wire where A1, A3, C, and T are as defined in figure 13.*

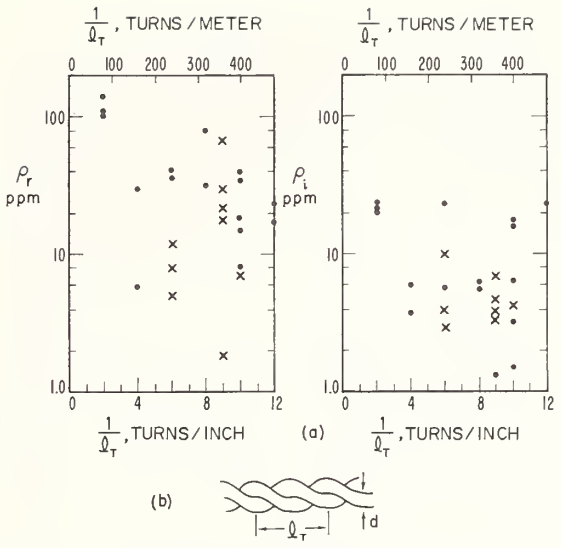


FIGURE 16. *Results of air core solenoid type (dots) and bobbin type (X's) transformers wound with cables twisted a different number of turns per inch.*

This equation should also be a reasonable guide for twisting wires other than size No. 24.

5.6. Lead Connections

One more important consideration in winding the transformers is how the leads are connected. The first transformers that we wound were connected as shown in figure 17a. The ratio errors were quite unstable and very sensitive to slight movement of the leads. To find a better way of connecting the leads, a number of air core transformers were made identical except for the lead connections. Figure 17 shows the three different ways the leads were connected. Of these three ways, the one shown in figure 17c was most stable and gave the smallest errors. The improvement in stability is probably due to the smaller mutual inductance and capacitance between leads. The lower error is probably a result of the ease with which the corresponding leads can be cut to the same length. It is important that in any arrangement of the leads that the leads on one side of the center tap be the same length as the corresponding leads on the other side. A difference in lead length of 0.001 in will give a ΔL of about 40 pH for No. 24 wire.

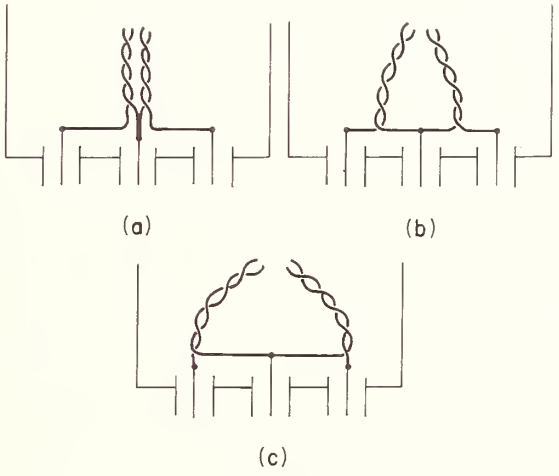


FIGURE 17. *Three different arrangements of the transformer leads.*

5.7. Cup-Core Design

One of the easiest ways of constructing a transformer with magnetic cores is to wind the cable on a bobbin and then clamp magnetic cup cores to the bobbin such as shown in figure 18. Figures 19 and 20 show the errors for three cup core transformers constructed like the one shown in figure 18. These transformers were made with all the optimum parameters that were found. The cables were made with No. 24 thermocouple wire twisted to six turns/inch (240 turns/meter). The cores are high permeability, fairly low-loss ferrite cores. The increase in ρ_r and ρ_i at frequencies above 100 kHz is due to the capacitance between leads not being equal. If these capacitances were balanced as discussed in part 3, the errors would be less than 1 ppm to above 1 MHz.

With care, the error can be made less than 0.1 ppm. A fourth cup core transformer was adjusted to have a ρ_r error less than 0.1 ppm at 100 kHz by changing the length of one lead slightly. Then ρ_r was adjusted to less than 0.1 ppm at 1 MHz by adjusting small trimmer capacitors placed between the leads. As figure 21 shows, the resulting ρ_r error is less than 0.1 ppm from 30 kHz to above 1 MHz.

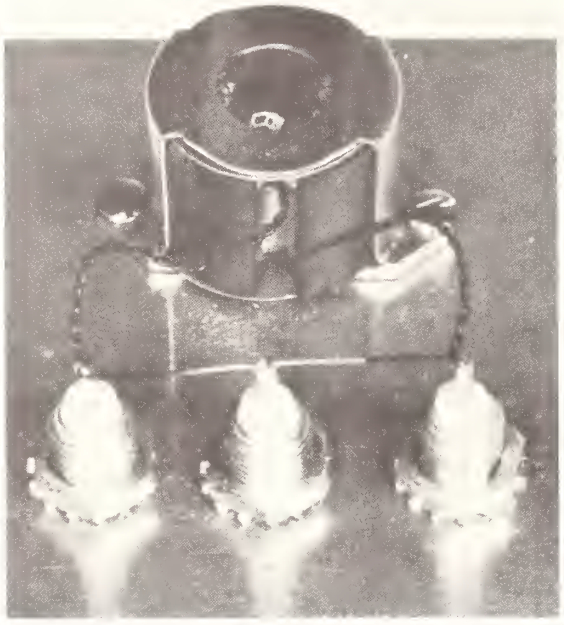


FIGURE 18. *A cup-core transformer having very low ratio errors like that shown in figures 19 and 20.*
The cover of the shield is not shown.

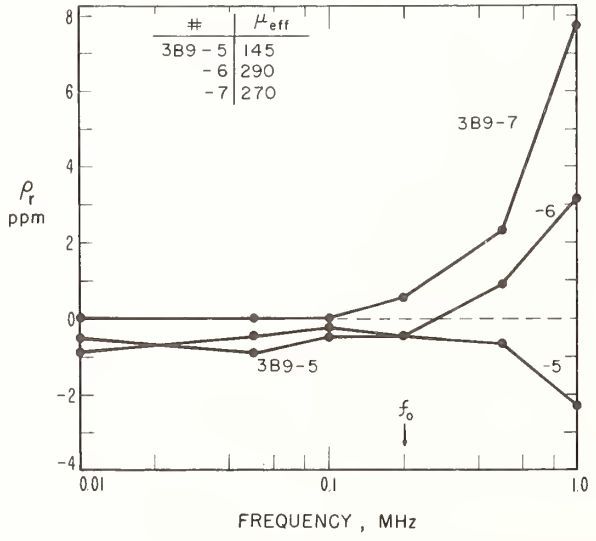


FIGURE 19. *Results from three cup-core transformers wound with No. 24 copper thermocouple wire twisted to six turns/inch.*

5.8. Tape-Core Transformers

Very broad band 2:1 ratio dividers can be made with high permeability magnetic tape cores. Figure 22 shows the ratio error of a divider wound on a supermalloy tape core having a tape thickness of 1 mil. The

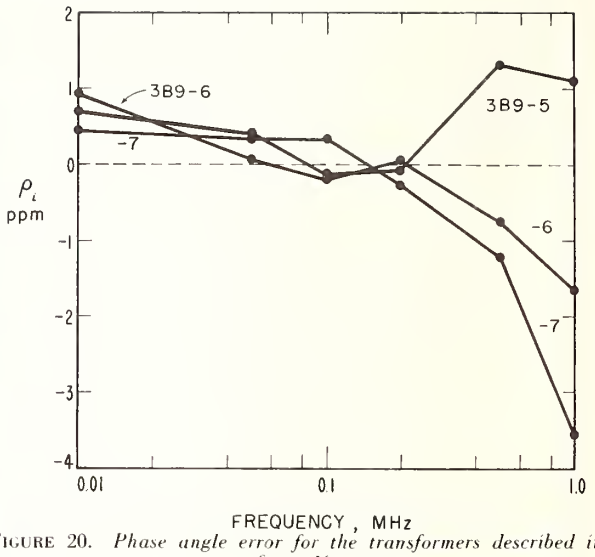


FIGURE 20. *Phase angle error for the transformers described in figure 19.*

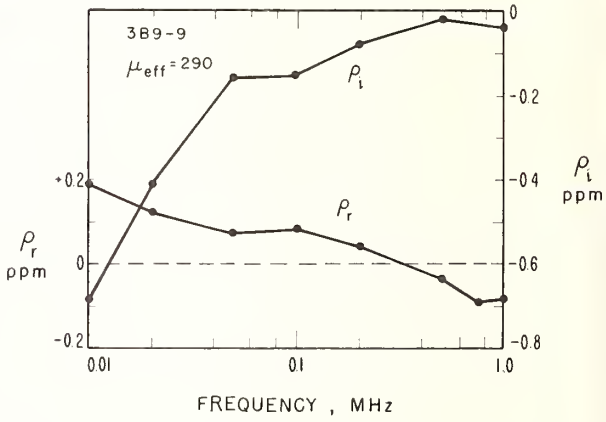


FIGURE 21. *Ratio error of a cup core transformer which has been adjusted to have a minimum ρ_r by adjusting lead length and by balancing the capacitance between leads.*

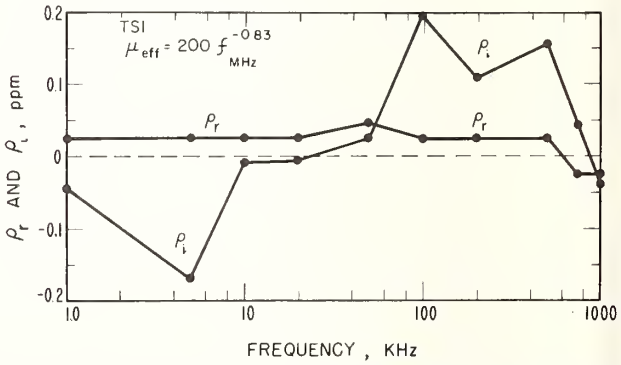


FIGURE 22. *Ratio error of a transformer wound on a supermalloy 1 mil tape core having dimensions OD=1.150, ID=0.650 and H=0.375 in.*

core has 64 turns of No. 24 copper thermocouple wire twisted to about 8 turns/inch (320 turns/meter). The capacitance between leads has been adjusted to be equal. The ρ_r error is less than 5 parts in 10^8 from 1 kHz to 1 MHz.

6. Conclusion

The 2:1 ratio transformer is a simple, yet accurate, divider which is easy to construct. Using the optimum parameters discussed in part 5, it is easy to keep the ratio error less than 1 ppm from 1 kHz to 1 MHz. Air core transformers having no magnetic cores can be made with errors less than 10 ppm over a reduced frequency range. With extra care, ρ_r can be made less than 0.1 ppm to above 1 MHz. The ratio errors are quite insensitive to temperature, time, shock, and most any change in the environment since both windings of the transformer are influenced equally by any change in the environment.

Five cup-core transformers such as described in part 5.7 have been combined as shown in figure 2b to make an accurate 30 dB attenuator at 1 MHz. Details of this type of attenuator together with a complete error analysis will be given in a future publication.

7. References

[1] Zapf, T. L., C. H. Chinburg, and N. K. Wolf (September 1963), Inductive voltage dividers with calculable relative corrections, IEEE Trans. on Instrumentation and Measurement **IM-12**, 80–85.

[2] EE Staff, MIT (1943), Magnetic circuits and transformers (John Wiley & Sons, Inc., New York, N.Y.), 400.

[3] McGregor, M. C., J. F. Hersh, R. D. Cutkosky, F. K. Harris, and F. R. Kotter (Dec. 1958). New apparatus at the National Bureau of Standards for absolute capacitance measurement, IRE Trans. on Instrumentation **I-7**, No. 3 and 4, 253–261.

[4] Cutkosky, R. D., and J. Q. Shields (September 1960), The precision measurement of transformer ratios, IRE Trans. on Instrumentation **I-9**, No. 2, 243–250.

(Paper 71C2–249)

A Precision RF Attenuation Calibration System*

C. M. ALLRED†, MEMBER, IRE, AND C. C. COOK†, MEMBER, IRE

Introduction

THE NEED for higher accuracy in attenuation measurements has kept pace with the advances in electronics. In an effort to meet the demands in this field, a precise and accurate system for the measurement of attenuation has been developed.

The system discussed operates at 30 Mc, but similar systems under construction are designed to operate at other frequencies. Choosing a single frequency of operation allows designs that increase the stability, extend the measuring range, lessen the problems of leakage, and enhance the over-all accuracy.

The System

Description of the System

Achieving a system with sensitivities of 0.001 db, accuracies of a few thousandths of a db, and a range exceeding 120 db requires careful design and special techniques. With such large ranges of attenuation to be measured, the monitor should be capable of voltage gains of 10^8 or more. Because of the stability problem of such a monitor and the desired sensitivity of 0.001 db, it was decided to operate in a two-channel system and use a null response. As is well known, the use of two-channel techniques greatly lessens the problems of level instability of the source and gain instability of the monitor. High sensitivity is also inherent in such systems. One may consider the increased complexity of the system and the necessity of making adjustment in both phase and magnitude as payment for the advantages. The block diagram of the system is shown in Fig. 1.

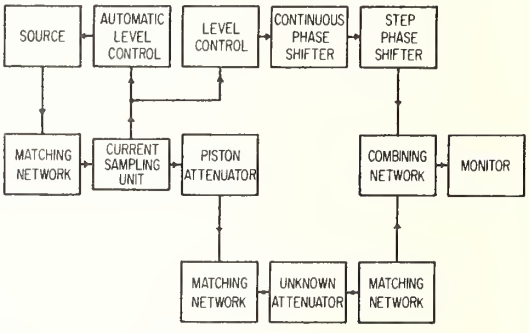


Fig. 1—Block diagram for attenuation measuring system.

The RF source is a 30-Mc crystal-controlled transmitter capable of 200 watts output with low residual modulation. The low resistive component of the piston attenuator launching coil is coupled to the 50-ohm source by a special impedance-matching network. The

* Received by the PGI, June 24, 1960. Presented at the 1960 Conference on Standards and Electronic Measurements as paper 6-2.
† National Bureau of Standards, Boulder, Colo.

current in the launching coil energizes the channel containing the phase shifter and also supplies the input to an automatic level control network connected to the source. This network is used to keep the current in the launching coil constant. The continuous precision phase shifter having low incidental level changes connects to a step adjustable phase shifter. The attenuator to be tested is inserted between matching networks in the channel containing the piston attenuator. The two channels are brought together in a special combining unit which presents correct impedances to the respective channels, minimizes interaction between the channels and has a low insertion loss. The system is completed by a high-gain monitor.

Measurements are made by adjusting both the standard attenuator and phase shifter until a null response is obtained. This is done with and without the attenuator under calibration in the system. The change in the standard attenuator and phase shifter readings gives the value of attenuation and phase shift of the unknown.

In its present form, the phase shifter has no accurate displacement measuring device and no attempt will be made to evaluate the phase measuring characteristics of the system.

Front and rear views of the complete system mounted in a console are shown in Figs. 2 and 3, respectively.

Fig. 2—Front view of attenuation measuring system.

Fig. 3—Rear view of attenuation measuring system.

Standard Attenuator

Properties: Piston attenuators are used quite extensively for precision attenuation measurements.[1-3] The major advantage of such attenuators is that the attenuation is determined (except for secondary effects) by the fundamental units, length and time (frequency). Another advantage is that close impedance matching is usually not a stringent requirement.

The attenuator has its disadvantages. It is essentially a nondissipative instrument and reflects varying amounts of energy rather than absorbing it as the attenuation is varied. As a result, the minimum insertion loss is quite high for linear operation and the terminal impedances are reactive. Usually, frequency sensitive networks are used for impedance matching. An additional problem occurs because of the permeability and finite conductivity of the material used for the waveguide. These cause a second-order effect on the attenuation constant. This requires a determination of the permeability and conductivity of the material used in the guide wall. Since the attenuation constant α is a slowly varying function of frequency, a single scale cannot be used over wide frequency ranges without correction. Only output signal ratios are measured; the relationship between input and output, if desired, must be determined by other means.

Theory: Fields existing within a uniform perfectly-conducting waveguide can be expressed as a linear combination of terms called modes. The field components for each mode are proportional to the quantity, $e^{-\gamma z+j\omega t}$, where the propagation constant γ is purely imaginary and is determined entirely by the guide dimensions and frequency.

Below a certain frequency, called the cutoff frequency, γ becomes purely real and the fields decay exponentially at a rate dependent upon the mode. To in-

[1] D. E. Harnett and N. P. Chase, "The design and testing of multirange receivers," Proc. IRE, vol. 23, pp. 578–593; June, 1935.
[2] R. E. Grantham and J. J. Freeman, "A standard of attenuation for microwave measurements," *Trans. AIEE*, vol. 67, pp. 535–537; June, 1948.
[3] C. G. Montgomery, "Technique of Microwave Measurements," M.I.T. Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York. N. Y., vol. 11, ch. 11, pp. 679–719; 1947.

sure mode purity at high attenuation levels, the mode with the smallest decay rate is commonly used.

In the actual case, the permeability and finite conductivity of the guide walls and the properties of the medium within the guide add a small imaginary and real part to the above propagation constant.

A good approximate equation for the propagation constant of the TE_{11} mode is given by:[4,5]

$$\gamma = \frac{p_{11}}{a}\sqrt{1-\left(\frac{2\pi af}{p_{11}C}\right)^2} - \frac{1}{a\sqrt{\pi\mu f\sigma}} \cdot \left[1 + j\left(1 - \frac{1}{1 - \frac{1}{a\left[1-\left(\frac{2\pi af}{p_{11}C}\right)^2\right]\sqrt{\pi\mu f\sigma}}}\right)\right]$$

where:

a = physical radius of guide

p_{11} = 1.8411838, a constant which is the first root of the first derivative of the first-order Bessel Function of the first kind

f = frequency

C = velocity of light in an unbounded medium having the electrical properties equal to that within the guide

μ = permeability of guide wall

MKS system of units are used throughout. With the operating frequency so far below the cutoff frequency, neglecting the effect of even moist air within the guide as compared to vacuum gives rise to an error less than 1 part in 10^7. (C for vacuum = 299793.0 ± 0.3 km/sec.[6])

Circular Waveguide: In the system described, circular waveguide was used with the excited field consisting of the TE_{11} mode. The displacement of the receiving coil relative to the launching coil is measured by a ruled scale and optical projector. The scale is accurate to 0.0001 inch in any 6 inches.

The guide is composed of brass and has an outside diameter of 3.9 inches and a length of 22.3 inches. The internal diameter of the guide was chosen so as to give an attenuation rate of 10 db per inch. This diameter was measured at 2-inch intervals along the guide. The average diameter was 3.19725 inches with a maximum variation of 0.00003 inch. The accuracy of the measurements are within ±50 millionths of an inch. These measurements were made at 68°F while the system is operated about 4°F above this temperature. The thermal coefficients of expansion of the stainless steel scale and the brass guide are approximately the same. If they were equal, temperature variations would have negligible effect. Assuming a conservative 5 parts per million (20 per cent) difference in their coefficients, a calculable error of 0.0001 db per 10 db could arise for the 4°F temperature variation. An uncertainty of 0.0001 inch in the diameter of the guide at 68°F adds an error of 0.0003 db per 10 db. The dc conductivity of the guide was found to be 1.35 mhos per meter. From Chambers and Pippard[7] and Bussey,[8] it is conservative to say that the RF conductivity does not decrease more than 10 per cent from the bulk dc value. Assuming the RF conductivity has decreased from the dc value by 5 per cent, this value of conductivity will be fairly certain to ±5 per cent. Such an uncertainty gives rise to an error of 0.00026 db per 10 db.

The inside wall of the guide has a thin coating (5 to 10 microinches) of rhodium. This provides a hard surface for the silver sliding contacts of the piston to bear upon and prevents corrosion of the brass surface. The effective conductivity of the guide surface is insignificantly altered from that of the brass itself.

Input System

The input system of a piston attenuator is important, as it plays a major role in the over-all measurement range. The high initial insertion loss inherent in piston attenuators arises from lack of mode purity and the interaction of the output on the input when the coil separation is small. What is needed is an input system that launches the single TE_{11} mode with an intensity independent of the position of the output coil, or a system that launches the desired mode with such a strong intensity that the output coil need not come close to the input coil. Both approaches are simultaneously used. Barlow and Cullen[9] reduce the interaction effect by making the magnitude of the generator and load resistive impedances equal to the characteristic impedance of the attenuator. Weinschel, Sorger, and Hedrich[10] reduce the effect by considering the entire attenuator as a mutual inductive coupled tuned bandfilter and adjust the generator impedance for minimum deviation from linearity.

A different approach is used here. If one could achieve a current sheet with a spatial distribution in the transverse plane of the guide having the same form as the familiar electric field pattern of the TE_{11} mode, the desired evanescent mode would be launched in both directions from the current sheet. Furthermore, if the current sheet were constant (infinite impedance gener-

[4] J. Brown, "Corrections to the attenuation constants of piston attenuators," *Proc. IEE*, vol. 96, pt. 3, pp. 491–495; November, 1949.

[5] C. M. Allred, "Chart for the TE_{11} mode piston attenuator," *J. Res. NBS*, vol. 48, pp. 109–110; February, 1952.

[6] J. W. M. DuMond, "Present status of precise information on the universal physical constants. Has the time arrived for their adoption to replace our present arbitrary conventional standards?" IRE Trans. on Instrumentation, vol. I-7, pp. 136–175; December, 1958.

[7] R. G. Chambers and A. B. Pippard, "The Effect of Method of Preparation on the High-Frequency Surface Resistance of Metals," Inst. of Metals, London, Eng., Monograph No. 13, pp. 281–293; 1953.

[8] H. E. Bussey, "Standards and measurements of microwave skin depth-conductivity, surface impedance, and Q," this issue, p. 171.

[9] H. M. Barlow and A. L. Cullen, "Microwave Measurements," Constable and Co., Ltd., London, Eng., pp. 384–388; 1950.

[10] B. O. Weinschel, G. U. Sorger, and A. L. Hedrich, "Relative voltmeter for VHF/UHF signal generator attenuation calibration," IRE Trans. on Instrumentation, vol. I-8, pp. 22–31; March, 1959.

ators) and independent of the fields reflected from the pickup coil, the interaction effect would cease to exist. The above conditions are approached by constructing the launching coil after the above pattern (see Fig. 4). The constant current sheet feature is approached by sampling the launching current and keeping this current constant by means of negative feedback to the RF source.

The attenuator guide has been extended beyond the launching coil for a twofold purpose. The pickup coil must produce a reflected field since it absorbs energy. This reflected field passes through the launching coil with little interaction and is terminated by the guide extension. The field excited by the launching coil in the guide extension is also terminated.

The problem of matching the low resistive component (estimated to be between 0.01 and 0.1 ohm) of the launching coil is quite severe. This is accomplished by means of a special network utilizing small ceramic capacitors in a ladder arrangement. These capacitors have a high RF voltage and current rating; this permits operation of the matching network at inputs to 200 watts. A 100-watt input produces a current of the order of 35 amperes in the launching coil.

The launching unit is constructed of small stainless steel tubing and water is passed through the unit to prevent excessive heating of the guide which would change its dimensions.

Mode filters are also used, since the reactive impedance of the launching unit gives rise to some axial electric field.

Performance of the input system is discussed later along with the over-all system performance. A detailed report of such input systems is being prepared for publication.

Fig. 4—Launching coil assembly similar to the one described.

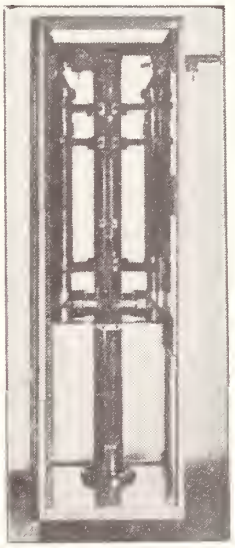


Fig. 5—Phase shifter.

Phase Shifter

A precision continuous phase shifter was needed which had level changes less than 0.001 db as the phase was varied. This placed the following restrictions on the phase shifter: the section of the phase shifter whose electrical length is varied must have a very uniform characteristic impedance, it must be terminated in this value of impedance, and the line must be sufficiently large to reduce the wall losses due to the finite conductivity of the metal.

The phase shifter is of the trombone type (see Fig. 5). This type consists of two stationary transmission lines connected together by a U-shaped transmission line which moves relative to the stationary section in a telescoping manner. Single-frequency operation permitted use of the simplified nonconstant impedance construction. The stationary sections of the phase shifter consist of precision coaxial lines which are terminated at both ends in their characteristic impedances. The moveable section is dimensionally stable but nonprecise. The undesirable level changes caused by imperfectly terminating the precision line are lessened if both ends are closely matched to the characteristic impedance of the line. A 10-db pad bracketed by impedance matching networks has been built into the center of the moving section. This provides isolation between the two halves of the trombone and the correct terminating impedances for the corresponding ends of the precision section. Additional pads and matching units are provided for terminating the other two ends.

The outer conductors of the precision section of the phase shifter are electroformed copper cylinders. A heavy silver plate with thin rhodium coating is on the inside walls. The inside diameters are 3.33 inches, outside diameters, 3.9 inches, and lengths, 22.3 inches. The center conductors are brass rods with silver plate and rhodium coating similar to that used on the outer con-

ductors. The diameters are 1.44 inches. Maximum departure of the critical diameters from the mean value is 0.00035 inch. The maximum change in characteristic impedance at any point due to these diameter variations is less than 0.04 per cent. These are calculated variations and represent pairing the worst possible combinations of diameter measurements. The averaging effect of the smooth variations in the diameters would produce a much smaller variation in characteristic impedance.

The complete phase shifter was made operational by adjusting the four terminating matching networks until input and output impedance changes were very small as the moving section was varied. These changes were less than 0.01 ohm for both real and reactive components as measured on an RF bridge having high differential sensitivity.

The constancy of output level of the phase shifter was measured using crystal diode detectors in a very sensitive differential circuit. The maximum level variation was less than 0.0005 db. This occurred for a 33° phase shift. The phase shifter level change will be significantly less than this in the case of most piston attenuators and dissipative attenuators operating where the attenuation is essentially independent of frequency since the phase shift is usually small in these attenuators. In any case, corrections can be made.

The Combining Junction

The two channels are brought together by means of a special network. This network minimizes interaction between the two channels which would otherwise require excessive padding to maintain impedance match. The network[11] is based on the bridged-T circuit. Such a network, if ideal, completely isolates the two channels from each other, presents each channel with the correct matching impedances, and would have an insertion loss of only 3 db. The actual network has the following measured characteristics:

Isolation between inputs:	110 db
Input Impedance:	
Input 1	$50.4+j0$ ohms
Input 2	$50.9+j0$ ohms
Insertion Loss:	
Input 1 to output	4 db
Input 2 to output	4 db.

The combining junction with cover removed is shown in Fig. 6.

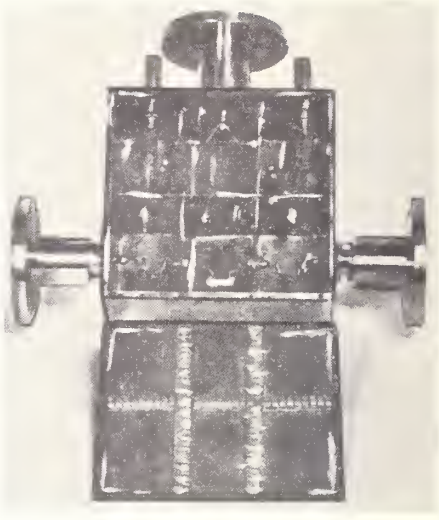


Fig. 6—Combining junction.

System Performance

The system is essentially a primary standard and therefore does not require calibration. In comparisons with other standards, agreement has been within the corresponding error limits. Tests to check the self-consistency of the system have been made by measuring a fixed value of attenuation at various initial positions of the attenuator, at different input power levels, and various initial positions of the phase shifter. The system is new and a few additions and modifications are yet to be made. The main addition will be a special low-noise stable monitor using phase detection designed to fit the system's special requirements. The present monitor is a slightly modified ordinary commercial communication receiver.

Sensitivity

The system has a sensitivity greater than 0.001 db at a level of 80 db below 1 volt output of the standard attenuator. In principle the sensitivity would increase by a factor of 10 for every 20-db increase in level. Instability of components and the fact that the signal is not strictly a pure single frequency sine wave prevent the sensitivity in a null system from completely achieving the ideal.

Stability

All components were constructed and mounted so as to give the system stability. For example, solid semi-flexible coaxial line is used throughout. The stability is indicated by the following calibration data of a commercial attenuator, Table I. The time sequence is from left to right and each row is a repeat of the attenuator settings. A lapse of about one hour occurred while data were being taken. This means that the time lapse between succcesive values in a given column represents roughly 12 minutes. All values are in db. Instabilities do occur sometimes, as indicated in the last two readings and a few components have had to be changed. Normally, quite a few readings can be taken before a drift of 0.001 db is noticed. As the difference between two readings gives the value of attenuation desired, the drift usually does not impair the accuracy significantly.

[11] C. M. Allred and C. C. Cook, "A multiple isolated-input network with common output," to be published in *J. Res. NBS.*

TABLE I

DATA SHOWING STABILITY OF SYSTEM

88.212	87.213	86.212	85.212	84.219	83.205
88.212	87.212	86.212	85.212	84.220	83.205
88.212	87.213	86.212	85.212	84.220	83.206
88.212	87.213	86.212	85.212	84.225	83.208

Linearity

The linearity of the attenuator at close coil separations is good but the point of departure from linearity (due to the interaction effect) could not be ascertained. The problem occurs because of heating effects in the pickup coil as the output level increases. This slow thermal drift occurs before any interaction effect is noticed.

The equivalent change in linearity produced by this high level thermal drift is about 0.006 db at the 3-volt level and is essentially absent at the 1-volt level.

A stable 10-db step of a dissipative attenuator was measured at various positions of the standard attenuator. The results are given in Fig. 7. The first point on the left, for example, gives the value of 9.971 for the nominal 10-db step of the dissipative attenuator when the standard attenuator is operated within the 23- to 33-db region. The 1-volt output level is at about the 33-db position. The first point deviates from the others because of the above-mentioned thermal drift in the pickup coil when operated at the 3-volt output level (23-db position). The last four points show deviations due to poor signal-to-noise ratio. When these data were taken, the temperature control system for the water flowing in the launching unit was not installed and the input end of the attenuator was a few degrees colder than the output end. A rough calculation of this effect is indicated by the dashed line and shows that, when in proper operation, the system will be more linear. There has been insufficient time to take new data.

Measuring Range

The data in Fig. 7 do not represent the range of the system, as additional padding had to be added to the attenuator channel at higher output levels (less attenuation) to permit nulling with the phase shifter leg. The last nine positions, however, represent roughly the level of the signal at the monitor. The last point, for example, is taken at 140 db below the 1-volt level. Thus the data on the graph represent a measurement range of 150 db. Each point was obtained by averaging 6 measurements. The standard deviations of the last five points were, respectively: 1.2, 1.6, 8.1, 6.2, and 30.3 thousandths of a db. The data include any nonreproducibility in the 10-db step of the dissipative attenuator.

Accuracy

The errors discussed do not include those due to mismatched impedances presented to the attenuator under calibration. There are situations where, because of high initial insertion loss, the measurement is essentially independent of impedance mismatch. Such cases occur in most piston attenuators and also in those dissipative attenuators having sufficient padding as an integral part of the attenuator.

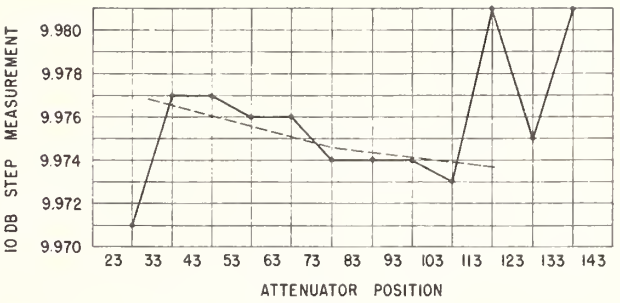


Fig. 7—Attenuator linearity test.

The total system error, which is a function of the range of measurement, A, can be separated into its parts.

Let

$$E_t = E_m + E_g + E\Phi + E_n$$

where

E_t = total system error.

E_m = error in measuring system. This can be subdivided into those errors due to the scale and those due to lack of resolution in the projection equipment.

Scale error = 0.001 db for $0 < A < 60$ db, 0.002 db for $60 < A < 120$, 0.003 for $120 < A < 180$ db.

Resolution error is somewhat less than 0.001 db for any value of A.

E_g = errors due to uncertainties in values of guide diameter, guide conductivity, and temperature effects. As indicated previously, this is well within $0.001\ A/10$ db.

E_Φ = errors due to level changes in the phase shifter. These are known sufficiently well to leave the uncertainties small compared to other errors.

E_n = errors due to the uncertainties caused by the random fluctuations of noise. If at large values of A ($A \geqq 100$ db) six measurements are averaged, the results of tests shown in Fig. 7 would indicate reasonably safe values of E_n to be:

$$\begin{aligned} E_n &= \pm 0.002 \text{ db for } 90 < A \leqq 100 \text{ db} \\ &= \pm 0.004 \text{ db for } 100 < A \leqq 110 \text{ db} \\ &= \pm 0.01 \text{ db for } 110 < A \leqq 120 \text{ db} \\ &= \pm 0.02 \text{ db for } 120 < A \leqq 130 \text{ db} \\ &= \pm 0.05 \text{ db for } 130 < A \leqq 140 \text{ db.} \end{aligned}$$

These could be reduced by averaging over a greater number of measurements. The above values of E_n are

based on the assumption that the input signal level to the unknown attenuator (attenuation values, A) is 1 volt. If this is raised to 3 volts:

$$E_n' = \pm 0.03 \text{ db for } 130 < A \leqq 140 \text{ db}$$
$$E_n' = \pm 0.06 \text{ db for } 140 < A \leqq 150 \text{ db}$$

where E_n' includes the effect of noise and the thermal drift in the pickup coil.

The combination of these errors into a total error gives:

$$E_t = \pm \left(0.002 + 0.001 \frac{A}{10}\right) \text{ db for } 0 \leqq A \leqq 60 \text{ db}$$
$$E_t = \pm \left(0.003 + 0.001 \frac{A}{10}\right) \text{ db for } 60 < A \leqq 90 \text{ db}$$
$$E_t = \pm 0.015 \text{ db for } 90 < A \leqq 100 \text{ db}$$
$$E_t = \pm 0.018 \text{ db for } 100 < A \leqq 110 \text{ db}$$
$$E_t = \pm 0.025 \text{ db for } 110 < A \leqq 120 \text{ db}$$
$$E_t = \pm 0.037 \text{ db for } 120 < A \leqq 130 \text{ db}$$
$$E_t = \pm 0.068 \text{ db for } 130 < A \leqq 140 \text{ db.}$$

Again, if the 3 volt level is used instead of the 1 volt:

$$E_t = \pm 0.048 \text{ db for } 130 < A \leqq 140 \text{ db}$$
$$E_t = \pm 0.079 \text{ db for } 140 < A \leqq 150 \text{ db.}$$

Acknowledgment

Others have helped in various phases of this work and the authors gratefully acknowledge the contributions of the following: D. H. Russell, E. C. Wolzien, R. L. Peck, R. Barkhaus, R. Brunner, M. A. Mulligan, A. L. Schuster, P. A. Nichols, R. C. Darr, W. H. Long, and P. L. London.

Reprinted from

IRE TRANSACTIONS ON
INSTRUMENTATION

Volume I-9, No. 2, Sept. 1960

6.a. **Some basic microwave phase shift equations,** Robert W. Beatty, Radio Sci. J. Res. 68D, No. 4, 349-353 (April 1964).

The phase differences between terminal variables (voltage, current, or traveling wave types) at the output with respect to similar ones at the input of a 2-port are expressed in terms of the scattering coefficients of the 2-port and of the reflection coefficients of the system into which it is inserted.

A variety of phase shifts may be defined for a given 2-port, depending upon which of the terminal variables are considered, whether or not generator and load reflection coefficients are assumed to vanish, and in which direction the 2-port has been inserted into the system. It is shown that a reasonable choice for one of the two "characteristic" phase shifts of a 2-port is ψ_{21}, the argument of S_{21}, one of the scattering coefficients. It follows that the other "characteristic" phase shift is ψ_{12}. The corresponding change in characteristic phase shift for variable phase shifters is the change in either ψ_{21} or ψ_{12} (whichever is appropriate) from the initial to the final setting.

Ideal phase shifters are discussed, and expressions for the change in output level of variable phase shifters are given. The importance of using a nonreflecting system in phase shift measurements is emphasized.

6.b. **Microwave attenuation standards and measurements,** R. W. Beatty, NBS Monograph 97, (April 1967).

A comprehensive and commentarial review of microwave attenuation measurement methods and standards is presented. In addition, a relatively new and more precise way of representing and analyzing an attenuation measurement is presented. This is turn permits more rigorous definitions and error analyses than were previously possible. Expressions for both mismatch and connector errors are presented.

The referral of microwave attenuation measurements to standards operating at lower frequencies is discussed with particular attention to the errors in the referral processes as well as the errors in the standards themselves. Standards operating at d-c, audio frequencies, and higher frequencies are included in this discussion which covers waveguide-below-cutoff attenuators and rotary vane attenuators. Desirable characteristics are listed for attenuators which are suitable for calibration, and examples of these are given.

Measurement methods are classified and described, giving greatest emphasis to the intermediate-frequency substitution method using a waveguide-below-cutoff standard attenuator, and to d-c substitution techniques. Methods for measurement of small attenuations as well as methods not requiring reference to any standard attenuators are covered. Comments are made on the accuracy and convenience of various methods, and references are given which cover most of the basic and important research in this field.

6.c. **Analysis of a differential phase shifter,** Doyle A. Ellerbruch, IEEE Trans. Microwave Theory & Tech. MTT-12, No. 4,453-459 (July 1964).

This paper presents the theory and analysis of a ganged pair of "line stretcher" microwave phase shifters. The error analysis shows that some of the errors inherent in a single phase shifter of this type can be reduced through the use of a differential system; however, the magnitudes of other errors may more than offset the reduction. Graphical data are included to facilitate the rapid determination of the limit of error for any specified angle measurement.

6.d. **Evaluation of a microwave phase measurement system,** Doyle A. Ellerbruch, J. Res. NBS 69C (Eng. & Instr.), No. 1, 55-65 (Jan.-Mar. 1965).

The best phase measurement sensitivity and accuracy are attained in a dual-channel balanced bridge type system. Two similar dual-channel systems are discussed in this report; these being the homodyne system and the modulated subcarrier. Both require low-frequency amplitude modulation in one of the channels; however, the homodyne system uses a balanced modulator. The difference in system theory due to the amplitude modulation difference is discussed, and the advantages of each are pointed out.

The standard phase shifter considered for these systems consists of a modified reflectometer terminated in a sliding short circuit. The requirements for a noncontacting short circuit are established, and its theoretical and experimental design data are given.

An error analysis is included, so the total error for any given system can quickly be determined. The maximum error can be derived by calculating the limits of the individual errors and adding them together. Equipments and tuning techniques that have been developed to reduce many errors to a minimum are discussed.

It is shown that either of these measurement systems are readily adaptable to automation. Samples of phase data that were taken from a few commercial components are included.

6-e. **Further analysis of the modulated subcarrier carrier technique of attenuation measurement,** William E. Little, IEEE Trans. Instr. & Meas. IM-13, No. 2 & 3, 71-76 (June-Sept. 1964).

The modulated subcarrier technique of attenuation measurement offers an accurate and convenient means of calibrating rotary vane attenuators. The technique also presents the possibility of being extended to other types of attenuation measurement. This paper gives an analysis of the output of the microwave detector, assuming it is completely linear; an error analysis of the technique; and a comparison of the technique with the 30-Mc IF substitution technique of attenuation measurement.

6-f. A method for the self-calibration of attenuation-measuring systems, Robert L. Peck, J. Res. NBS 66C (Eng. & Instr.), No. 1, 13-18 (Jan.-Mar. 1962).

The theory and experimental procedures are given for the self-calibration of insertion loss of attenuation-measuring systems. Four circuit configurations are developed. The calibrations may be obtained by simple graphical means or by an analytical solution. Experimental results are given which demonstrate that, by using the techniques outlined, attenuation calibrations of high accuracy may be made without reference to any previously calibrated attenuator.

6-g. A modulated subcarrier technique of measuring microwave phase shifts, G. E. Schafer, IRE Trans. Instr. I-9, No. 2, 217-219 (September 1960).

6-h. Error analysis of a standard microwave phase shifter, G.E. Schafer and R. W. Beatty, J. Res. NBS 64C (Eng. & Instr.), No. 4, 261-265 (Oct.-Dec. 1960).

A standard microwave phase shifter has been proposed which utilizes an adjustable short circuit attached to a tunable three-arm waveguide junction. Ideally, the change of phase of the emergent wave from the third port can be made to equal the change of phase of the equivalent load attached to the second port, whether the generator and detector are matched or not. The difference between the change of phase of the emergent wave from port 3 and the change of phase of the equivalent load attached to arm 2 because of imperfect tuning is termed tuning error. This analysis relates the tuning error to amplitude changes which are observed at the detector attached to arm 3 during the tuning procedure. Graphs are presented for determining the parameters needed to estimate limits of tuning error from the observations of amplitude changes during the tuning procedure. One calculates the change in phase of the reflection coefficient produced by the short circuit from its observed displacement and from the guide wavelength. This procedure results in errors called dimensional errors because of uncertainties in determining the axial motion of the short circuit and because of slight variations from the nominal broad dimension of the waveguide. Limits of these dimensional errors are calculated for WR-90 waveguide in the recommended frequency range of 8.2 to 12.4 kilomegacycles per second, and presented in graphical form.

See also 7.a, 10.1.

7. Impedance

Papers

Abstracts

* Private communication

7. Impedance—Continued

* Private communication.

Definitions of v, i, Z, Y, a, b, Γ, and S

DAVID M. KERNS

Abstract—Concepts and conditions underlying the establishment and use of immittance- and scattering-matrix descriptions of waveguide n-ports are discussed. TEM modes are considered and no restriction to high or microwave frequencies is implied. The discussion is intended to be critical and intensive rather than general. Needed results of electromagnetic (and waveguide) theory are assumed. Emphasis is placed on defining basic quantities needed in the matrix scheme. These include: generalized voltage and current, v and i, for waveguide modes; modal impedance; modal characteristic impedance; the immittance matrices Z and Y; traveling-wave amplitudes a and b; reflection coefficient Γ; and the scattering matrix S. Properties of Z, Y, and S are not discussed; applications are indicated but not discussed.

I. Introduction

IF THE TITLE of this article appears to be wholly in a secret code, then this article may well be needed! The quantities symbolized are generalized voltage and current v and i; impedance and admittance matrices Z and Y; traveling-wave amplitudes a and b; reflection coefficient Γ; and, finally, scattering matrix S. All these quantities are to be considered in a context of waveguides and waveguide junctions. Coaxial transmission line, in particular, is considered as one form of waveguide and no restriction to high or microwave frequencies is implied.

It is hardly necessary to say that the elements of Z, Y, and S are important measurable quantities. A simple and very important example is the single element of the scattering "matrix" of a passive one-port, S_{11}. This quantity is commonly denoted by Γ and is referred to as the reflection coefficient of a termination. Another example is afforded by the off-diagonal elements of the scattering matrix of a passive two-port, S_{12} and S_{21}, which determine the flat-line power transmission and phase-shift characteristics of the device. Now, other papers in this issue of the PROCEEDINGS are concerned with *methods* of measuring these and related quantities. This paper is concerned with the meaning of the quantities being measured rather than with methods of measurement.

More specifically, the purpose of this paper is to outline the principal concepts and conditions underlying a scheme of definition and use of the quantities listed in the title. In a short article with this subject it is, of course, quite impossible to give a complete theoretical background; indeed, most theorems and relationships directly needed will not be proved herein. References to the literature will be given at appropriate points in the text.

II. Electromagnetic Fields in Waveguide

Our first tasks are to define the types of waveguide to be considered and to describe the general character of the possible electromagnetic fields that can exist in a section of waveguide of the types considered.

Somewhat arbitrarily, we limit the waveguides to be considered to those consisting either of one hollow conductor or of two conductors, one of which is hollow and encloses the other. Thus we shall be considering types of waveguide, such as hollow rectangular pipes, which do not support a principal ($\equiv$TEM) mode, as well as types, such as coaxial line, which do support a principal mode. For the most part we can consider both of these classes of waveguide simultaneously and from a unified point of view.

The waveguides to be considered are to be, by hypothesis, *ideal* waveguides; that is, waveguides made of perfectly conducting material, having perfectly cylindrical walls, and filled with a medium that is homogeneous, isotropic, nondissipative, and linear. For ordinary waveguides this hypothesis frequently represents a good approximation. Moreover, this approximation is the first of only two basic approximations needed in the establishment of the descriptive scheme with which this article is concerned.

We consider only the common and basic case in which the field quantities vary harmonically with time t at frequency $\omega/(2\pi)$. The time dependence is represented by the suppressed complex exponential factor $\exp(j\omega t)$, and the treatment will thus involve complex amplitudes (both vector and scalar), rather than instantaneous real quantities. Within a waveguide, then, the complex electric-field vector $\boldsymbol{E}$ and the complex magnetic-field vector $\boldsymbol{H}$ satisfy Maxwell's equations in the form

$$\operatorname{curl} \boldsymbol{E} = -j\omega\mu\boldsymbol{H},$$
$$\operatorname{curl} \boldsymbol{H} = +j\omega\varepsilon\boldsymbol{E},$$

where the parameters μ and ε (representing, respectively, permeability and permittivity) are positive real scalars independent of $\boldsymbol{E}$, $\boldsymbol{H}$, position, and time (μ, ε may depend upon ω). The field is, moreover, subject to the boundary condition that the tangential component of $\boldsymbol{E}$ vanish on the surface of the waveguide. The sources of the field are outside the region under consideration.

The most general electromagnetic field satisfying the above differential equations and the boundary condition can be expressed as a superposition of an infinite number of elementary particular solutions (TE, TM, and TEM modes) characteristic of the cross section of the waveguide. (The fact that the waveguide modes cannot easily be calculated in detail except for a few simple cross-sectional shapes is here immaterial. For our purpose, only general results of the theory of waveguides, holding for waveguides of arbitrary cross section, are needed.) With each mode (other than a TEM mode) is associated a critical or cutoff fre-

Manuscript received March 31, 1967; revised April 7, 1967.
The author is with the National Bureau of Standards, Boulder, Colo.

quency, at which its axial propagation constant changes from pure real to pure imaginary as the frequency increases. The modes form a sequence such that, for any given frequency, only a finite number can have a pure imaginary propagation constant (and thus be "propagated"). Often just one mode is propagated: the TEM mode in two-conductor waveguide, and the dominant mode in uniconductor waveguide.

We make the important assumption that only one mode (a propagated mode) need be considered in the section of waveguide under consideration. This represents a restriction in the generality of the discussion and an element of approximation. In the multimode case, each mode can be treated in much the same way as the one mode considered here will be treated, and the principle of superposition can be applied.[1] The element of approximation is rarely intrusive. It must be expected that higher (evanescent) modes will be excited by discontinuities at either end of the section of waveguide under consideration. Essentially all that is required is that the section of waveguide being considered be long enough to provide very high attenuation of these higher modes. Ordinarily a distance equal to a few times a cross-sectional dimension would be sufficient.

It will be convenient to refer to the "section of waveguide under consideration," whose role in interconnecting two devices will eventually emerge, as a waveguide "lead."[2] The term is suggested by the usage applying to wires interconnecting electrical devices.

We emphasize that a waveguide lead is to be a section of ideal waveguide. This is not to say that an imperfect physical joint between waveguides cannot be involved or cannot be considered; an imperfect joint, like any other discontinuity, should properly be considered as a device by itself, or be combined with some other device.

The conditions that have been imposed (harmonic time dependence, ideal waveguide, and the one-mode hypothesis) are directly concerned only with the situation within a waveguide lead; no condition is directly imposed on the nature of the devices at either end of the waveguide lead.

III. Definitions of v, i, and Related Quantities

In this section we wish to give meaningful and fairly rigorous definitions of v, i, impedance, and characteristic impedance—all with reference to one- and two-conductor waveguides. The symbols v and i are used advisedly, for these quantities represent generalizations of conventional transmission-line voltage and current, which are conventionally defined by means of certain line integrals. Here no attempt is made to retain line-integral concepts as basic; instead, a rather different approach, which has been described in essence a number of times in the literature [1]–[11], is used.

[1] For discussion including consideration of the multimode case, see Kerns and Beatty [1].

[2] This usage is apparently not common, but it has been found useful locally, and it seems that some such term is needed. It is suggested as a convenient usage applicable also to multimode waveguide serving an interconnecting role.

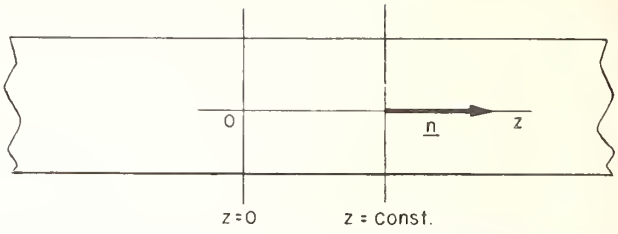


Fig. 1. Schematic of section of waveguide, showing z-axis of coordinates, unit vector $\boldsymbol{n}$ in z-direction, reference plane z = const, and arbitrary devices at either end.

We choose the z-axis of a system of coordinates parallel to the generators of the cylindrical waveguide boundary, and let $\boldsymbol{n}$ be a unit vector in the positive z-direction. The transverse coordinates will be denoted by x and y, but these need not be rectangular coordinates. A plane z = const will be termed a reference plane; the portion of this plane within the waveguide cross section will also be referred to as a terminal surface. This usage is common for reasons that will become apparent. The terminal surface is assumed to be located within the section of waveguide being considered at some distance from any discontinuity, as suggested in Fig. 1.

For our purpose it is sufficient as well as necessary to consider the transverse components (i.e., the components perpendicular to z) of the electromagnetic field in the waveguide, and so the z-components will be ignored. From waveguide theory it is known that (under the prescribed conditions) the transverse components $\boldsymbol{E}_t$, $\boldsymbol{H}_t$ of the electromagnetic field on the surface z = const are expressible in the form

$$\boldsymbol{E}_t = v(z)\boldsymbol{e}^o(x, y), \qquad \boldsymbol{H}_t = i(z)\boldsymbol{h}^o(x, y). \tag{1}$$

The quantities v and i, which are defined by these equations, are scalar coefficients, in general complex, and (for fixed excitation) are functions only of z. The entities $\boldsymbol{e}^o(x, y)$, $\boldsymbol{h}^o(x, y)$ are transverse vectorial field-pattern (or "basis") functions characteristic of the mode involved.[3] These functions are subject to normalization, as will be described; as the first step in this process it is permissible and convenient to assume that the functions are chosen real.

Now, we may regard the left-hand sides of (1) as being physically determined. Maxwell's equations do not—indeed, cannot—tell us how to allocate numerical factors in the representations on the right-hand sides. Should we, for example, use a "large" $\boldsymbol{e}^o$ and a correspondingly small v, or what? This is the question of normalization, and normalization is essentially a matter of convenience. What is appropriate at this stage is to set up a general and flexible scheme in which important quantities are exhibited explicitly and which will be convenient for later use. A scheme that seems to meet these requirements is given in the following paragraphs.

[3] Superscript (and later subscript) o's are used not to indicate a mode but to distinguish the set of entities most directly involved in the normalization scheme.

We impose a *power-normalization* condition

$$\frac{1}{2}\int_S \boldsymbol{e}^o \times \boldsymbol{h}^o \cdot \boldsymbol{n}\, dS = W_o; \tag{2}$$

here the integration goes over the cross section S of the waveguide, and W_o is a real, positive constant to be chosen at convenience. We note an immediate consequence of (1) and (2): by using Poynting's theorem for the complex Poynting's vector, we find

$$\int_S \frac{1}{2}\boldsymbol{E} \times \bar{\boldsymbol{H}} \cdot \boldsymbol{n}\, dS = W_o v\bar{\imath}, \tag{3}$$

where the superposed bar denotes the complex conjugate. The time-average power-flux in the positive z-direction across the terminal surface is the real part of $W_o v\bar{\imath}$. Thus we arrive at a fundamental network equation. One might suppose that $W_o = 1$ (or perhaps $W_o = 1/2$) would nearly always be the most convenient choice for W_o, but in some other less systematic formulations, with less emphasis on the introduction of a complete descriptive scheme, other choices of W_o have been made implicitly.

Another normalizing condition is needed (since there are two functions to be normalized). Having committed ourselves to (2), which normalizes the product of $\boldsymbol{e}^o$ and $\boldsymbol{h}^o$, the only possibility is something that essentially normalizes the ratio of $\boldsymbol{e}^o$ and $\boldsymbol{h}^o$. Waveguide theory tells us that for any mode, $\boldsymbol{h}^o$ will be proportional to $\boldsymbol{n} \times \boldsymbol{e}^o$. Hence we impose the *impedance normalizing* condition

$$\boldsymbol{h}^o = Z_o \eta \boldsymbol{n} \times \boldsymbol{e}^o, \tag{4}$$

where η is the wave-admittance of the mode involved and Z_o is the impedance normalizing constant. For TEM modes the wave-admittance has the value $(\varepsilon/\mu)^{1/2}$, for TM modes it is $\omega\varepsilon/\beta$, and for TE modes it is $\beta/(\omega\mu)$; here β is the phase constant for the mode involved. Since these quantities are real and positive, consistency with (2) requires that Z_o be real and positive.[4] Z_o is otherwise arbitrary. (Note that in (4) we have chosen to write an arbitrary constant in the form of a definite constant times an arbitrary constant.)

The normalization constants are called "constants" because: 1) once chosen, they are not to be changed without notice; and 2) they definitely do not depend upon the position coordinates (x, y, z), although they may be chosen, at convenience, to depend upon parameters such as frequency and waveguide dimensions. The process of changing from one chosen set of normalization constants to another chosen set may be called "renormalization," and is discussed briefly in Section V.

The basis functions $\boldsymbol{e}^o$ and $\boldsymbol{h}^o$ are derived essentially from the eigenvalue problem for the waveguide cross section; by (2) and (4) these functions are normalized up to the choice of a sign. Example: For a rectangular waveguide having its walls in the planes $x = 0$, a and $y = 0$, b $(a > b)$, it can be shown that

$$\begin{aligned} \boldsymbol{e}^o &= \pm (Z_o\eta)^{-1/2} \cdot 2\left[W_o/(ab)\right]^{1/2} \sin(\pi x/a)\boldsymbol{e}_y \\ \boldsymbol{h}^o &= \mp (Z_o\eta)^{+1/2} \cdot 2\left[W_o/(ab)\right]^{1/2} \sin(\pi x/a)\boldsymbol{e}_x \end{aligned} \tag{5}$$

where $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$ are the x and y unit vectors. In these expressions the roles of Z_o and W_o are evident.

We may now define the quantity

$$v/i = Z \tag{6}$$

as the *impedance* of the field of the mode being considered in the transverse surface $z = \text{const}$. We note that this is a modal quantity (it would be defined the same way for the mode involved even if other modes were allowed to be present) and that a direction is involved in its definition. In the most general situation, with arbitrary sources at either end of the waveguide lead, Z can take on any value, real or complex, at a given terminal surface. If the device at the right-hand end of the waveguide section is a passive one-port, a unique value of Z/Z_o is determined (by Maxwell's equations). We denote this unique value by Z_{11}/Z_o and say that Z_{11} is the impedance of the one-port relative to the chosen normalization and referred to the chosen reference plane. Example: If the waveguide is terminated by a perfectly conducting wall at a distance l from the reference plane, then $Z_{11} = jZ_o \tan \beta l$.

In order to describe an n-port, we must introduce the concept of transfer impedance. This will be done in the next section. First, however, we wish to consider further consequences of the definitions laid down thus far.

Having defined impedance as in (6), we can now define characteristic impedance. We shall do this in such a way that characteristic impedance will possess what appears to be the most useful and most widely accepted fundamental property: The *characteristic impedance* of the mode involved is defined to be that particular value of v/i that corresponds to a field consisting solely of a wave traveling in the positive z-direction. The formulation has been arranged so that this characteristic impedance is precisely the normalization constant Z_o—which was introduced as and still is an essentially arbitrary normalization constant.[5] Usually the arbitrariness of the value of the characteristic impedance is not stressed as much as it is here. Also, the term characteristic impedance has been applied to other quantities, defined on the basis of different fundamental properties and with different motivation. These points will be touched upon in the discussion that follows.

We record here an immediate consequence of (3), (6), and the definition of Z_o. Namely, if we choose $W_o = 1/2$, the power carried by a pure traveling wave is given by

[4] Evanescent modes (for which η is pure imaginary) or even a lossy medium in the otherwise ideal waveguide lead (in which case η is complex) can be dealt with conveniently by choosing Z_o so that the *product* $Z_o\eta$ remains real and positive. No other changes would be required in the present formulation. However, the properties of the eventually resulting scattering matrices would be changed markedly [12].

[5] Although Z in general depends upon the position-coordinate z, when $Z = Z_o$, it is independent of z. This important property could be taken as the basis for an alternative, but equivalent, definition of characteristic impedance. We see also that the definition given is consistent with the notion of characteristic impedance as iterative impedance.

$$W = \frac{1}{2} v\bar{i} \tag{7a}$$

$$= \frac{1}{2} Z_o |i^2| \tag{7b}$$

$$= \frac{1}{2} |v^2|/Z_o. \tag{7c}$$

Choosing $W_o = 1/2$ gives v and i "peak-value" normalization and is appropriate for later reference to these equations.

We note that characteristic impedance is characteristic not of the waveguide but of a particular state of the (modal) electromagnetic field in the waveguide. Despite the wide use of the phrase "characteristic impedance of a waveguide," a little consideration shows that the phrase cannot mean what it seems to say, for in general a multiplicity of modes, all with different characteristic impedances, can be present simultaneously in a waveguide. One might (arbitrarily) assign all the modes the same characteristic impedance, but this would hardly make it characteristic of the waveguide!

What has been said does not reduce the real significance of the conventional definition of characteristic impedance for a principal mode, for which voltage and current are conventionally defined by means of certain line integrals. If one chooses W_o to be unity and Z_o to have the conventional value (for coax, say), then the present scheme defines v and i identically equal to the conventionally defined (root-mean-square) voltage and current, respectively. The equality includes equality of units and dimensions: v and i come out in volts and amperes (as happens whenever Z_o is given the dimensions of ohms and W_o is dimensionless). Moreover, a two-conductor line, such as a coaxial line, may be considered as a standard of impedance: the conventional Z_o, in the absence of standing waves, prescribes directly what the value of the corresponding ratio of conventional v and i must be.

Earlier it was stated that normalization constants could be chosen at convenience. For a principal mode the conventional choice is clearly convenient. Nevertheless, the price in inconvenience that would have to be paid for some other choice is not great. Differences in impedance normalization show up as a slight complication in the "joining equations." These are discussed in Section V.

Line integrals have been used to some extent to define generalized voltage and current for modes other than principal modes. We shall discuss this approach briefly and conclude that it is not really relevant to our purpose. Line integrals of transverse $\boldsymbol{E}$ and transverse $\boldsymbol{H}$ possess the essential property of being linear measures of these field components, respectively, just as v and i defined in (1) are. Because such integrals are path-dependent, it is in general necessary to specify not only the termini but also the path between the termini. (This means that one needs a knowledge of the field patterns and thus the construction of line integrals does not seem feasible for an arbitrary mode in a waveguide of arbitrary cross section.) Line-integral definitions of this type are discussed by Schelkunoff [13]. For example, for rectangular waveguide, of dimensions a and b ($a > b$), operating in the dominant (TE_{10}) mode, he chooses line integrals such that (for a pure traveling wave),

$$V = bEe^{-j\beta z}, \qquad I = \frac{2}{\pi} a\eta E e^{-j\beta z}, \tag{8}$$

where E is a constant. These definitions would lead to $W_o = \pi/8$ and $Z_o = \pi b/(2a\eta)$ in the present scheme. This normalization would be, of course, perfectly workable.

Schelkunoff also defines a set of three impedances, called " 'integrated' characteristic impedances." One of his definitions is the same as that adopted here for characteristic impedance; the other two are different, being based on a different property taken as fundamental. Namely, taking the *form* of the power equations (7b) and (7c) as fundamental, he defines $Z_{W,I}$ and $Z_{W,V}$ so that

$$W = \frac{1}{2} Z_{W,I} |I^2| \tag{9a}$$

$$= \frac{1}{2} |V^2|/Z_{W,V}. \tag{9b}$$

For principal modes, with conventional line-integral definitions of V and I, it turns out that $Z_{W,I} = Z_{W,V} = Z_o$. For modes other than principal modes these equalities do not necessarily hold; and in particular, for the TE_{10} mode in rectangular waveguide, with V and I given by (8), the equalities do not hold. In this case, in other words, one needs different integrated characteristic impedances according as one chooses to calculate power in terms of I or V, and both of these impedances are different from Z_o.

Now, for rectangular waveguide (at least), all three of the impedances discussed in the preceding paragraph are proportional to the wave-impedance and they depend upon the cross-sectional dimensions of the waveguide in the same way. Hence they are all equally useful in calculating reflections at discontinuities in rectangular waveguide, to the limited extent that this can be done simply in terms of such quantities. Other choices of impedances for other types of waveguide as well as for rectangular waveguide exist and may be proposed.[6] An underlying motivation is to find impedance definitions that are helpful in solving or approximately solving what are properly electromagnetic field problems. Such problems are not our present concern. The present purpose is merely to set up a descriptive scheme, into which results of measurements (or calculations) may be placed, and with the aid of which calculations (of network type) may be made. In such a scheme $W_o = 1$ and $Z_o = 1$ often serve admirably.

IV. Impedance, Admittance, and Scattering Matrices for Waveguide Junctions

Once the basic definitions pertaining to v and i are established, the definitions of Z, Y, and S for waveguide junc-

[6] Recently a definition differing slightly from the three discussed here was proposed by R. M. Walker in a correspondence entitled "Waveguide Impedance—Too Many Definitions" [14].

tions are straightforward. A waveguide junction is defined to be a linear electromagnetic system possessing ideal waveguide leads and is considered to be subject to excitation solely through the effects of propagated modes in the waveguide leads. (See Fig. 2.) For simplicity in notation and discussion we shall consider primarily a waveguide junction with just two leads, with just one mode propagating in each. Thus we shall be considering a two-port. This will probably serve as well as, or better than, the general case to indicate the basic ideas for both one-ports and n-ports.

In each of the two waveguide leads of the two-port we choose a terminal surface and denote it S_m; here, and subsequently, port indices such as $m, n, \cdots$ may be understood to take on the values 1, 2. From the preceding section we know that the transverse components of $\boldsymbol{E}$ and $\boldsymbol{H}$ on the terminal surfaces can be expressed in the form

$$\begin{aligned} \boldsymbol{E}_{mt} &= v_m \boldsymbol{e}_m^o, \\ \boldsymbol{H}_{mt} &= i_m \boldsymbol{h}_m^o, \end{aligned} \tag{10a}$$

where $\boldsymbol{e}_m^o$, $\boldsymbol{h}_m^o$ are real basis fields subject to the impedance normalization

$$\boldsymbol{h}_m^o = Z_{om} \eta_m \boldsymbol{n}_m \times \boldsymbol{e}_m^o \tag{10b}$$

and to the power normalization

$$\frac{1}{2} \int_{S_m} \boldsymbol{e}^o \times \boldsymbol{h}^o \cdot \boldsymbol{n}_m \, dS = W_{om}, \tag{10c}$$

where $\boldsymbol{n}_m$ is the unit vector on S_m directed *into* the junction; η_m is the wave admittance of the mode in waveguide m; and Z_{om}, W_{om} are real, positive normalization constants.

It should be clear that not only do the equations (10a) determine $\boldsymbol{E}_{mt}$, $\boldsymbol{H}_{mt}$ when the v's and i's are given but also they determine the v's and i's when $\boldsymbol{E}_{mt}$ and $\boldsymbol{H}_{mt}$ are given. On the basis of certain existence and uniqueness theorems of electromagnetic theory, we can say that the prescription of *either* transverse $\boldsymbol{E}$ *or* transverse $\boldsymbol{H}$ on both terminal surfaces is just sufficient to determine $\boldsymbol{E}$ and $\boldsymbol{H}$ throughout the junction. Thus if the v's are given, $\boldsymbol{E}$, $\boldsymbol{H}$, and the i's are determined; if the i's are given, $\boldsymbol{E}$, $\boldsymbol{H}$, and the v's are determined. Furthermore, since the junction being considered is linear and passive, the i's and the v's in the two cases will be related by homogeneous, linear equations.

To see this in more detail, suppose that $v_1 \neq 0$, $v_2 = 0$ is prescribed. Then the whole field will be proportional to v_1; in particular, $\boldsymbol{H}_{1t}$ and $\boldsymbol{H}_{2t}$, and hence i_1 and i_2, will be proportional to v_1. We write the constants of proportionality as follows:

$$\begin{aligned} i_1 &= Y_{11} v_1, \\ i_2 &= Y_{21} v_1. \end{aligned}$$

Similarly, if v_1 is zero and v_2 is not zero, we write

$$\begin{aligned} i_1 &= Y_{12} v_2, \\ i_2 &= Y_{22} v_2. \end{aligned}$$

Finally, if neither v_1 nor v_2 is zero, we use the principle of superposition and write

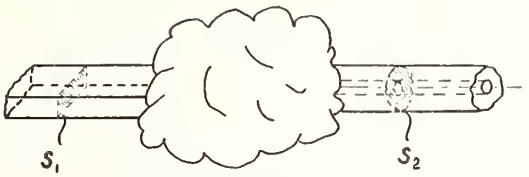


Fig. 2. Illustrating a two-port.

$$\begin{aligned} i_1 &= Y_{11} v_1 + Y_{12} v_2, \\ i_2 &= Y_{21} v_1 + Y_{22} v_2; \end{aligned} \tag{11}$$

the four admittances Y_{mn} are the elements of the admittance matrix of the two-port.

In an entirely similar way we can arrive at the equations

$$\begin{aligned} v_1 &= Z_{11} i_1 + Z_{12} i_2, \\ v_2 &= Z_{21} i_1 + Z_{22} i_2, \end{aligned} \tag{12}$$

thereby defining the elements of the impedance matrix of the two-port.

The elements Y_{mn} are analogous to the short-circuit input and transfer admittances of circuit theory; similarly the elements Z_{mn} are analogous to the open-circuit input and transfer impedances of circuit theory. (The quantities with indices $m = n$ are the input quantities, and those with $m \neq n$ are the transfer quantities.)

As a matter of notation we define the matrices

$$Y = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix}, \qquad Z = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix}. \tag{13}$$

These are the promised definitions of Y and Z matrices.

For a one-port, the matrices are one-dimensional:

$$Y = (Y_{11}), \qquad Z = (Z_{11});$$

Z_{11} was virtually defined in Section III, but without benefit of the important ideas discussed briefly in the third paragraph of this section.

We turn now to the definitions of a, b, S (and Γ). We consider first a waveguide lead by itself, as in Section III, without reference to devices at either end. The quantities a and b may be defined formally in terms of v and i by means of the equations

$$\begin{aligned} a &= \frac{1}{2}(v + Z_o i), \\ b &= \frac{1}{2}(v - Z_o i), \end{aligned} \tag{14a}$$

or by the inverse equations

$$\begin{aligned} v &= a + b, \\ Z_o i &= a - b. \end{aligned} \tag{14b}$$

Using the fact that $v/i = \pm Z_o$ for pure progressive waves traveling in the $\pm z$-direction, one can deduce that a and b are respectively the amplitudes (relative to $\boldsymbol{e}^o$) of the electric field of the progressive components of the field traveling in the + and − z-directions. The meaning of these quantities

is more clearly shown if we make the z-dependence explicit. With the help of waveguide theory it can be shown that our definitions lead to

$$v(z) = Ae^{-j\beta z} + Be^{j\beta z},$$
$$Z_o i(z) = Ae^{-j\beta z} - Be^{j\beta z}, \tag{15}$$

where A and B are constants. Comparison of these equations with (14b) shows that $a = A \exp(-j\beta z)$ and $b = B \exp(j\beta z)$. Thus, if we recall the suppressed time factor $\exp(j\omega t)$, the role of a and b as measures of the traveling-wave components of the field becomes evident. [Equations (15) incidentally imply that v and i, as functions of z, satisfy differential equations of transmission-line form, with propagation constant $j\beta$ and characteristic impedance Z_o.]

Using (3) and (14b), and choosing $W_o = 1$, we find the basic expression for complex power-flux in terms of a and b,

$$W = \frac{1}{\bar{Z}_o}(|a|^2 - |b|^2) + \frac{1}{\bar{Z}_o}(b\bar{a} - a\bar{b}).$$

The time-average power-flux is the real part of this expression; when Z_o is real (as it is in this paper), $\mathrm{Re}(W) = (|a|^2 - |b|^2)/Z_o$.

We now associate wave amplitudes a_m and b_m with the terminal surfaces of our two-port. The a's and b's may be characterized as *incident* and *emergent* wave amplitudes, respectively (because in (10b) $\boldsymbol{n}_m$ was chosen as the inward normal!). One may expect on physical grounds that there will exist a set of homogeneous, linear equations determining the emergent wave-amplitudes b_1, b_2 as functions of the incident wave-amplitudes a_1, a_2. Such a relationship is also implied mathematically by (11) or (12) and (14). Therefore we may write

$$b_1 = S_{11}a_1 + S_{12}a_2,$$
$$b_2 = S_{21}a_1 + S_{12}a_2. \tag{16}$$

One may analyze these equations by considering alternately $a_2 = 0$ and $a_1 = 0$. One sees, in view of the physical meaning of the a's and b's as traveling-wave amplitudes, that S_{11} and S_{22} are of the nature of reflection (or back-scattering) coefficients and that S_{12} and S_{21} are of the nature of transmission coefficients. The coefficients are often referred to collectively as scattering coefficients, and the square array

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

is the scattering matrix that was to be defined.

Γ is nothing more or less than a common notation for the element of the one-dimensional scattering matrix of a passive one-port: $S = (S_{11})$, $\Gamma = S_{11}$.

In the recent literature an interesting generalization of the scattering-matrix concept is discussed. The formalism, as well as some of the terminology, is similar to that long employed in microwave work (and employed here), but the physical meaning can be substantially different. The generalized concept does not appear to enable one to do anything essentially new, but it does offer an interesting, elegant, and sometimes expeditious way of arranging details. A brief quantitative discussion is given in the Appendix to this paper; for further discussion the reader is referred to the literature [15]–[17].

V. Renormalization; Joining Equations

Both of the topics to be discussed in this section came up in the discussion in an earlier section; both are important for the purpose of this paper because they have to do directly with the relationships between terminal fields and our representation of these fields.

A. Renormalization: We regard (10) to (16) as having been set up for some particular set of normalization constants; we recognize that these equations could have been set up for some other set of normalization constants; and we inquire how the quantities v, i, Z, etc., change—or transform—when the normalization is changed. The guiding principle to be used in finding the transformations is simply that the key physical entities being represented shall not be changed by the transformation.

We shall illustrate this for the case of v, i, and Z, taking Z to be the impedance matrix of a two-port.

The effect of a change of characteristic impedance on basis fields is for any $\boldsymbol{e}_m^o$ and $\boldsymbol{h}_m^o$ just that exemplified in (5). Let us indicate a change of characteristic impedances by the notation

$$Z_{om} \to Z'_{om},$$

where the unprimed and the primed quantities are, respectively, the old and the new ones. Since $\boldsymbol{E}_{mt}$ and $\boldsymbol{H}_{mt}$ must be invariant, (1) and (5) imply that the v's and i's must transform as follows:

$$v_m \to v'_m = \alpha_m^{-1} v_m, \qquad i_m \to i'_m = \alpha_m i_m,$$

where $\alpha_m = (Z_{om}/Z'_{om})^{-1/2}$. From these equations it follows further that the transformation of the elements of the impedance matrix must be

$$Z_{mn} \to Z'_{mn} = \alpha_m^{-1}\alpha_n^{-1} Z_{mn}.$$

(The transformation of the admittances Y_{mn} is just the inverse of that of the Z_{mn}.)

The wave amplitudes a, b and the off-diagonal elements of S are also affected by impedance renormalization. Further details, including a consideration of power renormalization, may be found in the literature [1].

B. Joining equations: We consider that a waveguide lead of one system is to be joined to a waveguide lead of another system. The waveguide leads must be of the same size and shape, of course, and it is assumed that there is no physical discontinuity. We assume that the terminal surfaces associated with each system have been so located that they coincide when the connection is made. (See Fig. 3.) The transverse components of $\boldsymbol{E}$ and $\boldsymbol{H}$ on the common terminal surface are then given by the equations

$$\boldsymbol{E}_t = v\boldsymbol{e}^o, \qquad \boldsymbol{H}_t = i\boldsymbol{h}^o \tag{17}$$

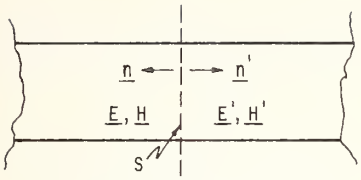


Fig. 3. Coincident terminal surfaces.

associated with the one system and also by the equations

$$\boldsymbol{E}_t' = v'\boldsymbol{e}^{o'}, \qquad \boldsymbol{H}_t' = i'\boldsymbol{h}^{o'} \tag{18}$$

associated with the other. Let us assume that we have $W_o = W_o'$ but *not* necessarily $Z_o = Z_o'$. Choosing signs so that $\boldsymbol{e}^o$ and $\boldsymbol{e}^{o'}$ are polarized in the same sense, we find with the aid of (5) and (10b) that

$$\boldsymbol{e}^{o'} = \alpha\boldsymbol{e}^o, \qquad \boldsymbol{h}^{o'} = -\alpha^{-1}\boldsymbol{h}^o; \tag{19}$$

where $\alpha = \sqrt{Z_o/Z_o'}$ and the negative sign in the second equation arises because $\boldsymbol{n} = -\boldsymbol{n}'$.

Now, for the electromagnetic fields corresponding to (17) and (18) to be continuous across the terminal surface, as required by electromagnetic theory, it is necessary and sufficient that the transverse components be continuous. Thus, in view of (19), it is necessary and sufficient that

$$v' = \alpha^{-1}v, \qquad i' = -\alpha i. \tag{20}$$

These are the desired joining equations. They are of the same form as the equations describing an ideal transformer with turns ratio α. We prefer *not* to illustrate the relations with a diagram showing an ideal transformer because there is some danger of confusion. As long as we bear in mind the relationship of the terminal variables to the field quantities, it is clear that the joining equations have the form they do have, not because of any physical discontinuity, but because of the difference in impedance normalizations.

VI. Conclusion

At this point, the purpose of this paper, as announced in the introduction, has been accomplished—perhaps in a rather literal sense. Here, in order to tie things together better, and to make the motivation for some of the work more apparent, we should like to indicate how the definitions that have been made do form a basis for calculations of network type. To accomplish this we must first say a few words about the representation of active devices, or sources.

For simplicity we do not consider active devices of negative resistance or negative conductance type. We assume linear behavior of the device (at least from an external point of view), and so an active device will be represented by linear equations. The only difference from the previous equations describing passive devices is that now the equations must be inhomogeneous. We illustrate this for one-port sources. A one-port source may be represented by an impedance and an "open-circuit" voltage, by an admittance and a "short-circuit" current, or by a reflection coefficient and a generated-wave amplitude:

$$v = Z_g i + v_g \tag{21a}$$

$$i = Y_g v + i_g \tag{21b}$$

$$b = \Gamma_g a + b_g. \tag{21c}$$

The quantities bearing the subscript "g" characterize the given device relative to chosen normalizations and reference plane. The equations are simply the most general linear equations connecting pairs of variables. Since the equations must be of the form shown, we have in fact proved Thevenin's and Norton's theorems, the content of which is exhibited in (21a) and (21b), respectively. Apparently no name is attached to the important theorem represented by the third equation.

Let us now consider a two-port provided with a source joined at port 1 and with a passive termination at port 2. We choose to apply scattering equations and we assemble the equations descriptive of the three elements of the system. For the two-port itself we have (16), which is repeated here for convenience,

$$\begin{aligned} b_1 &= S_{11}a_1 + S_{12}a_2, \\ b_2 &= S_{21}a_1 + S_{22}a_2; \end{aligned} \tag{16}$$

for the source we have

$$b_1' = \Gamma_g a_1' + b_g;$$

and for the passive termination

$$b_2' = \Gamma_l a_2'.$$

The polarizations of the various a's and b's are indicated in Fig. 4. We notice that in (16) and in the last two equations, the number of unknowns is in each case twice the number of equations; to obtain determinate solutions we must, of course, utilize the joining equations. Joining equations for a's and b's are easily derived from (20); in the present instance, if we assume no "discontinuities in normalization" ($W_{om} = W_{om}'$, $Z_{om} = Z_{om}'$), one finds

$$a_1' = b_1, \; b_1' = a_1, \qquad a_2 = b_2', \; b_2 = a_2',$$

(as might be expected). Thus the number of equations is equal to the number of unknowns and the system will in general possess a unique solution. In fact, by eliminating all the variables but b_1 and b_2, one obtains the two equations

$$\begin{aligned} (1 - \Gamma_g S_{11})b_1 - \Gamma_l S_{12} b_2 &= S_{11} b_g, \\ -\Gamma_g S_{21} b_1 + (1 - \Gamma_l S_{22}) b_2 &= S_{21} b_g \end{aligned} \tag{22}$$

for the determination of b_1 and b_2.

As an example, we can use (22) to obtain a formula for a quantity of immediate physical significance, known as transducer power gain, and defined as the ratio of (net) power delivered to the load on terminal 2 to the available power of the source. Solving (22) for b_2, one finds

$$b_2 = S_{21} b_g / D,$$

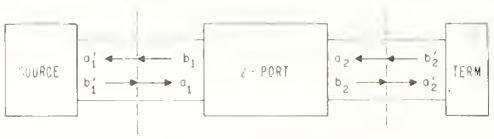


Fig 4. Schematic of system consisting of source, two-port, and passive termination.

where $D=(1-\Gamma_g S_{11})(1-\Gamma_l S_{22})-\Gamma_g\Gamma_l S_{12}S_{21}$. Using the expression for power in terms of a and b, given in Section IV, we find for the power P_l delivered to the load,

$$\begin{aligned} P_l &= (|b_2|^2 - |a_2|^2)/Z_{02} \\ &= |b_2|^2(1 - |\Gamma_l|^2)/Z_{02} \\ &= |S_{21}b_g/D|^2(1 - |\Gamma_l|^2)/Z_{02}. \end{aligned} \tag{23}$$

The available power of the source is, by definition, the power that would be delivered to a conjugate-match termination directly connected to the source. Finding P_A is simpler than finding P_l and so we merely state the result:

$$P_A = \frac{|b_g|^2}{Z_{01}} \frac{1}{1 - |\Gamma_g|^2}. \tag{24}$$

The transducer power gain is thus

$$\frac{P_l}{P_A} = \frac{Z_{01}}{Z_{02}} \frac{|S_{21}|^2(1 - |\Gamma_g|^2)(1 - |\Gamma_l|^2)}{|(1 - \Gamma_g S_{11})(1 - \Gamma_l S_{22}) - \Gamma_g\Gamma_l S_{12}S_{21}|^2}. \tag{25}$$

Equations (24) and (25) are given as (2.41) and (2.52), respectively, of Kerns and Beatty [1], where these and other related quantities are discussed.

In a similar way, one could show how Z or Y might be used for calculations, but we shall not do this here.

The ramifications of the scheme, the basis of which we have described, are most numerous. Additional matrices are defined; an important type, useful with either v, i or a, b variables, is known variously as a transmission, transfer, or cascading matrix. All the matrices are interrelated. All may be subject to constraints, such as reciprocity, realizability, losslessness, which are useful in both measurement and calculations. Applications to the description of particular devices, such as circulators, directional couplers, T-junctions, and to analysis and design of systems (emphatically including measurement systems) are innumerable. The references that have been listed are intended partly as suggestions for further reading.

Appendix

In this Appendix we give a brief quantitative discussion of the generalized scattering matrix mentioned in the text. We consider a system consisting of source, two-port, and passive termination (as in the preceding section, and with normalizations as in the preceding section). As will be seen, the generalized scattering matrix is a system matrix for such a system, and it will embody the properties of the terminations as well as those of the two-port.

We define "power waves" (to use Kurokawa's term) associated with port m of the two-port by means of the equations

$$\begin{aligned} \hat{a}_m &= \frac{1}{2\sqrt{R_m}}(v_m + Z_m i_m), \\ \hat{b}_m &= \frac{1}{2\sqrt{R_m}}(v_m - \bar{Z}_m i_m), \end{aligned} \tag{26}$$

where Z_m is the impedance (of source or load) terminating port m, and R_m is the real part of Z_m. Strictly, $|R_m|$ should appear in (26), but we are for simplicity considering only positive R_m. Sometimes Z_m is referred to as a "normalizing number" and (26) is said to represent "complex normalization." Whether or not these terms are used, we emphasize that (26) has nothing to do with the normalizations in our definitions of v and i. As far as (26) is concerned v and i are given a priori. This is the point of view in the references cited, where v and i are in fact not defined explicitly.

The power waves are in general linear combinations of the a's and b's defined in Section IV, and they are so contrived that the $\hat{a}$'s are essentially given quantities. Thus in the present instance,

$$\hat{a}_1 = v_{g1}/(2\sqrt{R_1}), \qquad \hat{a}_2 = 0.$$

To verify these equations one may use (21a) and (20) to obtain the equations $v_1=v_{g1}-Z_1 i_1$, $v_2=-Z_2 i_2$ pertaining to the terminations, and then substitute these into (26). With $W_a=1$, the time-average power input to the two-port at port m is given in terms of the power waves by

$$P_m = |\hat{a}_m|^2 - |\hat{b}_m|^2, \tag{27}$$

as can be verified by using (3) and (26).

The generalized (or power-wave) scattering matrix is defined formally by

$$\begin{aligned} \hat{b}_1 &= \hat{S}_{11}\hat{a}_1 + \hat{S}_{21}\hat{a}_2 \\ \hat{b}_2 &= \hat{S}_{21}\hat{a}_1 + \hat{S}_{22}\hat{a}_2 \end{aligned}$$

in the same manner as S was defined in (16). Inasmuch as the $\hat{a}$'s are knowns, the scattering equations do not have to be solved; they give the $\hat{b}$'s immediately. The actual problem is to find the $\hat{S}_{mn}$. These can be expressed as functions of Z_m, Z_{om}, and the Z, Y, or S matrix of the two-port.

In the present instance we have, for example,

$$\hat{b}_2 = \hat{S}_{21}a_1 = \hat{S}_{21}v_g/(2\sqrt{R_1}).$$

Now, according to [27], $|\hat{b}_2|^2$ represents the power delivered to the load ($\hat{a}_2$ being zero); $|v_g|^2/(4R_1)$ is recognized as the available power at the source; and so $|\hat{S}_{21}|^2$ is seen to be just the transducer power gain that was evaluated in the preceding section.

There is, of course, much more that can be said about power waves and the associated scattering matrix, but for

this the reader is referred to the references already cited. Perhaps enough has been said here to indicate the relationship and the distinction between S and $\hat{S}$.

Acknowledgment

The author is indebted to R. W. Beatty, L. B. Elwell, and P. F. Wacker for many helpful discussions.

References

[1] D. M. Kerns and R. W. Beatty, *Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis*. New York: Pergamon Press, 1967.

[2] D. S. Saxon, M.I.T. Radiation Lab., memorandum, consisting of an introductory section for notes on lectures by J. Schwinger, "Discontinuities in waveguides," February 1945.

[3] J. W. Miles, "The equivalent circuit for a plane discontinuity in a cylindrical wave guide," *Proc. IRE*, vol. 34, pp. 728–742, October 1946.

[4] W. Altar, "Q circles—A means of analysis of resonant microwave systems," pt. I, *Proc. IRE*, vol. 35, pp. 355–361, April 1947; and pt. II, *Proc. IRE*, vol. 35, pp. 478–484, May 1947.

[5] *Principles of Microwave Circuits*, M.I.T. Radiation Lab. Ser., vol. 8, C. G. Montgomery, R. H. Dicke, and E. M. Purcell, Eds. New York: McGraw-Hill, 1948.

[6] *Waveguide Handbook*, M.I.T. Radiation Lab. Ser., vol. 10, N. Marcuvitz, Ed. New York: McGraw-Hill, 1951.

[7] N. Marcuvitz and J. Schwinger, "On the representation of the electric and magnetic fields produced by currents and discontinuities in waveguides," *J. Appl. Phys.*, vol. 22, pp. 806–819, June 1951.

[8] M. Sucher and J. Fox, *Handbook of Microwave Measurements*. Brooklyn, N. Y.: Polytechnic Press of the Polytechnic Institute of Brooklyn, 1963.

[9] R. E. Collin, *Field Theory of Guided Waves*. New York: McGraw-Hill, 1960.

[10] H. A. Atwater, *Introduction to Microwave Theory*. New York: McGraw-Hill, 1962.

[11] R. N. Ghose, *Microwave Circuit Theory and Analysis*. New York: McGraw-Hill, 1963.

[12] H. Haskal, "Matrix description of waveguide discontinuities," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-12, pp. 184–188, March 1964.

[13] S. A. Schelkunoff, *Electromagnetic Waves*. New York: Van Nostrand, 1943, p. 318.

[14] R. M. Walker, "Waveguide impedance—too many definitions," *Electronic Communicator*, vol. 1, p. 13, May–June 1966.

[15] H. J. Carlin and A. B. Giordano, *Network Theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1964.

[16] K. Kurokawa, "Power waves and the scattering matrix," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-13, pp. 194–202, March 1965.

[17] R. W. Newcomb, *Linear Multiport Synthesis*. New York: McGraw-Hill, 1966.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967
pp. 892-900
The Institute of Electrical and Electronics Engineers, Inc.
PRINTED IN THE U.S.A.

Impedance Measurements and Standards for Uniconductor Waveguide

ROBERT W. BEATTY, FELLOW, IEEE

Abstract—A tutorial review is presented of the measurement of impedance and reflection coefficient in uniconductor waveguide. Normalized impedance in a waveguide operating in a particular mode is defined and related to measured quantities such as the reflection coefficient and the VSWR.

Emphasis is given to the rectangular waveguide operating in its dominant mode and to the tuned reflectometer as an instrument for achieving the most accurate results. The evolution of the tuned reflectometer at NBS is outlined and recent techniques are discussed. Different types of standards of reflection coefficient are described and the advantages and limitations of each are mentioned.

Introduction

ACTUALLY, no one measures impedances, in the usual sense, in uniform uniconductor waveguides. Instead, one usually measures the VSWR (voltage standing-wave ratio) and the position of a minimum in the voltage standing wave, or the reflection coefficient which exists when a certain mode propagates in a certain waveguide and reflection takes place from a termination. Usually, the propagating mode is the dominant one (having the lowest cutoff frequency) and the waveguide is a uniform one having a standard set of cross-sectional dimensions.

In a given waveguide, at a given frequency and given mode, there is only one VSWR corresponding to a given load or termination. Also there is only one value of reflection coefficient and one value of normalized impedance[1] (to be defined later). However, there may be various values of impedance, depending upon the value chosen for Z_0, the "characteristic impedance" or normalizing impedance.

Thus the title of this paper is not really a fraud, since the normalized impedance which one might wish to measure in a uniconductor waveguide is uniquely determined by the characteristics of the termination, the waveguide, the propagating mode, and the frequency.

It is shown in another paper [1] in this issue that any value may be arbitrarily chosen for Z_0; however, there may be advantages and disadvantages associated with each choice. No one choice will satisfy all requirements in the case of uniconductor waveguide. In the case of the TEM (Transverse Electric and Magnetic) mode in two-conductor waveguide, such as coaxial line, the choices of Z_0 have always been made according to established formulas. (It is not widely recognized that other choices could be made, and that there might occasionally be advantages to doing so.)

Manuscript received March 31, 1967; revised April 4, 1967.

The author is with the National Bureau of Standards, Boulder, Colo.

[1] This assumes that the same scale is used to measure the amplitude of the waves traveling in each direction.

In the following, reflection coefficient and VSWR in waveguide will be defined before reviewing methods of measuring impedances, VSWR's, and reflection coefficients. Emphasis will be given to the tuned-reflectometer method, which has the greatest potential accuracy. The slotted line will be mentioned but is discussed in greater detail in another paper [2] in this issue. Finally, standards of reflection coefficient for uniconductor waveguide systems will be discussed. The presentation begins at a rather elementary level, with material well known to workers in the field, then tapers into a more advanced discussion. Most of the technical details are left to the references, however.

Definitions

In making careful measurements, one needs first to precisely define the quantity to be measured. Otherwise, one cannot precisely analyze or evaluate the uncertainties in the measurement. The paper by D. M. Kerns entitled "Definitions of v, i, Z, Y, a, b, Γ, and S," in this issue, has been written with this in mind. The reader is referred[2] to that paper [1] for definitions of generalized voltage v and current i in waveguide, impedance Z, admittance Y, incident and reflected wave amplitudes a and b, reflection coefficient Γ, and scattering coefficient S.

It has been assumed that waveguides are uniform and made of metal having infinite conductivity; therefore, the definitions are intimately associated with an idealized model for a waveguide. The errors caused by the difference between an actual waveguide and its model can be kept small and usually are of no concern. They might become of concern in the future, however, as the precision of measurement increases.

In measurements with a slotted section of waveguide, one observes the VSWR[3] or voltage standing-wave ratio ㄷ. In a lossless line, this is usually defined as the ratio of the maximum to the minimum value of the standing wave of $|v|$. One may correct or compensate for the effect on v of the attenuation of the waveguide, and refer the result to a conveniently located terminal surface in the waveguide. Phase information is obtained by oserving the position of a

[2] See also Kerns and Beatty [3].

[3] The Greek letter σ has been used to denote VSWR, but the same symbol is often used to denote electrical conductivity. Conflicts of this kind can be avoided by using symbols which have seldom, if ever, been used to denote electrical quantities. A rich source of such symbols is the Japanese Katakana alphabet. From it, the symbol ㄷ has been selected. It means "ratio," and is pronounced "hee."

minimum point of the standing wave relative to this terminal surface.

In measurements with a reflectometer, one observes $|\Gamma|$, the magnitude of the voltage reflection coefficient. The phase ψ of Γ is usually not measured with the reflectometer, but can be obtained if desired using modified techniques.

The relationships among the quantities (which are measured and desired) are given as follows:

$$\left.\begin{aligned} v &= a + b \\ Z_0 i &= a - b \end{aligned}\right\} \tag{1}$$

$$Z = \frac{v}{i}. \tag{2}$$

(Note that Z is a function of position along the waveguide and is indeterminate until the normalizing impedance Z_0 has been specified.)

$$Z_0 = \left(\frac{v}{i}\right)_{b=0} \tag{3}$$

$$[\text{VSWR}] \text{ or } \rho = \left|\frac{v_{\max}}{v_{\min}}\right| = \frac{|a| + |b|}{|a| - |b|} = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{4}$$

$$\Gamma = \frac{b}{a} = \frac{v - Z_0 i}{v + Z_0 i} = \frac{Z - Z_0}{Z + Z_0} \tag{5}$$

$$|\Gamma| = \frac{\rho - 1}{\rho + 1}. \tag{6}$$

The normalized impedance Z/Z_0 is a dimensionless and, in general, complex quantity z. It is clear that

$$z = \frac{1 + \Gamma}{1 - \Gamma} \quad \text{and} \quad \Gamma = \frac{z - 1}{z + 1}. \tag{7}$$

Effect of Change of Position of Terminal Surface

It is emphasized that Z and Γ are referred to a definite terminal surface or reference plane in a waveguide. The effect on Z and Γ of changing the position of the reference plane is therefore of considerable interest in impedance measurements.

A change of position of the terminal surface, at which the terminal variables v and i (or a and b) are determined, changes the reflection coefficient Γ in a simple manner, but the normalized impedance z is changed in a more complicated way.

The diagram shown in Fig. 1 applies to the case where Γ_2 is known and we wish to determine Γ_1 at a distance l toward the generator, which is along a uniform waveguide having a propagation constant $\gamma = \alpha + j\beta$.

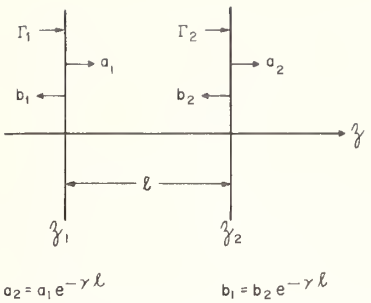


Fig. 1. Change of position of terminal surface.

$$a_2 = a_1 e^{-\gamma l}, \quad \text{and} \quad b_1 = b_2 e^{-\gamma l}. \tag{8}$$

Therefore,

$$\Gamma_1 = \frac{b_1}{a_1} = \frac{b_2}{a_2} e^{-2\gamma l} = \Gamma_2 e^{-2\gamma l}. \tag{9}$$

The magnitudes are related by

$$|\Gamma_1| = |\Gamma_2| e^{-2\alpha l}, \tag{10}$$

and the phase of Γ_1 equals the phase of Γ_2 minus $2\beta l$, or

$$\psi_1 = \psi_2 - 2\beta l + 2n\pi, \tag{11}$$

where n is an integer.

Thus one can "refer" a reflection coefficient Γ_2 which is known at one position on a section of waveguide to a different position by rotating counter-clockwise the vector representing Γ_2 by the angle $2\beta l$, and multiplying the length of the vector by a factor $e^{-2\alpha l}$.

One can derive the conventional transmission-line formulas from (9), remembering that

$$z = \frac{Z}{Z_0} = \frac{1 + \Gamma}{1 - \Gamma}, \quad \text{and} \quad e^{\gamma l} = \cosh \gamma l + \sinh \gamma l. \tag{12}$$

One obtains

$$z_1 = \frac{z_2 + \tanh \gamma l}{1 + z_2 \tanh \gamma l}, \tag{13}$$

and for the special case of no losses,

$$z_1 = \frac{z_2 + j \tan \beta l}{1 + j z_2 \tan \beta l}. \tag{14}$$

Instead of using the above formulas, it is simpler to plot z_2 or Γ_2 on the well-known Smith Chart, rotate the vector representing Γ_2 by the angle $2\beta l$, and change the magnitude by the factor $e^{-2\alpha l}$ to plot Γ_1. Then z_1 can be easily read from a chart of moderate size with moderate precision. In order to obtain greater precision in certain cases, the central region of the chart is often expanded. A rotating and sliding scale is also used to facilitate rotating Γ_2 and making appropriate changes in its magnitude.

Measurement Methods

A number of measurement methods and instruments have been developed for measuring impedance in uniconductor waveguides, and each has its particular advantages and uses. In this paper, most of them are only briefly mentioned in passing, in order to devote more space to two selected ones, the slotted-line standing-wave machine and

the tuned reflectometer. However, references are given to enable the reader to investigate details of other methods. In addition, much valuable information is available on request from instrument manufacturers.

The Idealized Slotted Section

In order to integrate the concept of VSWR with *generalized* voltage v on uniconductor waveguides, the theory of the idealized slotted-section standing-wave machine is briefly reviwed.

The propagation of certain waveguide modes is not greatly disturbed if a thin slot is made in the outer conductor of the waveguide in a direction parallel to the axis of the waveguide, Examples of these modes are the TEM mode in coaxial line, the TE_{10} mode in rectangular waveguide, and the TE_{11} mode in waveguide of circular cross section.

A probe may be inserted in the slot to sample the fields, and movement of the probe along the slot will provide data on the variation of the field strength as a function of probe position.

Probes may sample either the electric or the magnetic field, or both, and may sample either transverse components, axial components, or both. The usual type of probe encountered in commercially available instruments is one which samples the transverse component of the electric field and does not respond to the magnetic field or to axial components of the electric field (if present).

Ideally, the output level $|b_3|$ of the probe to the detector is proportional to the strength of the field sampled. For example, if the transverse component E_t of the electric field is sampled,

$$|b_3| = K_1|E_t| = K_2|v| = K_2|a + b| = K_2|a| \cdot |1 + \Gamma|, \quad (15)$$

where K_1 and K_2 are constants. Thus, the probe samples the voltage standing wave which is set up as the sum of the oppositely directed traveling waves having amplitudes a and b.

Suppose that the probe is located at terminal surface 1, whose position is variable, and the load Γ_L is at terminal surface 2, whose position is fixed. The probe output level is

$$\begin{aligned} |b_3| &= K_2|a_1| \cdot |1 + \Gamma_1| \\ &= K_2|a_2|e^{\alpha l}|1 + |\Gamma_L|e^{(j\psi_L - 2\gamma l)}|. \end{aligned} \quad (16)$$

In the idealized slotted line, which is lossless ($\gamma = j\beta$, $\alpha = 0$),

$$|b_3| = K_2|a| \cdot |1 + |\Gamma_L|e^{j(\psi_L - 2\beta l)}|. \quad (17)$$

As the probe changes position, the smaller of the two vector components shown in Fig. 2 rotates, and goes in and out of phase with the fixed vector component of unit magnitude.

Thus the ratio of maximum-to-minimum probe output is

$$\frac{|b_3|_{\max}}{|b_3|_{\min}} = \frac{1 + |\Gamma_L|}{1 - |\Gamma_L|} = \text{VSWR}. \quad (18)$$

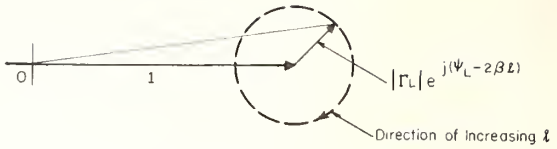


Fig. 2. Vector diagram of probe output level.

One sees from the diagram that, as the probe moves in a direction away from the load, the first minimum in its output occurs when $l = l_m$, and l_m satisfies the following relaship:

$$\psi_L - 2\beta l_m = \pm \pi. \quad (19)$$

The phase angle ψ_L of the reflection coefficient of the load is then given by

$$\psi_L = 2\beta l_m \pm \pi. \quad (20)$$

The Actual Slotted Line

Although the idealized slotted section of waveguide is simple, any actual slotted section is quite complicated, and a rigorous analysis of its performance is difficult. Errors in slotted-line measurements of impedance are discussed in another paper [2] in this issue. See also [4]–[9].

Measurements of Reflection Coefficient in Waveguide of Various Cross Sections

Usually, slotted sections and traveling-probe standing-wave machines are widely available in standard sizes of waveguide having rectangular cross section. When it is desired to use such an instrument to make measurements of reflection coefficient in a waveguide having a different cross section, an adapter is often used. One can either determine the characteristics of the adapter and calculate their transforming effect on the reflection coefficient, or employ a tuner [10] to avoid this procedure. The latter procedure may be more convenient and is described as follows.

As shown in Fig. 3, a nonreflecting load is connected to section 2. (Such loads usually are not readily available, but can be made by suitably tapering some appropriate absorbing material.) The tuner is then adjusted so that there are no standing waves indicated by the traveling probe. This makes $S_{11} = 0$ for the two-port[4] representing the adapter. To a good approximation, $S_{22} = 0$, also, if the adapter is nearly lossless. One then replaces the nonreflecting load by a short circuit, noting the probe position (reference plane) at which the probe response is minimum. The two-port bounded by this plane in the slotted section and the plane of the short circuit terminating section 2 then has $S_{21} = 1$, to a very good approximation. One next replaces the short circuit by the device under test. Since, in general,

$$\Gamma = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L},$$

[4] For discussion of scattering coefficients S_{11}, S_{12}, S_{21}, and S_{22} of two-ports, see Kerns and Beatty [3].

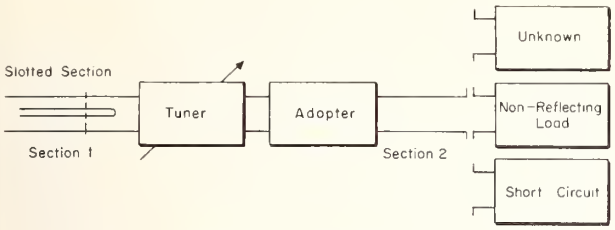


Fig. 3. Tuning an adapter.

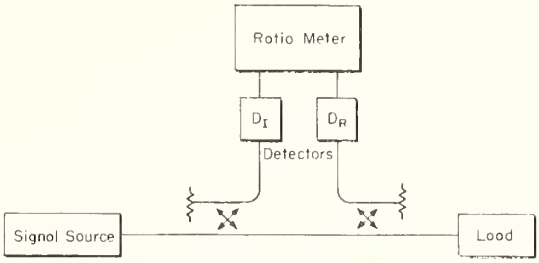


Fig. 4. Conventional reflectometer.

it follows that the measured reflection coefficient Γ (referred to the plane so located in the slotted section) equals quite closely the reflection coefficient Γ_L at the desired location in section 2 due to the device under test.

The error due to losses in the adapter has yet to be analyzed, but is likely to be small. It is told later how this type of error can be avoided [11] by the use of a hybrid tuned reflectometer.

Reflectometer Techniques in General

It is commonly understood that a reflectometer is an instrument used to determine the magnitude $|\Gamma|$ of the reflection coefficient, or the corresponding VSWR. Phase information is usually not provided, but can be obtained by appropriate modifications of the instruments and techniques [5]. The use of reflectometers in swept-frequency measurements is an important subject, and is covered elsewhere in this issue in the paper [12] entitled "Swept-Frequency Techniques." With equipment presently available, one can expect to make swept-frequency measurements of $|\Gamma|$ over the range 0.1 to 0.35 with an uncertainty less than 5 percent at frequencies up to 140 GHz.

Evolution of the Tuned Reflectometer at NBS

The tuned reflectometer, for single-frequency operation, offers the highest accuracy today in the measurement of reflection-coefficient magnitudes. Reasons for this are probably best introduced by tracing the evolution of the tuned reflectometer at the National Bureau of Standards. Similar developments carried out in other laboratories will also be mentioned.

Background Prior to 1955

It was recognized at NBS at an early date that the conventional slotted line had inherent limitations such as the irregularities in probe response associated with a probe traveling in a slot, and the discontinuities at the end of the slotted section caused by the end of the slot, transition section, or taper, etc. (as well as other limitations). It was hoped that other methods could be found which would avoid these limitations without introducing other ones which might be equally serious. A number of methods were considered, and although there are applications where the advantages of each are important, many did not appear to be adaptable to ultimate refinement for highest accuracy.

The resonance line technique of Chipman [13], [14] was investigated because it avoided problems connected with the slot, but there remained difficult problems in maintaining constant probe coupling as the plunger moved inside the waveguide.

Rotating probes as introduced by Tischer [6], [9], [15] were located in the top wall of the waveguide and also eliminated problems connected with the slot. However, they introduced a fixed discontinuity at the point of measurement and although probe loading problems were simplified, they were not eliminated. Use of two or more fixed probes spaced out along the waveguide axis as suggested by Samuel [16] avoided the slot, but did not avoid loading of the waveguide by the probes, and other fairly serious problems were introduced.

Reflectometers employing directional couplers were investigated because slots and probe loading effects were avoided. The conventional reflectometer shown in Fig. 4, as employed for swept-frequency measurement of VSWR, is a valuable instrument, but does not permit attainment of extremely accurate results, even at a fixed frequency, because of the limited directivities and the residual reflections in the directional couplers. As early as 1946, Korman (of RCA) showed how to adjust a reflectometer to obtain ideal response [17]. He adjusted it until the response remained constant as a short circuit was slid along the output transmission line. The use of a tuner to overcome the finite-directivity limitation of directional couplers was described by Barnett [18] (of the Hewlett-Packard Co.). A phasable load having a small reflection coefficient was moved back and forth in the output waveguide while adjusting a tuner until the side-arm output of the directional coupler remained constant. This was the same procedure used by Pomeroy (of the Western Electric Co.) and later by Engen and Beatty (of NBS).

Barnett was concerned with reflections in the reflectometer system and reduced them by attention to design. Later, Hunton and Pappas [19] (of the Hewlett-Packard Co.), used a sliding short circuit in the output waveguide of the reflectometer to separate the effects of the sources of reflection. They suggested a technique for avoiding the errors from these reflections by taking the mean of maximum and minimum readings of the output meter as the short-circuit slides, and setting the ratio meter at unity for this condition. They noted that the reflectometer compared a small quantity to a large one when measuring small reflections, while in the slotted-line technique the ratio of two large quantities was observed. Thus they concluded that the reflectometer technique was potentially more accurate than the slotted-line technique for the measurement of small reflections.

385-582 O - 70 - 20

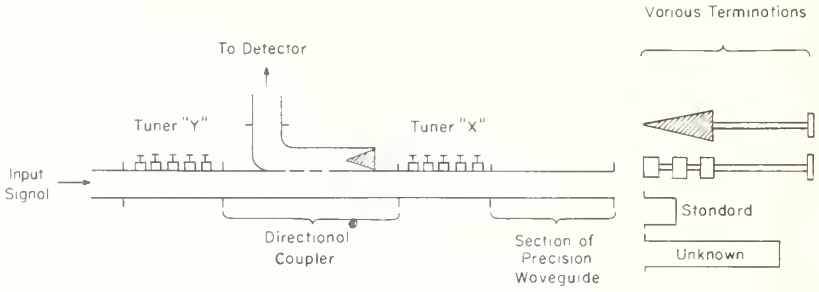


Fig. 5. Tuned-reflectometer arrangement.

First Approach—Phasable Load Technique

The first really promising and stimulating approach to high-precision measurement of VSWR was made by MacPherson and Kerns [20] (of NBS). They employed a three-arm waveguide junction (three-port) apparatus for measuring the VSWR of phasable loads. The analysis by Kerns of the three-port in terms of gathering coefficients[5] showed that results of a previous analysis [21] of probe loading effects in slotted lines could be applied more rigorously in this case than to the slotted line. A feeling for the precision attained with this technique is given by the following results. Six measurements of VSWR of a termination yielded a mean of 1.0166 with a standard deviation of the mean of 0.0003. Fixed (nonphasable) loads could be measured by employing a line stretcher, designed so that the discontinuities that it introduced were fixed with respect to the three-arm junction.

Phasable-Load Technique—With Tuners

In attempting to improve the accuracy of the phasable-load technique, Beatty analyzed the three-port in terms of scattering coefficients and obtained a simple equation for the amplitude b_3 of the wave incident on the detector [22]:

$$b_3 = c\,\frac{1 + K\Gamma_u}{1 - \Gamma_{2i}\Gamma_u}. \tag{21}$$

In terms of the scattering coefficients of the three-port,

$$K = \frac{\begin{vmatrix} S_{21}S_{22} \\ S_{31}S_{32} \end{vmatrix}}{S_{31}}, \tag{22}$$

and Γ_{2i} is the reflection coefficient of the equivalent source in the output waveguide lead. He noted that one could obtain a squared VSWR response [23] by using certain kinds of lossless junctions for which $K \approx \Gamma_{2i} \approx 1$. Both of these quantities could be adjusted by appropriate tuning of the junctions. He also noted that a magnified response could be obtained by adjusting $|K|$ to have large values. One type of waveguide junction (not lossless), the directional coupler, has an inherently large $|K|$, so that not much tuning was required. In addition, it was noted that a second tuner could adjust Γ_{2i}, and that the two adjustments were nearly independent.

[5] The gathering matrix is the reciprocal of the scattering matrix. See Kerns and Beatty [3].

This led to a technique [22] employing a directional coupler and two tuners in which K was adjusted for a detector null corresponding to $K\Gamma_u = -1$. Then Γ_{2i} was adjusted to vanish. The termination having a reflection coefficient Γ_u was then replaced by a standard having a known $|\Gamma_s|$. The line stretcher was moved to change the phase of the load and the ratio of maximum-to-minimum detector wave amplitude was observed. This ratio equalled

$$\frac{|\Gamma_u| + |\Gamma_s|}{|\Gamma_u| - |\Gamma_s|}.$$

The magnitude $|\Gamma_u|$ was then calculated. This technique was very successful in accurately measuring magnitudes of reflection coefficients, and very closely checked calculated values for half-round inductive obstacle [24], [25] standards of reflection coefficient.

The Tuned Reflectometer Evolves

A more convenient technique not requiring the use of a line stretcher evolved from a suggestion by Engen [26] that the tuners could be adjusted for the conditions $K = \infty$, $\Gamma_{2i} = 0$. Engen had developed tuning techniques for a reflectometer employing two directional couplers. He suggested that one achieve these conditions by sliding two different terminations inside a waveguide and tuning for constant detector output in each case. It is interesting that Pomeroy had already developed the same technique and had written internal reports on it (this was not learned at NBS until 1966) but did not publish an exact description of it. However, he did publish descriptions of similar techniques [27], [28]. Upon achieving the above conditions, the magnitude of b_3 is proportional to $|\Gamma_L|$, the magnitude of the reflection coefficient of the load. The measuring technique is indicated in Fig. 5 and consists of alternately connecting a standard and the unknown termination to the reflectometer and observing the ratio of the detector outputs. When the magnitude of the reflection coefficient of the standard is unity, the measured ratio, expressed in decibels, is the return loss of the termination under test. Tables of return loss versus $|\Gamma|$ are available [29]. Detailed accounts of tuning and measurement techniques and analyses of errors have been published [26], [30].

Beatty and Anson introduced the concept of a hybrid system [31] in which all of the equipment consisted of rectangular waveguide components except for a wave-

guide-to-coaxial adapter, an output section of coaxial line, and the sliding loads. The discontinuities introduced in going from one type of waveguide to another do not need to be separately compensated, but are automatically taken care of when tuning the reflectometer in the usual way. The hybrid reflectometer principle should be valuable [11] in making accurate measurements in strip line, circular waveguide, ridged waveguide, reduced-height rectangular waveguide, or other types of waveguide.

Later Developments

Refinements have been made in the tuned-reflectometer technique including a semi-automatic method [32] for tuning a reflectometer. If very accurate results are required, the tuning operations are a reasonable price to pay, but one would still prefer to avoid tedious adjustments, if possible. One way to avoid adjustments of the tuners without losing too much accuracy is to employ a section of waveguide having a length of one-quarter guide wavelength [33]. Measurements are made with this section first inserted between the reflectometer and the unknown and standard terminations. The measurements are then repeated with the quarter-wave section removed. Although it is not specifically mentioned in the reference, it is quite important that insertion of the quarter-wavelength section be accomplished without adding significant reflections. This is presently possible using high-precision coaxial connectors or waveguide joints, but cannot be done using ordinary connectors or joints. Comparison of results with the tuned reflectometer indicates that the accuracy is not degraded more than a few percent over a range of $|\Gamma|$ from 0.02 to 0.33. Thus at probably a small sacrifice in accuracy, one eliminates the tuning adjustments, but doubles the number of observations of data in a single measurement. At present, no complete error analysis of this technique has been made.

A tuned reflectometer has been constructed and used for measurements in the 60 to 90 GHz frequency range [34]. Special problems that were encountered and solved included the design and construction of suitable tuners, and frequency stabilization by phase locking of the oscillators.

The tuned-reflectometer technique has been further developed and is presently used to calibrate [35] standards of reflection coefficient based upon a "step" in the narrow dimension in rectangular waveguide. The accuracy obtainable[6] exceeds that obtained with the highest quality slotted sections of waveguide [36]. For example, a termination having a VSWR of 1.05 can be measured with a tuned reflectometer with an uncertanty less than 0.05 percent. It is also rather easy to accurately measure terminations having a high reflection. One simply compares the reflections from the unknown and a quarter-wavelength short circuit. The limiting source of error is caused mainly by the residual variation in $|b_3|$ which occurs when sliding the short circuit. If the residual variation is less than 0.02 decibel, the limit of uncertainty in the measured reflection coefficient of the short circuit due to imperfect tuning is approximately 0.25 percent. The uncertainty may be further reduced by improving the stability of the measurement system. At least an order of magnitude improvement is presently possible.

Reflection-Coefficient Standards

In order to produce a known reflection coefficient or normalized impedance in a given waveguide, one may use either fixed or variable standards. When the phase of the reflection is of no concern, they are sometimes called standards of VSWR or standard mismatches.

Fixed Standards

Most of the reflection-coefficient standards normally encountered are fixed and have either a low value of VSWR (1.5 or less) or a high VSWR such as from a short circuit.

In rectangular waveguide, many reflection-coefficient standards have taken the form of waveguide discontinuities of simple geometry. The reflection coefficient is calculated from the geometry. Examples of these are the step in the narrow dimension and the half-round inductive obstacle.

Some workers have used thin diaphragms (irises, slits, or windows) inside waveguides as standards of reflection coefficient. They are easily constructed and formulas are available for calculating their reflection coefficients. However, the formulas are derived assuming perfectly conducting, infinitely thin diaphragms, which can only be approximately realized in practice. It is difficult to assess the error in the reflection coefficient caused by departure from the assumed conditions. Therefore, this type of standard is presently not recommended for precise work.

Most commercially available reflection-coefficient standards employ a step in the narrow dimension (Fig. 6) formed by joining a reduced-height waveguide to one having standard dimensions. The VSWR is, to a good approximation, equal to the ratio of the narrow inner dimensions, and is insensitive to frequency. There is a frequency-dependent, second-order correction for the discontinuity capacitance, but this is small if the reflection coefficient is 0.2 or less, which is usually the case. The termination in the reduced-height waveguide has a small, unavoidable reflection. It is usually made phasable so that its effect on the total reflection can be taken into account and subtracted. Although these types of standards are convenient and have good accuracy and stability, there is no separation of the reflection due to the discontinuity from that due to the waveguide joint, and the effects of losses are neglected. More accurate standards can be constructed such as half-round inductive obstacles.

Half-round inductive obstacle reflection-coefficient standards [24] are shown in Fig. 7. The geometry is amenable to precise calculation and permits fabrication by electroforming as well as by machining. A geometry suitable for electroforming tends also to be suitable for calculation, because sharp edges in the interior of the electroformed product are avoided. In general, such edges would 1) require mathematical singularities in the representation of the electromagnetic field at an edge of an infinitely conducting

[6] For accuracy charts on microwave impedance measurements at NBS, see "Accuracy charts for RF measurements," by Wildhack, Mason, and Powers, in this issue.

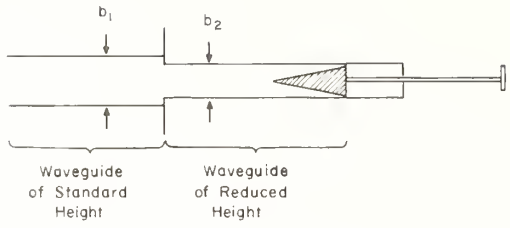


Fig. 6. Step in narrow dimension of waveguide.

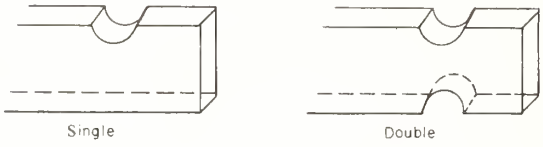


Fig. 7. Half-round inductive obstacles.

obstacle, 2) greatly complicate perturbation calculations needed to evaluate effects of finite conductivity, and 3) tend to exaggerate the effect of finite conductivity. Accurate calculated values of VSWR and $|\Gamma|$ are available for WR-90, WR-187, and WR-284 rectangular waveguide [25]. The reflection from the joint is distinct, and nonstandard waveguide is not required. Formulas are available [37] for corrections for finite conductivity of the metal. The reflection from this type of standard is frequency-sensitive.

Standards of low reflection have been constructed at NBS as shown in Fig. 8. At present there are no formulas for accurate calculation of reflections from the hole, which may be located in either the broad or narrow wall, and either on or off center. An approximate formula, based upon Bethe's theory [38], for the return loss L_R from a hole centered in the broad wall of a rectangular waveguide is

$$L_R = 20 \log_{10} \frac{6ab\lambda_G}{\pi d^3 \left[1 + \frac{1}{2}\left(\frac{\lambda_G}{\lambda}\right)^2\right]}, \qquad (23)$$

where d is the diameter of the hole, a and b are the inside width and height, λ is the wavelength, and λ_G the guide wavelength. Leakage from the hole can be decreased at will by increasing the length of the section of below-cutoff waveguide formed by the hole.

The below-cutoff hole in the waveguide may be easily fabricated to good accuracy. It is possible to slide either a tapered termination or a short circuit in the waveguide past the hole. Thus the reflections from the termination, the hole, and a waveguide joint may be easily separated [39]. It may be possible to plug the hole with a rod in such a way that the discontinuity will be temporarily removed.

Perhaps the most accurate type of reflection-coefficient standard is the short circuit, or more specifically, the quarter-wavelength section of short-circuited waveguide [22], [40] as shown in Fig. 9. This standard is frequency-sensitive, but is not sensitive to small dimensional errors, or to reflections or losses at the waveguide joint. This freedom from losses at the joint is a result of the vanishing of the axial component of current at the joint. Although the nominal reflection coefficient is unity, the actual reflection coefficient $|\Gamma_s|$ can be determined to a very good accuracy from a knowledge of the microwave conductivity of the waveguide walls. For example, an error of 4 percent in determining conductivity will result in an error of only 0.001 percent in $|\Gamma_s|$ when $|\Gamma_s| \approx 0.99950$. The conductivity is obtained from independent measurements. The reflection coefficient of a standard of this type has also been determined by two other independent methods, which gave close agreement with the calculated value. These methods are 1) measuring the Q of a cavity which incorporates the quarter-wavelength reflection-coefficient standard, and 2) measuring the losses of the standard by means of a microcalorimeter.

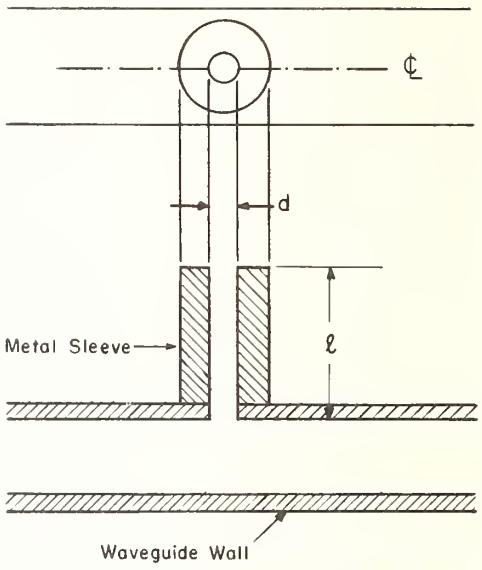


Fig. 8. Standard of low reflection for waveguide.

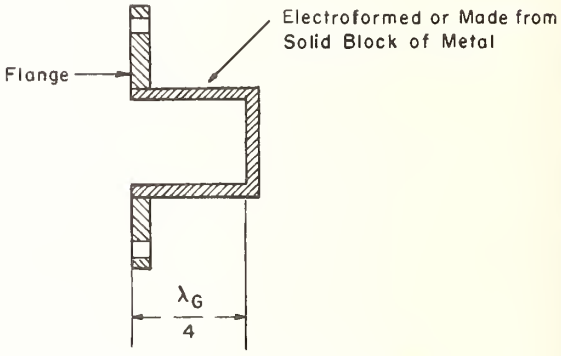


Fig. 9. Quarter-wavelength short circuit.

When using this type of standard to measure an unknown having a small reflection, the difference in return loss is large, say 60 dB, and consequently difficult to measure accurately. However, an adjustable load may be used as an intermediate source of reflections, making the measurement in two smaller steps, say 30 dB each.

An impedance standard having nominally zero reflection coefficient consists of a waveguide terminated in a load which can be adjusted [41] to have no reflection. In this

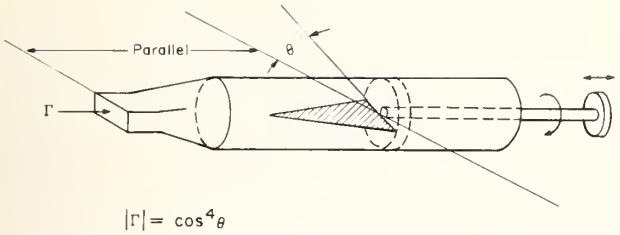


Fig. 10. Variable impedance standard employing rotary dissipative vane.

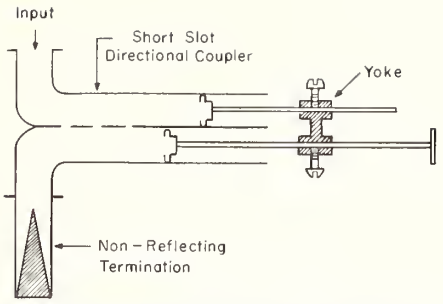


Fig. 11. Variable impedance standard employing directional coupler.

case, it is important that the waveguide have the same standard cross-sectional configuration and dimensions as the waveguide in which the reflection coefficients are to be observed. Such an impedance standard has a normalized impedance of unity.

Variable Standards

Several types of variable standards of reflection coefficient have been developed. One type which has been produced commercially [42], [43] consists of an absorbing strip attached to a sliding short circuit in waveguide of circular cross section as shown in Fig. 10. One rotates the strip to vary the magnitude of the reflection and slides the short circuit to vary its phase. The magnitude and phase angle obey calculable laws, so this device ordinarily needs no calibration unless especially accurate results are required. A tapered section designed to introduce little dissipative loss and little or no reflection couples the waveguide to one of rectangular cross section. It is presently possible to design transition sections which are compensated over the full frequency range of the waveguide [44] and have very low reflections.

Another type of variable standard which has been produced commercially consists of a short-slot 3-dB directional coupler having ganged sliding short circuits in two of the arms as shown in Fig. 11. In practice, one places a nonreflecting termination on one of the remaining arms and adjusts the relative position of the short circuits to obtain the desired magnitude of reflection coefficient. The short circuits are then clamped in this relative position and slid together to vary the phase of the reflection coefficient. The reflection coefficient of this device is not accurately predictable due to imperfections in the directional coupler, and therefore calibration is required. In principle, the device can be adjusted to present any reflection coefficient but, in practice, one cannot quite achieve this objective.

Another type of variable standard, which has a wide range but must be calibrated, is shown in Fig. 12. It is based upon the adjustable termination for rectangular waveguide previously mentioned [41]. Its possibilities have been neither fully explored nor exploited.

A variable standard of reflection coefficient may be quickly assembled from a slide-screw tuner and a nonreflecting (fixed) termination, which are readily available in many laboratories. If the tuner is driven by a micrometer, the standard can be precisely set and calibrated.

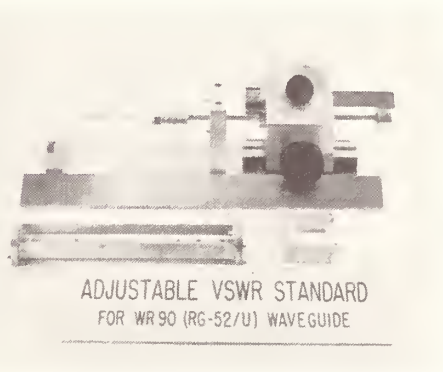


Fig. 12. Variable impedance standard employing adjustable termination.

Conclusions

Measurements of VSWR and of magnitude of reflection coefficient $|\Gamma|$ in uniconductor waveguide can today be made with least uncertainty using tuned-reflectometer techniques. The price that one pays for high accuracy in this case is tedious tuning procedure. Some work has been done to reduce the tedium without sacrificing accuracy, but more remains to be accomplished. Another challenge for future RF metrologists is to achieve the same accuracies with swept-frequency systems as are obtained at fixed frequencies with tuning techniques. The measurement of the phase of the reflection coefficient has been of less concern, but is becoming of greater interest. It is presently best determined using refined versions of standing-wave machines. However, it is conceivable that more accurate phase determinations might be made in the future using modified reflectometer techniques.

A number of types of fixed or variable standards of reflection coefficient have been devised. Of the fixed standards, the quarter-wavelength short-circuited section of waveguide and the half-round inductive obstacles give the least uncertainty. Of the variable standards, the rotating strip in circular waveguide gives a wide range and follows a calculable law. Usually, fixed standards may be constructed to have smaller uncertainties than variable standards. Some of these standards such as the below-cutoff circular hole in rectangular waveguide need a more rigorous analysis leading to more accurate calculations of their reflection.

Standards of reflection for uniconductor waveguides having other than rectangular cross section apparently have not been designed or manufactured for commercial use, and have received little if any attention, to date. Although the need for accurate measurements in these other types of waveguide does not presently appear to be great, standards of reflection coefficient for waveguide having circular, or some other cross section, will eventually be needed.

Acknowledgment

Helpful discussions with D. M. Kerns are gratefully acknowledged, as well as suggestions from the reviewers, B. C. Yates, W. J. Anson, H. M. Altschuler, and B. P. Hand.

References

[1] D. M. Kerns, "Definitions of v, i, Z, Y, a, b, Γ, and S," this issue.

[2] R. L. Jesch and R. M. Jickling, "Impedance measurements in coaxial waveguide systems," this issue.

[3] D. M. Kerns and R. W. Beatty, *Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis*, New York: Pergamon, 1967. This work contains additional theoretical background.

[4] E. M. Wareham, "The slotted section standing wave meter," *J. Brit. IRE*, vol. 15, pp. 539–564, November 1955.

[5] R. W. Beatty, "Microwave impedance measurements and standards," NBS Mono. 82, p. 32, August 12, 1965.

[6] A. B. Giordano, "Measurement of standing-wave ratio," in *Handbook of Microwave Measurements*, M. Sucher and J. Fox, Eds. New York: Polytechnic Press, 1963, Ch. II, pp. 73–133.

[7] S. Hopfer, "Precision measurements with slotted sections," *PRD Repts.*, vol. 2, October 1953.

[8] E. L. Ginzton, *Microwave Measurements*. New York: McGraw-Hill, 1957.

[9] F. J. Tischer, *Mikrowellen-Messtechnik*. Berlin: Springer-Verlag, 1958. See ch. 4, "Anpassungs- und Impedanzmessung," pp. 73–197.

[10] H. F. Mathis, "Measurement of reflection coefficients through a lossless network," *IRE Trans. on Microwave Theory and Techniques (Correspondence)*, vol. MTT-3, p. 58, October 1955.

[11] R. W. Beatty, "Measuring impedance through an adapter without introducing additional error," *Proc. IEEE (Correspondence)*, vol. 53, pp. 656–657, June 1965.

[12] P. C. Ely, Jr., "Swept-frequency techniques," this issue.

[13] R. A. Chipman, "A resonance curve method for the absolute measurement of impedance at frequencies of the order of 300 Mc," *J. Appl. Phys.*, vol. 10, pp. 27–38, January 1939.

[14] A. S. Meier and W. P. Summers, "Measured impedance of vertical antennas over finite ground planes," *Proc. IRE*, vol. 37, pp. 609–616, June 1949.

[15] S. Hopfer and L. Nadler, "Waveguide rotary standing-wave indicators," *PRD Repts.*, vol. 6, pp. 1–6, January 1959.

[16] A. L. Samuel, "An oscilloscope method of presenting impedances on the reflection-coefficient plane," *Proc. IRE*, vol. 35, pp. 1279–1283, November 1947.

[17] N. I. Korman, "Note on a reflection-coefficient meter," *Proc. IRE*, vol. 34, pp. 657–658, September 1946.

[18] E. F. Barnett, "More about the -hp- precision directional couplers," *Hewlett-Packard J.*, vol. 4, January–February, 1953.

[19] J. K. Hunton and N. L. Pappas, "The -hp- microwave reflectometers," *Hewlett-Packard J.*, vol. 6, September–October, 1954.

[20] A. C. MacPherson and D. M. Kerns, "A new technique for the measurement of microwave standing-wave ratios," *Proc. IRE*, vol. 44, pp. 1024–1030, August 1956.

[21] W. Altar, F. B. Marshall, and L. P. Hunter, "Probe error in standing-wave detectors," *Proc. IRE*, vol. 34, pp. 33–44, January 1946.

[22] R. W. Beatty and D. M. Kerns, "Recently developed microwave impedance standards and methods of measurement," *IRE Trans. on Instrumentation*, vol. I-7, pp. 319–321, December 1958.

[23] R. W. Beatty, "Magnified and squared VSWR responses for microwave reflection coefficient measurements," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-7, pp. 346–350, July 1959.

[24] D. M. Kerns, "Half-round inductive obstacles in rectangular waveguide," *J. Res. NBS*, vol. 64B, pp. 113–130, April–June, 1960.

[25] "Precision measurement and calibration—Electricity and electronics," in *NBS Handbook 77*, vol. 1, pp. 695/1–717/8, February 1, 1961.

[26] G. F. Engen and R. W. Beatty, "Microwave reflectometer techniques," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-7, pp. 351–355, July 1959.

[27] A. F. Pomeroy, "Precision measurement of impedance mismatches in a waveguide," *Bell Sys. Tech. J.*, vol. 26, pp. 446–459, July 1947.

[28] A. F. Pomeroy, "Evaluation of uncertainties in measurement of electrical echoes in waveguide lines," *Microwave J.*, vol. 5, pp. 72–77, November 1962.

[29] R. W. Beatty and W. J. Anson, "Table of magnitude of reflection coefficient versus return loss ($L_R = 20 \log_{10} 1/|\Gamma|$)," NBS Tech. Note 72, September 19, 1960.

[30] W. J. Anson, "A guide to the use of the modified reflectometer technique of VSWR measurement," *J. Res. NBS*, vol. 65C, pp. 217–222, October–December 1961.

[31] R. W. Beatty and W. J. Anson, "Application of reflectometer techniques to accurate reflection measurements in coaxial systems," *Proc. IEE (London)*, vol. 109B, pp. 345–348, July 1962.

[32] M. H. Zanboorie, "A semi-automatic technique for tuning a reflectometer," *IEEE Trans. on Microwave Theory and Techniques (Correspondence)*, vol. MTT-13, pp. 709–710, September 1965.

[33] W. E. Little and D. A. Ellerbruch, "Precise reflection coefficient measurements with an untuned reflectometer," *J. Res. NBS*, vol. 70C, pp. 165–168, July–September 1966.

[34] O. L. Patty, W. E. Little, and M. H. Zanboorie, "Waveguide connector measurements with a millimeter wave reflectometer," to be published. (The authors are at NBS Boulder, Colo.)

[35] R. E. Larson, "Microwave calibration techniques at the National Bureau of Standards," *Acta IMEKO Conf. Rec.*, Stockholm, Sweden, pp. 383–393, 1964.

[36] "Accuracy in measurements and calibrations, 1965," W. A. Wildhack, R. C. Powell, and H. L. Mason, Eds. NBS Tech. Note 262, June 15, 1965.

[37] D. M. Kerns and W. T. Grandy, Jr., "Perturbation theorems for waveguide junctions, with applications," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-14, pp. 85–92, February 1966.

[38] R. L. Kyhl, "Directional couplers," in *Technique of Microwave Measurements*, M.I.T. Radiation Lab. Ser., vol. 11. New York: McGraw-Hill, 1947, ch. 14.

[39] R. W. Beatty, G. F. Engen, and W. J. Anson, "Measurement of reflections and losses of waveguide joints and connectors using microwave reflectometer techniques," *IRE Trans. on Instrumentation*, vol. I-9, pp. 219–226, September 1960.

[40] R. W. Beatty and B. C. Yates, "Formulas and tables for quarter wavelength short-circuit impedance standards," to be published. (The authors are at NBS, Boulder, Colo.)

[41] R. W. Beatty, "An adjustable sliding termination for rectangular waveguide," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-5, pp. 192–194, July 1957.

[42] F. C. de Ronde, "Un élément simple pour mesurer des impédances en ondes centimétriques et millimétriques: La terminaison variable à réglages indépendants et à lecture directe du module et de l'argument du coefficient de réflexion," *Arch. Sci. (Geneva)*, vol. 10, fasc. spéc., 1957.

[43] C. W. van Es, M. Gevers, and F. C. de Ronde, "Waveguide equipment for 2 mm microwaves, I. Components," *Philips Tech. Rev.*, vol. 22, pp. 113–125, 1961.

[44] F. C. de Ronde, "Full-band matching of waveguide discontinuities," presented at the Internat'l Microwave Symp., Palo Alto, Calif., May 1966.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

Impedance Measurements in Coaxial Waveguide Systems

R. L. JESCH AND R. M. JICKLING, MEMBER, IEEE

Abstract—**This article is a tutorial review of impedance measurements and standards in coaxial waveguide systems propagating a TEM wave. It describes the development of coaxial air lines as impedance standards, reviews representative measurement methods, and discusses the errors and measurement techniques of the slotted line in detail.**

I. Introduction

ACCURACIES of impedance measurements in waveguide of coaxial cross section have been improved by a factor of ten in less than two decades, an achievement based primarily on the parallel development of precision coaxial-line standards and precision coaxial connectors.[1] The development of these coaxial-line standards and their contribution toward improved impedance-measurement systems will be reviewed in this paper. Improvements in bridge, slotted-line, and reflectometer measurement techniques that have taken place during this same period would not have been accomplished without concurrent improvements in precision connectors. These connectors have extremely low VSWR's and provide well defined reference planes.

To give some background on the early development of coaxial impedance measurements, the basic methods and techniques will be mentioned. These start with the method of measuring impedances by the use of slotted lines [1]–[3]. A single fixed probe can also be used by separating the probe from the load by a length of line containing a variable phase shifter allowing one to observe the standing-wave pattern [4] as it sweeps past the probe. Reflectometer systems using directional couplers [5] and hybrid junctions [6], operated at a single frequency, provide a way to measure the amplitude of reflection coefficient, but often without information about relative phases of the two traveling waves. Reflectometers operating at single frequencies are described by Beatty [7]. New methods employing swept-frequency techniques, described by Ely [54], can be used to obtain the amplitude and the phase of microwave parameters.

II. General Background

Coaxial lines can be used from dc to above 40 GHz, but are widely utilized beginning with the audio frequencies up through KU-band (18 GHz). Propagation is ordinarily restricted to the TEM mode. The wavelength at the operating frequency should be less than the mean circumference of the transmission line to prevent the propagation of higher modes. These higher modes, if generated at discontinuities in the system, are rapidly attenuated to an undetectable level within the distance of a diameter of the coaxial line. When good connectors are used, leakage is usually no problem in impedance measurements.

Transmission-line parameters are frequency-dependent. As the frequency is increased the skin effect becomes noticeable, i.e., current tends to become concentrated at the surfaces of the conductors. At high frequencies this current occupies a very small portion of the volume of the conductor, resulting in an increase in effective resistance. Skin effect also results in a decrease of inductance as the frequency is increased. Although in many instances these small variations in the parameters may be neglected, coaxial-line standards used for measurements and calibration of instruments must be corrected for these variations to keep the uncertainties of the standards sufficiently small.

Impedance in coaxial-line systems can be defined as the complex ratio of voltage v to current i. However, most measurements of high frequency and microwave impedance in coaxial lines are based on a sampling of the standing or traveling waves on the transmission line, rather than on voltage and current. Sinusoidal time variation and single frequencies are normally assumed.

The solution of the wave equations and the derivation of the transmission-line equations which are basic to impedance measurements are given, for example, by Everitt and Anner [8], Johnson [9], or Ramo and Whinnery [10].

When the transmission line is not terminated in its characteristic impedance Z_0, the power incident upon the load Z_L is reflected, as shown in Fig. 1(a). The ratio of the complex traveling-wave amplitudes, i.e., the reflection coefficient Γ, is related to Z_L and Z_0 by

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0}. \tag{1}$$

The incident and reflected waves interfere to form standing waves [Fig. 1(b)] which can be characterized by the ratio of the maximum to the minimum voltages (the voltage standing-wave ratio σ), where

$$\sigma = \frac{E_{\max}}{E_{\min}} = \frac{1 + |\Gamma|}{1 - |\Gamma|}. \tag{2}$$

Manuscript received March 31, 1967; revised April 6, 1967.

The authors are with the National Bureau of Standards, Boulder, Colo.

[1] The term "precision coaxial connectors" refers to those connectors which have very low reflection and meet other strict specifications of the IEEE Instrumentation and Measurement Group Technical Subcommittee on Standardization of Precision Coaxial Connectors. Many laboratory connectors in wide use today, although of low reflection, do not meet the specifications of the committee. For connector details, see the paper "Standardization of Precision Coaxial Connectors," elsewhere in this issue [14].

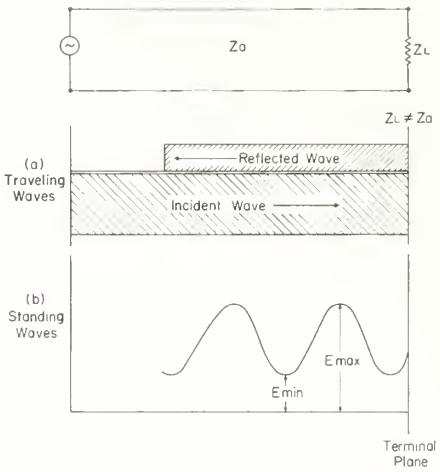


Fig. 1. Incident and reflected wave of a transmission line terminated in a mismatched condition showing the resulting standing wave.

In general, the impedance Z_p at any point along the line is related to the load impedance by the expression

$$Z_p = Z_0 \frac{Z_L + Z_0 \tanh \gamma l}{Z_0 + Z_L \tanh \gamma l}, \tag{3}$$

where

Z_0 = characteristic impedance of the transmission line,
Z_L = complex load impedance,
γ = complex propagation constant, $\alpha + j\beta = [(R + j\omega L) \cdot (G + j\omega C)]^{\frac{1}{2}}$ (α is the attenuation constant and $\beta = 2\pi/\lambda$ is the phase constant),
l = length of line between the point and load.

For the special case when line losses are ignored, $\alpha = 0$, $\gamma = j\beta$, (3) becomes

$$Z_p = Z_0 \frac{Z_L + jZ_0 \tan \beta l}{Z_0 + jZ_L \tan \beta l}. \tag{4}$$

The load impedance at the end of a transmission line can be calculated from the VSWR present on the line and the position of a voltage minimum with respect to the load. The above equation can be reduced to an expression for the load impedance Z_L in terms of the VSWR and the distance l:

$$Z_L = Z_0 \frac{1 - j\sigma \tan \beta l}{\sigma - j \tan \beta l}. \tag{5}$$

The load impedance, $Z_L = R + jX$, can be calculated directly from (5) which can be rationalized as follows:

$$R = Z_0 \frac{\sigma(1 + \tan^2 \beta l)}{\sigma^2 + \tan^2 \beta l} \tag{6}$$

$$X = Z_0 \frac{(\sigma^2 - 1) \tan \beta l}{\sigma^2 + \tan^2 \beta l}. \tag{7}$$

One of the more useful graphical aids for transmission-line computations is the Smith Chart [11]. This chart not only enables one to relate impedances to VSWR or reflection coefficient and position of a voltage minimum but is also useful for impedance transformations and the plotting of data.

III. Rigid Air-Spaced Coaxial Lines As Impedance Standards

Sections of rigid coaxial transmission line have been used as standards of impedance for a number of years. Polystyrene foam supported the center conductor of those early lines [12]. Later, incremental coaxial standards with solid polystyrene or teflon support proved to be useful [13].

Dimensional and dielectric-constant uncertainties limited the accuracy of those standards to a few tenths of a percent at best. However, with the improved machining techniques developed during the last ten years, the mechanical uncertainties in the manufacture of accurate tubing diameters have been reduced to less than a hundredth percent. The present-day precision air-spaced line standards are constructed from this tubing; the center conductors are normally supported at their ends by adjacent coaxial connectors or dielectric beads.

The accuracy of laboratory precision line standards also depends on a definable electrical length, a problem greatly simplified by the precision coaxial connectors developed during the past dozen years both in this country and abroad. Precision coaxial connectors are described in another paper [14] in this issue. The advent of general precision coaxial connectors has also simplified the problem of defining the reference plane for impedance measurements, as both inner and outer conductors of the coaxial line end in the same transverse mechanical plane. Improved standards depending on precision connectors are described in earlier articles [15], [16].

A. Characteristics of Coaxial Air-Spaced Lines

The equation for the characteristic impedance of a coaxial line, $Z_0 \cong 60 \ln (b/a)$, should be corrected for electrical skin depth and environmental conditions in order to be applicable to precision line standards. Since the characteristic impedance is defined as

$$Z_0 = \left(\frac{R + j\omega L}{G + j\omega C}\right)^{1/2} \text{ohms}, \tag{8}$$

the accuracy with which Z_0 can be determined will depend on the knowledge of R, L, G, and C, the resistance, inductance, conductance, and capacitance per unit length, respectively.

Stratton [17] has provided convenient approximate equations for R and L:

$$R \cong \left(\frac{f}{\pi}\right)^{1/2} \left[\frac{(\mu_a \rho_a)^{1/2}}{a} + \frac{(\mu_b \rho_b)^{1/2}}{b}\right] \text{ohms per meter}, \tag{9}$$

$$L \cong \frac{\mu_0}{2\pi} \ln \frac{b}{a} + \frac{1}{2\pi^{3/2} f^{1/2}} \cdot \left[\frac{(\mu_a \rho_a)^{1/2}}{a} + \frac{(\mu_b \rho_b)^{1/2}}{b}\right] \text{henrys per meter}, \tag{10}$$

where

a = outer diameter of the inner conductor, in meters,
b = inner diameter of the outer conductor, in meters,
μ_0 = permeability of free space,
μ_a = permeability of the inner conductor material,
ρ_a = resistivity of inner conductor material,
μ_b = permeability of the outer conductor material,
ρ_b = resistivity of the outer conductor material,
f = frequency in hertz.

For the great majority of electrical conductors, $\mu_a = \mu_b = \mu_0$. The resistance per unit length R is the sum of the resistance of the inner and outer conductors and, due to skin effect, is proportional to $(f)^{1/2}$. Equation (9) is valid to an uncertainty of a few percent at 100 kHz and above. This uncertainty, however, is not a serious obstacle to accurate characteristic impedance calculation because of the small contribution of R to the value of Z_0.

The first term in (10) is the zero skin-depth form of coaxial inductance, the external inductance per unit length. The second and third terms comprise the internal inductance of the conductors and are frequency-sensitive because of skin effect.

The calculation of the exact solutions of R and L, published by Russell [18] in 1909 in terms of a series of Bessel functions, is more difficult and reasonably requires the use of a computer. In view of the uncertainty in the physical dimensions of the conductors, Stratton's approximations are quite satisfactory above 1 MHz [19].

The dielectric loss, $G = \omega C \tan \delta$, in the medium (air) between the conductors is negligibly low below approximately 80-percent relative humidity.

The capacitance per unit length can be calculated from the physical dimensions by the exact equation

$$C = \frac{2\pi\varepsilon}{\ln b/a} \text{ farads per meter,} \tag{11}$$

where

$\varepsilon = k_\varepsilon \varepsilon_0$
k_ε = relative dielectric constant of the medium (usually air)
$\varepsilon_0 \cong 8.8542 \times 10^{-12}$ farads per meter.

The value of k_ε may be computed from $k_\varepsilon = n^2$, where n is the refractive index of air [20]:

$$(n-1)10^6 = \frac{103.49}{T} p_1 + \frac{177.4}{T} p_2 + \frac{86.26}{T}\left(1 + \frac{5748}{T}\right) p_3. \tag{12}$$

The total barometric pressure is the sum of the partial pressures p_1, p_2, and p_3 of dry carbon-dioxide-free air, carbon dioxide, and water vapor, respectively, at absolute temperature T. For all but the most precise work, the effect of carbon dioxide may be neglected and p_2 set equal to zero. A handbook [21] gives p_3 for a known value of relative humidity. The value of k_ε at 760 mm, 50-percent relative humidity is approximately 1.000649: this reduces to 1.000558 at 628 mm (Boulder, Colo.).

In air-spaced lines when the skin depth is insignificant compared with the dimensional uncertainties in the conductors, (8) becomes the lossless case, $Z_0 = (L/C)^{1/2}$. Substituting the first term of (10) and (11) yields

$$Z_0 = \frac{2c \times 10^{-7}}{k_\varepsilon} \ln \frac{b}{a} = \frac{59.9585}{k_\varepsilon} \ln \frac{b}{a}. \tag{13}$$

Mechanical tolerances in the conductors cause an uncertainty in the calculated values of R, L, and C. Their effect on R may be disregarded. The effect on Z_0 of diameter tolerances Δa and Δb and center conductor eccentricity e can be estimated [22] by the sum of

$$\frac{\Delta Z_0}{Z_0} \cong \frac{59.96}{Z_0}\left(\frac{\Delta b}{b} - \frac{\Delta a}{a}\right) \tag{14}$$

and

$$\frac{\Delta Z_0}{Z_0} \cong -\frac{240e^2}{Z_0(b^2 - a^2)}. \tag{15}$$

MacKenzie and Sanderson have described the many practical difficulties in determining the effective diameters of the conductors; measurements are complicated by conductor ellipticity, triangularity, and lack of straightness. Although the surface finish resulting from high-quality machining practice may not interfere with the present state-of-the-art in air-gauging of conductor diameters, the measured value of characteristic impedance of a reference air line apparently does depend on the surface finish [16].

B. Applications of Standard Lines

Coaxial-line standards were used at Bell Telephone Laboratories, Inc., about 1950 to calibrate RF bridges [12]. These lines were constructed of 3/8-inch copper tubing and had the inner conductor supported at intervals with polystyrene foam. The frequency range of the bridges (30 MHz and lower) required long line standards.

Later Selby et al. [13] and other authors developed short sections of rigid line spaced with solid dielectric. These line sections, with an inner diameter of the outer conductor of 0.280 inch, approximately the same diameter as type N and type C cable connectors, were used for the study by the nodal-shift technique of the two-terminal pair network parameters of coaxial cable connectors.

Alford and Watts described a quarter-wavelength technique for the calibration of coaxial hybrids in which the standard was a section of air line of accurately known characteristic impedance [38].

With the development of improved machining techniques such as floating tools, cold forging on a precision steel mandrel, and electroforming, tubing finished to a uniformity of a few tens of microinches is now possible. Inner conductors can be constructed to these tolerances by centerless grinding. Air-spaced rigid coaxial standard lines can now be constructed to an accuracy for Z_0 of 0.05 percent or

better for improved impedance measurements [23], for precision bridge calibrations [24], slotted-line measurements [25]–[27], reflectometer measurements, and transcomparator calibrations [28]. Sliding-load techniques in coaxial systems also require precision lines [29], [30].

The calibration of the time-domain reflectometer is one of the more recent uses in which precision line standards have proved to be well suited [31]. For this special application, a short section of line is constructed with its outer conductor diameter the same as the outer conductor of the reflectometer system. Then a series of inner conductors of various diameters computed from (13) are ground, yielding a set of line standards of characteristic impedances ranging between 40 and 60 ohms.

IV. Bridge Methods of Impedance Measurements

Basically, in RF bridge techniques, an unknown impedance is compared with a known standard impedance. To be most useful, the standard must be variable in both magnitude and phase, not an easy accomplishment above a megahertz or so. Some measure of success has been achieved in the VHF region. For the progress on bridge circuits at the lower frequencies, see the paper, "Lumped Parameter Impedance Measurements," by Huntley and Jones, in this issue [32].

A. Modified Schering Bridge

A model of Sinclair's modified Schering bridge circuit using a coaxial terminal was devised by Soderman, extending the frequency utilization of this circuit to 165 MHz [33]. Careful design work to minimize the residual parameters of the bridge components resulted in a measurement uncertainty of $\pm$(2 percent + 1 ohm) in resistance for a range of zero to 200 ohms and $\pm$(5 percent + 2 ohms) in reactance over a range of -230 to $+230$ ohms at 100 MHz.

B. Twin-T Admittance Bridge

Woods extended the useful frequency range of the basic twin-T circuit. With a careful evaluation of the residual impedances in the various bridge arms, he was able to develop a precision instrument with which it is possible to measure complex admittances of any phase angle at frequencies up to 200 MHz to an uncertainty of $\pm$(0.2 percent + 0.005 pF) in capacitance and $\pm$(0.2 percent + 5 micromhos) in conductance. Conductance values of 0–50 millimhos and capacitance values of $\pm$50 pF over a frequency range of 3–300 MHz can be measured on the present model [34].

Of absolute necessity for this accuracy, Woods developed a precision set of rigid coaxial-line standards of characteristic impedance matching the low-impedance (24.3-ohm) terminals of the bridge and designed with laboratory connectors. A new scaled-down model designed to reach 500 MHz is under development.

C. Byrne Circuit

A bridge-like nulling method for the determination of impedance in the 5- to 500-MHz range designed by Byrne [35] is shown in Fig. 2. When power is applied to the circuit,

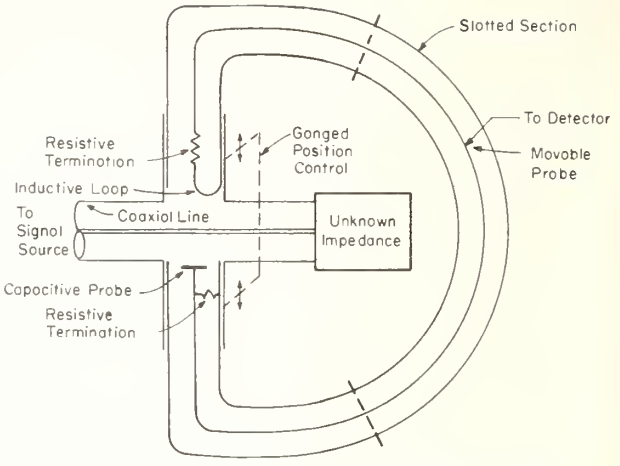


Fig. 2. Schematic diagram of the Byrne impedance bridge.

the inductive and capacitive probes, located in the coaxial line as near the unknown terminal as possible, respond to the magnetic and electric fields, respectively. The voltage induced in the inductive probe is proportional to the current flowing past the sampling point, while the voltage induced in the capacitive probe is proportional to the voltage at this point. The ratio of these two voltages is proportional to the impedance magnitude at the sampling plane and the phase difference between these two voltages is proportional to the impedance phase.

In its practical application [36], the inductive and capacitive probes are physically coupled as shown and the impedance magnitude is read directly from an appropriate front panel scale. The phase angle probe is also coupled to a front panel display. A correction for the length of coaxial line between the sampling plane and the actual unknown terminal is performed using (4), or more conveniently an impedance chart such as Carter ($Z-\Theta$) [3] or Smith [11].

Air-spaced 50-ohm coaxial lines in both open and short-circuited configuration are used in the initial calibration of this instrument. An uncertainty in impedance measurements of about 1 to 2 percent can be realized in practice with a commercial model of this bridge. Because of its wide impedance range, 2–2000 ohms and -90 to $+90$ degrees above 50 MHz, and its ease of balancing, the bridge is a useful tool for studying the electrical characteristics of an impedance in the VHF region. By the use of precision connectors and the time-domain reflectometer as an exploratory tool (see Section VI), higher accuracy up to 1.3 GHz is expected from a laboratory model now being developed.

D. Thurston Circuit

Thurston originated the rapid, convenient, direct-reading null device diagramed in Fig. 3. The instrument consists of a T-junction of four arms, the first terminated in the unknown admittance, the second terminated in a conductance standard $G_s = 1/Z_0$, and the third terminated in a susceptance standard (an eighth-wavelength shorted line) $jB_s = -j/Z_0$. The signal source from the fourth arm drives the three admittance arms in parallel; therefore, the current in each

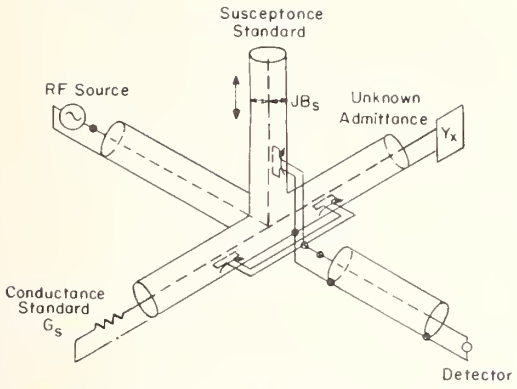


Fig. 3. Principal elements of Thurston admittance comparator.

line is proportional to the admittance of that line. The voltage induced in each of three adjustable current coupling probes is proportional to the current in the corresponding line and its magnitude depends on the orientation of the probe. The three probes are connected in parallel. When the probes are properly oriented, the combined output at the detector is zero, and the admittance components, G and B, can be read directly from the calibrated scales.

The improved version of this instrument described by Soderman [37] has a direct-reading range of 40 to 1500 MHz with an uncertainty of $\pm$(3 percent + 0.2 millimho) up to 1 GHz for both conductance and susceptance arms from 0 to 20 millimhos, and of $\pm(3\sqrt{M}$ percent + 0.2 millimho) from 20 to ∞ millimhos, where M is the scale multiplier. Above 1 GHz the series inductance of the common junction slowly increases the uncertainty to a value of $\pm$5 percent at 1.5 GHz.

E. Coaxial Hybrid Junction

Basically the four-arm circuits of the hybrid junction act as a bridge in the sense that if the impedances connected either to arms 1 and 2 or 3 and 4 are identical, zero power transmission will result between the remaining two arms. A family of bridges using hybrid characteristics was developed [38] in which the balance is independent of frequency. A wideband coaxial hybrid was also found to have a useful application in the development of an RF impedance plotter [55].

V. Slotted Lines

Most commercially available slotted lines use electric field probes. The variation of the voltage with probe position can be shown graphically, as in Fig. 4. As the vector Γ rotates uniformly with the angle $2\beta l$, where β is the propagation constant and l is the displacement of the probe from a reference plane, it can be seen that the ratio of maximum-to-minimum probe output (VSWR) is

$$\sigma = \frac{E_{max}}{E_{min}} = \frac{1 + |\Gamma|}{1 - |\Gamma|}.$$

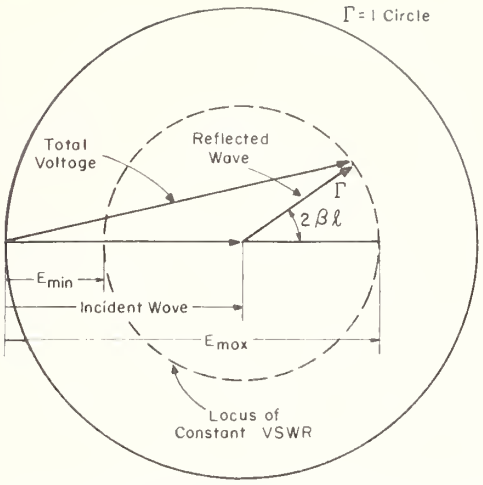


Fig. 4. Vector diagram on Smith Chart representing the probe output of an idealized slotted line.

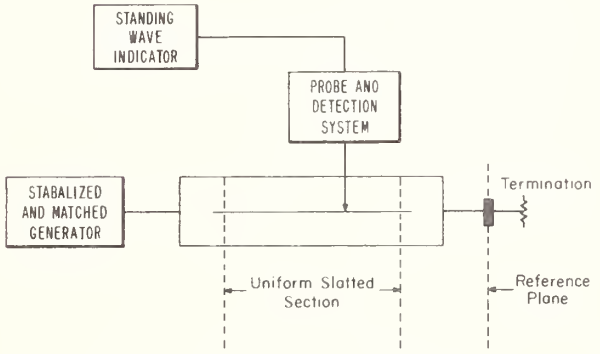


Fig. 5. Slotted-line measurement system.

A. Instrumentation

The main components of the apparatus which make up a slotted-line measurement system are shown in Fig. 5. An isolator helps to ensure that the frequency and output level of the signal source will not shift with changes in load impedance, while a low-pass filter rejects harmonics that may be generated in the signal source along with the fundamental frequency.

Depending upon application and accuracies required, one has a choice between slotted lines equipped with regular coaxial connectors and the coaxial slotted line equipped with the precision coaxial connector. Certain desirable characteristics of a coaxial slotted line include the following. The dimensions of the conductors must be held to very close tolerances, including the slot width. The slotted line should be equipped at the measurement end with a coaxial connector that has mechanical dimensions and electrical characteristics precisely compatible with those of the slotted line. The carriage assembly must be capable of moving the probe so that the travel of the probe's tip and the axis of the midpoints between the inner and the outer conductor of the slotted line are accurately parallel.

B. Slotted-Line Errors

Certain inherent errors in slotted lines must be considered if these lines are to be used for accurate impedance measurements. The most common sources of error are caused by the residual VSWR, probe effects, and line losses.

Residual VSWR is a measure of the standing waves present on a slotted line that is terminated by a standard nonreflecting load and excited by a single frequency from a nonreflecting source. This residual VSWR may be a result of discontinuities caused by variations in characteristic impedance, slot effects, or the connector on the end of the slotted line. The residual VSWR usually increases with frequency. Three measurement techniques that are useful in determining the residual VSWR of the slotted line are described.

1) Quarter-Wavelength Technique: Residual VSWR can be determined using quarter-wavelength precision reference air lines [27], [38]. One terminates the slotted line, as closely as possible, with a matched load and measures the impedance of the load with the slotted line. This will yield a reflection coefficient Γ_1 equal to the sum of the residual reflection coefficient of the slotted line Γ_b, plus the reflection coefficient of the load Γ_a:

$$\Gamma_1 \cong \Gamma_a + \Gamma_b, \tag{16}$$

This result is plotted as point A in Fig. 6 on an expanded Smith Chart. A section of reference air line, an odd multiple of a quarter-wavelength long, is then inserted between the slotted line and the load. The insertion of the reference air line causes the reflection coefficient of the load to be inverted on the Smith Chart. The slotted line will now indicate a reflection coefficient Γ_2:

$$\Gamma_2 \cong \Gamma_b - \Gamma_a, \tag{17}$$

which is plotted as point B. Solving (16) and (17) for Γ_b yields

$$\Gamma_b \cong \frac{\Gamma_1 + \Gamma_2}{2}, \tag{18}$$

where Γ_b is the residual error in the slotted line. The deviation of the center of the Smith Chart from C, midway between points A and B, is the residual VSWR of the slotted line.

2) Sliding-Load Technique: This technique of determining residual VSWR involves the use of a sliding load [39] with low VSWR. Ideally, the slotted line should be terminated with a load having no reflection. The reflection of any discontinuity in the connector of the load is lumped in with the residual VSWR of the slotted line. Repeated VSWR measurements are made at various positions of the sliding load to determine the highest obtainable VSWR, σ_{max}. The maximum reflection coefficient $|\Gamma_{max}|$ can then be calculated from

$$|\Gamma_{max}| = \frac{\sigma_{max} - 1}{\sigma_{max} + 1} \tag{19}$$

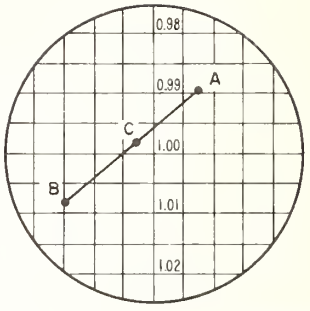


Fig. 6. Representation of residual VSWR on an expanded Smith Chart.

By returning the probe to the position of the original voltage maximum and readjusting the sliding load for minimum indicated output, a minimum VSWR σ_{min} is obtained. The minimum reflection coefficient $|\Gamma_{min}|$ can then be calculated from

$$|\Gamma_{min}| = \frac{\sigma_{min} - 1}{\sigma_{min} + 1} \tag{20}$$

The residual reflection coefficient is then

$$|\Gamma_r| \cong \frac{|\Gamma_{max}| \pm |\Gamma_{min}|}{2}. \tag{21}$$

The positive sign in (21) is chosen when the residual VSWR of the slotted line is greater than that of the load; the negative sign is chosen when it is smaller. When the relative magnitudes of the two VSWR's are not known, the load VSWR may be determined by one of the techniques described in Section V-F.

3) The Nodal-Shift Method: The third technique that can be used for determining the residual VSWR of a slotted line, but which is especially useful for correcting errors introduced by a coupling system, is the nodal-shift technique [40], [41]. The length of the slotted line is effectively changed by attaching a sliding short circuit or a different length of short-circuited coaxial lines. Referring to Fig. 7, if region II is terminated with either the sliding short circuit or one of the short-circuited coaxial lines, the voltage node of the standing wave can be located on the slotted line which is shown as region I. Electrical distances θ_1 and θ_2 are measured with respect to a reference plane. $\theta_1 - \theta_2$ is shown in Fig. 8 as a function of the line extension. From the nodal-shift curve, the residual-reflection coefficient of the slotted line is calculated by

$$|\Gamma_R| = \sin \frac{\pi D}{\lambda}, \tag{22}$$

where D is the peak-to-peak amplitude. The corresponding residual VSWR, σ_R, is then

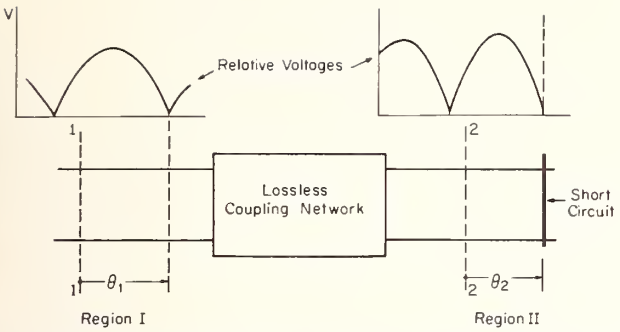


Fig. 7. Voltage standing-wave pattern on the two sides of the coupling network due to the short circuit in region II.

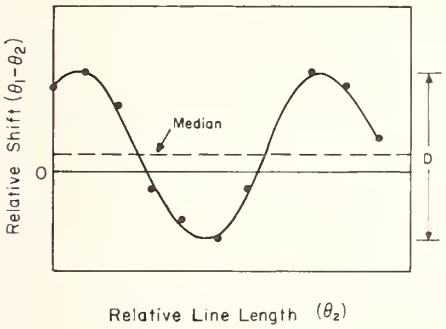


Fig. 8. Nodal-shift curve.

$$\sigma_R = \frac{1 + \sin\frac{\pi D}{\lambda}}{1 - \sin\frac{\pi D}{\lambda}} \cong 1 + \frac{2\pi D}{\lambda}, \tag{23}$$

where λ is the wavelength in the waveguide. Tischer improved the original nodal-shift method by utilizing mechanical coupling between the probe and short circuit [2]. An automatic method for obtaining data for the nodal-shift curve was developed by Beatty [42].

C. Probe Effects

1) Variation of Probe Coupling with Probe Position: This variation is the deviation from a "flat line" and, ideally, is indicated by the variation in probe output as a function of probe position when the slotted line is terminated in the characteristic impedance of the slotted portion and is excited by a nonreflecting source. This deviation is caused by imperfections in the slotted line which may be the result of lateral or vertical variation of the probe position with respect to the inner conductor, variation in the alignment or dimensions of the inner and outer conductors, or sag in the inner conductor. While these mechanical errors are independent of frequency, their effects on measured quantities are not. The deviation must be measured to determine the resulting uncertainty that will be present in the data measured with a slotted line. A plot of the deviation in probe coupling versus probe position of a typical precision slotted line measured at 1 kHz is shown in Fig. 9. The data were obtained by measuring with bridge techniques the capacitance between the probe and the center conductor. Such an investigation is quite helpful in determining which section of the line is most uniform for higher-frequency operation. A similar experiment should be repeated at the RF operating frequencies. Any additional deviation that may result from slotted-line losses will be observed. This check is accomplished by terminating the slotted line with a tunable load and adjusting for minimum probe output variation. Figure 10 shows a plot of the detected signal of the same typical slotted line well matched at 4 GHz. Once such a plot is made, a uniform portion of the slotted line should be selected for future impedance measurements. Even though the slope of the line will change with frequency, one can definitely see the correlation between 1-kHz and 4-GHz data.

2) Probe Loading of the Slotted Line: In the ideal situation, the probe penetrates the slotted line only very slightly for minimum disturbance of the fields within the slotted line. In practice, however, one must have sufficient coupling so that the required energy is extracted from the slotted line to operate the indicating system. It is recommended that one begin with more than adequate probe insertion and note the change in observed VSWR as the probe is gradually withdrawn in small increments. When the change in VSWR has become negligible compared to the desired measurement accuracy, the probe has been sufficiently decoupled. The effects of variation of probe coupling are much worse at the lower frequencies because of the decreased number of half-wavelengths falling within the linear portion of the slotted line.

The error caused by probe loading [6] has been estimated for the case when the probe dimensions are small compared with a wavelength and when the probe effect on the slotted section is that of a shunt admittance. It is also assumed that the generator is nonreflecting. A simple equivalent circuit, as shown in Fig. 11, is used to represent conditions at a reference plane in the slotted section at the probe position. Y_L' is the normalized admittance of the load at the same plane. It is assumed that there are no losses in the slotted section. The expression for the line voltage at the probe position is

$$V_p = e\frac{\dfrac{1}{Y_P + Y_L'}}{1 + \dfrac{1}{Y_P + Y_L'}} = \frac{e}{1 + Y_P + Y_L'}. \tag{24}$$

It is also assumed that the probe output voltage $|b_3|$ is proportional to $|V_P|$, so that (24) gives the probe response.

In the ideal situation, the probe admittance $Y_P = 0$, and

$$|b_3| = \frac{k}{1 + Y_L} = \frac{k}{2}(1 + \Gamma_L'), \tag{25}$$

where k is a constant. However, the probe admittance Y_P modifies the probe response so that the observed ratio of

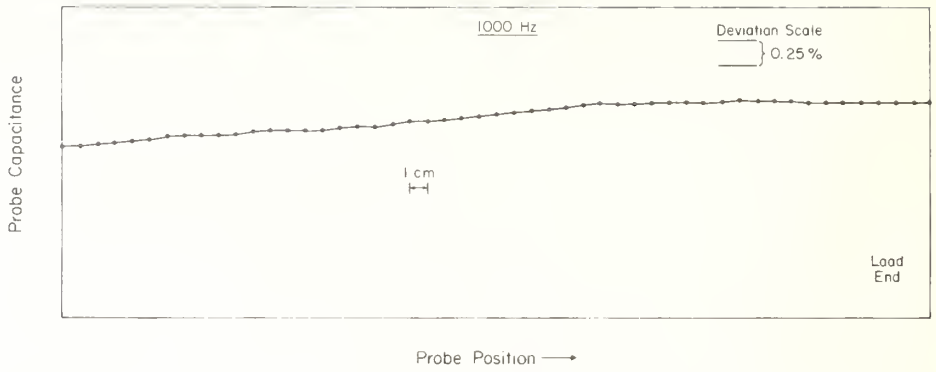


Fig. 9. Plot of the deviation in probe coupling of a typical precision slotted line measured at 1 kHz.

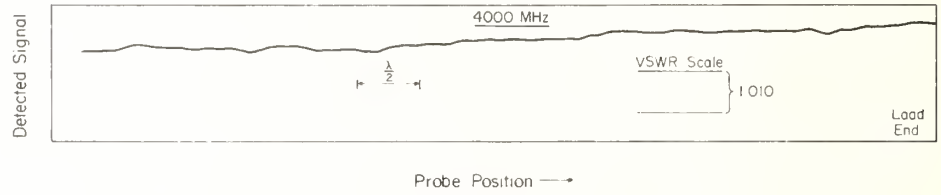


Fig. 10. Plot of detected signal of same line shown in Fig. 9 which is now terminated by a tunable load.

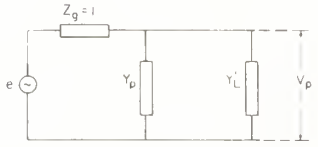


Fig. 11. Simplified equivalent circuit of a slotted line for analysis of probe loading effects.

maximum to minimum no longer equals the VSWR, and the distance of the first minimum no longer gives ψ_L by calculation. ($\psi_L = 2\beta l \pm \pi$ and is the phase angle of the load reflection coefficient.)

Calculated corrections for the effect of the probe admittance on measured VSWR and on the position of the voltage minimum and maximum can be found in Ginston [1], on page 244. To apply these corrections, Y_P must be known. Fortunately, a value for this probe admittance can be obtained by analysis of the response curve, since Y_P is difficult either to calculate or measure directly [43].

An analysis of the case when the generator is not matched to the slotted section has been made by Beatty [44]. He also gives an approach for the case when the probe dimensions are not small compared to a wavelength or cannot, for some other reason, be represented by a simple admittance shunted across the waveguide [44].

D. Line Losses

Measurements of higher VSWR's (> 10) can be seriously affected by line losses. In any slotted line, because of these losses, the amplitude of the standing wave decreases toward the generator end. Therefore, if ratios are taken of two adjacent voltage maxima to the minimum located between them, slightly different results will be obtained [45]. If the average of the two ratios is denoted by r, the VSWR, σ_L, of the load is approximately

$$\sigma_L \approx r\left[1 + \alpha l\left(r - \frac{1}{r}\right)\right], \tag{26}$$

where α is the attenuation in nepers per unit length and l is the distance from the terminal surface of the load to the minimum position used in the measurement.

E. Other Sources of Error

Other sources of error include detector characteristics deviating from the ideal, generator frequency and power instability, and inability to determine the voltage minimum of the standing-wave pattern accurately for the lower VSWR ranges [46].

F. Measurement Techniques

Proper choice of equipment and measurement techniques help to eliminate or reduce the random and systematic errors that degrade accuracy. Of the many techniques available, several will be described. Each technique has its own range of usefulness determined mainly by the magnitude of the impedance to be measured and the accuracy to be achieved.

For VSWR's of 1.2 or less, a slotted-line output recording system is recommended. Such a system graphically plots the standing-wave pattern, allowing one to view and choose the standing waves that have not been affected by variation in probe coupling of the slotted line. Figure 12 shows a slotted-line recording system. Another slotted-line recorder system is described by Sanderson [47].

If the detector requires frequent recalibration or if its square-law characteristics are not reliable, a method de-

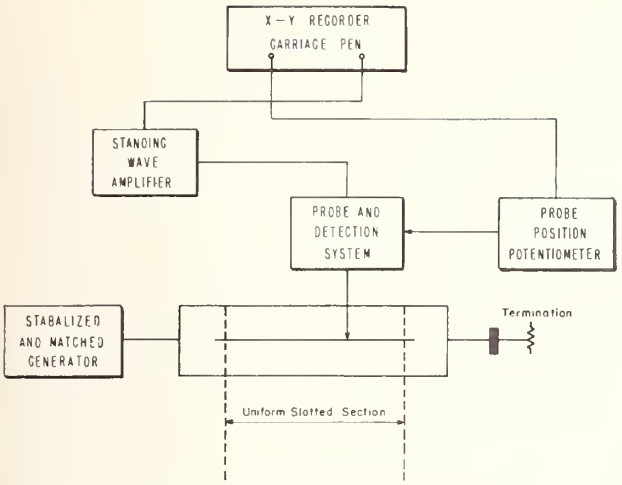


Fig. 12. Slotted-line recorder system.

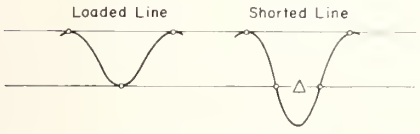


Fig. 13. Relative probe response of a loaded and short-circuited slotted line.

scribed by Winzemer [48] provides a way to determine the VSWR independent of the detector response. For VSWR's between 1.10 and 10, a useful approximate expression is given by

$$\sigma = \csc \frac{\pi \Delta}{\lambda}, \tag{27}$$

where Δ is the distance between probe positions having equal responses for a short-circuited slotted line. This Δ is measured at the original indicated level of the voltage minimum of the load as shown in Fig. 13. Other less simple expressions are described by Winzemer when the loaded voltage minimum is not available or a short circuit cannot be used.

For measurements of VSWR over ten, the "width-of-minimum" method is widely used [1]. The VSWR is given by

$$\sigma = \frac{\lambda}{\pi \Delta}, \tag{28}$$

where Δ is the measured distance between power points that are twice the amplitude of the minimum (3.01 dB above minimum). The assumption of square-law response of a crystal detector is usually adequate for this measurement if the probe signal is less than one millivolt, approximately. As discussed previously, unless line losses are corrected they can seriously affect the accuracy to which the VSWR can be measured.

A calibrated RF attenuator placed between the source and slotted line is the basis for an accurate method for measuring medium and large VSWR's that is also independent of the detector characteristics. The maximum-to-minimum voltage ratio can be measured directly in terms of the differential attenuator settings. A possible limitation is that more power output than is available from some sources may be needed in order to obtain the desired sensitivity.

The quarter-wave technique [27], employing precision reference air-spaced coaxial lines in conjunction with the Winzemer, "width-of-minimum," direct-recording, or other VSWR measurement methods, yields the highest attainable accuracy. The reflection coefficient Γ_a is given by

$$\Gamma_a = \frac{\Gamma_1 - \Gamma_2}{2}, \tag{29}$$

where Γ_1 is the measured reflection coefficient when the unknown termination is attached directly to the measurement end of the slotted line, and Γ_2 is the measured reflection coefficient when an odd multiple of a quarter-wavelength line section is inserted between the measurement end and the unknown termination.

The fixed-probe method of measuring small VSWR's of sliding loads is convenient and accurate. A slotted line is terminated in the sliding load and the probe is fixed in an arbitrary position. The maximum and minimum voltages are noted on a VSWR amplifier as the load is moved in its waveguide. The ratio of these voltages accurately yields the VSWR of the load despite the presence of residual VSWR.

The coupled sliding-load technique is also an accurate method for measuring small reflections [30]. The load is coupled mechanically to the slotted-line probe so that the two move together. The residual VSWR, which is caused by the reflection from the junction of the two lines, may be adjusted for a minimum by a slide-screw tuner. Subsequent measurements made with the slotted line at the same frequency indicate VSWR with respect to the reference air line. This method appears capable of making absolute impedance measurements on a low VSWR termination to a estimated error in reflection coefficient of 0.0012.

Another technique that has been used for measuring VSWR is the Tischer Modified Nodal-Shift Method [2]. Since mechanical coupling is used between the slotted-line probe and the sliding termination which is a short circuit, the method is related to the coupled sliding-load technique.

Once the VSWR has been measured by one of the above methods, the complex impedance of an unknown load can be determined, using a Smith Chart or (6) and (7), if the relative position of a voltage minimum with respect to a plane of reference is obtained.

The measurement of VSWR to an uncertainty of 0.1 to 1 percent and phase of reflection-coefficient magnitude from 0.1 degree to approximately 1 degree up to 8 GHz is possible when employing precision reference air-line techniques.

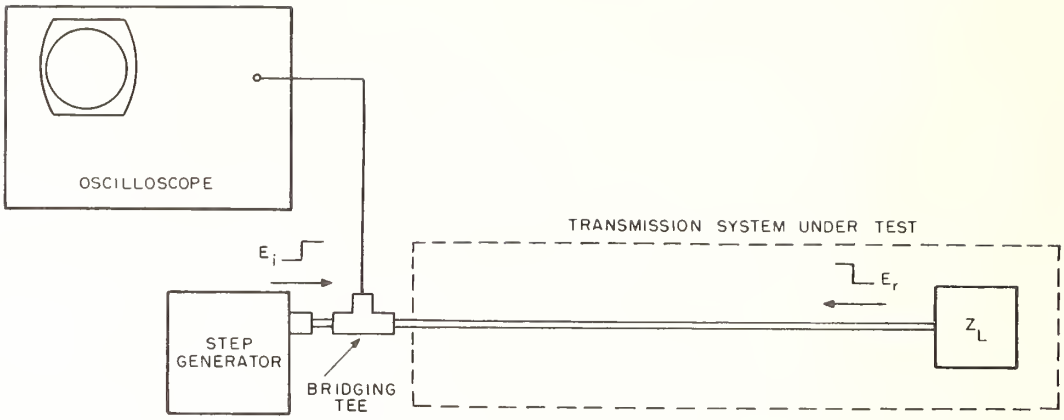


Fig. 14. Time-domain reflectometer.

VI. Time-Domain and Frequency-Domain Reflectometers

A. Time-Domain Reflectometer Systems

A good example of the present trend in RF measurements toward automation of techniques and rapid assembly of data covering a wide radio spectrum is shown in the recent development of a laboratory time-domain reflectometer (TDR) [49], diagramed in Fig. 14. Essentially, the TDR is a miniature closed-circuit radar set with an oscilloscope sampling the leading edge of both incident waves and any reflected waves at the bridging tee. Pulse techniques have long been used for the investigation of transmission-line discontinuities where the time scale has permitted measurements with microsecond pulses and megahertz bandwidths. In a laboratory system, where reflections may be separated by inches, nanosecond pulse generators and gigahertz sampling equipment are needed. The development of step-function generators with 30-ps rise times and sampling oscilloscopes with 12-GHz bandwidths now meets this need.

Small discontinuities separated by less than a centimeter can now be resolved in a transmission-line system. They are not only located physically, but each discontinuity, capacitive, inductive, or resistive, can be identified and quantitatively measured if desired. Where tuners can be inserted, the discontinuities may be removed directly one by one starting at the generator end. Shunt capacitance is added directly at the site of the discontinuity to eliminate the effects of series inductance, and conductor or dielectric material is removed to counteract shunt capacitance.

Since multiple reflections in a narrowband system prevent meaningful study with the TDR, it is more often used as a qualitative tool for the rapid adjustment of a wideband system. A quantitative measurement of capacitance can be performed by duplicating, with an adjustable calibrated capacitance probe in a uniform transmission-line section, the oscilloscope response to an unknown capacitance. And by careful comparison to line standards of incremental characteristic impedance, the impedance of transmission lines and terminations may be measured to 0.1 percent or better.

B. Frequency-Domain Reflectometers

Frequency-domain reflectometer techniques using auxiliary tuners [50] offer an accurate and sensitive means of measuring reflection coefficients. These techniques, originally developed in rectangular waveguide, can be applied to coaxial systems. An all-coaxial reflectometer has been designed and constructed to operate in the frequency range of 1–4 GHz [51]. A reflectometer method for measurements of coaxial components is also described by Spinney [52]. A hybrid reflectometer constructed of rectangular waveguide except for the coaxial precision waveguide and sliding loads has been used for coaxial measurements. The problems in setting up a coaxial system using reflectometer techniques were described by Beatty and Anson [29]. One such system was used for evaluating the laboratory and general precision coaxial connectors [51]. The VSWR of coaxial connectors has also been accurately measured by employing reflectometer swept-frequency techniques [53].

The evolution of the tuned reflectometer is outlined and recent techniques discussed in an article by Beatty, "Impedance Measurements and Standards for Uniconductor Waveguide," in this issue [7]. A detailed explanation of reflectometer techniques and applications is given by Beatty [44].

VII. Future Trends

New measurement systems need to be designed that offer high measurement accuracies and relative ease of operation, but still have enough system flexibility to make possible a wide frequency-range coverage. There are other areas where techniques and instrumentation are not keeping pace with the state-of-the-art. For example, the electrical length of precision lengths of coaxial air lines, used as absolute standards, can be calculated quite precisely from mechanical data, but cannot be measured to a high degree of accuracy, especially at the higher frequencies. A special

reflectometer system has been proposed that would have the capabilities of measuring these air-line lengths and the phase angle of other coaxial impedance standards to an uncertainty of 0.1 degree or less, but the cost of this system, for a wide frequency coverage, would almost prohibit its use for other than exacting standards work.

In the past several years, systems employing swept-frequency techniques have become quite popular in the measurement field. These are described by Ely [54], in "Swept-Frequency Techniques," in this issue. These systems enable one to obtain component and system measurements over a wide frequency range in a relative short period of time. Much can be done to improve the accuracies of these systems by the design of appropriate broadband standards and by development of better measurement techniques.

With the increased use of all coaxial components in the past several years, interest has been steadily growing in the field of miniature coaxial components employing connectors of the 3.5-mm and smaller sizes. A joint industry-NBS committee has been formed to study connectors that would permit optimum performance over a range of frequencies compatible with miniature line sizes up to 40 GHz.

Because of these new trends, the whole coaxial impedance field demands the development of improved standards, measurement techniques, and instrumentation.

Acknowledgment

Helpful discussions with R. W. Beatty are gratefully acknowledged, in addition to suggestions from A. J. Estin and R. C. Powell.

References

[1] E. L. Ginston, *Microwave Measurements*. New York: McGraw-Hill, 1957.

[2] F. J. Tischer, *Mikrowellen-Messtechnik*. Berlin: Springer-Verlag, 1958.

[3] F. E. Terman and J. M. Pettit, *Electronic Measurements*. New York: McGraw-Hill, 1952, pp. 122–192.

[4] A. C. MacPherson and D. M. Kerns, "A new technique for the measurement of microwave standing-wave ratios," *Proc. IRE*, vol. 44, pp. 1024–1030, August 1956.

[5] R. Levy, "Directional couplers," in *Advances in Microwaves*, vol. 1. New York: Academic, 1966, pp. 115–209.

[6] C. G. Montgomery, *Technique of Microwave Measurements*. New York: McGraw-Hill, 1947.

[7] R. W. Beatty, "Impedance measurements and standards for unicon-ductor waveguide," this issue.

[8] W. L. Everitt and G. E. Anner, *Communication Engineering*. New York: McGraw-Hill, 1956, pp. 294–315.

[9] W. C. Johnson, *Transmission Lines and Networks*. New York: McGraw-Hill, 1950.

[10] S. Ramo and J. R. Whinnery, *Fields and Waves in Modern Radio*. New York: Wiley, 1953.

[11] P. H. Smith, "Transmission line calculator," *Electronics*, pp. 29–31, January 1939; see also "An improved transmission line calculator," *Electronics*, p. 130, January 1944.

[12] R. A. Kempf, "Coaxial impedance standards," *Bell Sys. Tech. J.*, vol. 30, pp. 689–705, July 1951.

[13] M. C. Selby, E. C. Wolzien, and R. M. Jickling, "Coaxial radio frequency connectors and their electrical quality," *J. Res. NBS*, vol. 52, pp. 121–132, March 1954.

[14] B. O. Weinschel, "Standardization of precision coaxial connectors," this issue.

[15] I. A. Harris and R. E. Spinney, "The realization of high-frequency impedance standards using air-spaced coaxial lines," *IEEE Trans. Instrumentation and Measurement*, vol. IM-13, pp. 265–272, December 1964.

[16] T. E. MacKenzie and A. E. Sanderson, "Some fundamental design principles for the development of precision coaxial standards and components," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-14, pp. 29–39, January 1966.

[17] J. A. Stratton, *Electromagnetic Theory*. New York: McGraw-Hill, 1941, p. 550.

[18] A. Russell, "The effective resistance and inductance of a concentric main and methods of computing ber and bei and allied functions," *Philosophical Mag.*, vol. 17, pp. 524–552, April 1909.

[19] R. E. Nelson and M. R. Coryell, "Electrical parameters of precision, coaxial, air-dielectric transmission lines," NBS Mono. 96, p. 6, June 1966.

[20] L. Essen and K. D. Froome, "Dielectric constant and refractive index of air and its principal constituents at 24 GHz," *Nature*, vol. 167, p. 512, March 1951.

[21] "Vapor pressure of water below 100°C," in *Handbook of Chemistry and Physics*. Cleveland, Ohio: Chemical Rubber Co., any edition.

[22] B. O. Weinschel, "Air-filled coaxial lines as absolute impedance standards," *Microwave J.*, vol. 7, pp. 47–50, April 1964.

[23] T. E. MacKenzie, "Some techniques and their limitations as related to the measurement of small reflections in precision coaxial transmission lines," *IEEE Trans. Instrumentation and Measurement*, vol. IM-15, pp. 365–375, December 1966.

[24] D. Woods, "A coaxial connector system for precision RF measuring instruments and standards," *Proc. IEE (London)*, vol. 108, pt. B, pp. 205–215, March 1961.

[25] A. E. Sanderson, "A new high-precision method for the measurement of the VSWR of coaxial connectors," *IRE Trans. Microwave Theory and Techniques*, vol. MTT-9, pp. 524–528, November 1961.

[26] F. R. Huber and H. Neubauer, "Measurement techniques for the determination of the major characteristics of coaxial components," *Microwave J.*, vol. 5, pp. 196–203, September 1962.

[27] A. E. Sanderson, "Calibration techniques for one- and two-port devices using coaxial air lines as absolute impedance," presented at the Annual ISA Conf. and Exhibit, West Concord, Mass., General Radio Co., Reprint B21, 1964.

[28] R. C. Powell, R. M. Jickling, and A. E. Hess, "High-frequency impedance standards at the National Bureau of Standards," *IRE Trans. Instrumentation*, vol. I-7, pp. 270–274, December 1958.

[29] R. W. Beatty and W. J. Anson, "Application of reflectometer techniques to accurate reflection measurements in coaxial systems," *Proc. IEE (London)*, vol. 109, pt. B, pp. 345–348, July 1962.

[30] B. O. Weinschel, G. U. Sorger, S. J. Raff, and J. E. Ebert, "Precision coaxial VSWR measurements by coupled sliding-load technique," *IEEE Trans. Instrumentation and Measurement*, vol. IM-13, pp. 292–300, December 1964.

[31] J. E. Cruz and R. L. Brooke, "A variable characteristic impedance coaxial line," *IEEE Trans. Microwave Theory and Techniques (Correspondence)*, vol. MTT-13, pp. 477–478, July 1965.

[32] L. E. Huntley and R. N. Jones, "Lumped parameter impedance measurements," this issue.

[33] R. A. Soderman, "A new bridge for the measurement of impedance between 10 and 165 Mc," *Gen. Radio Exper.*, vol. 24, pp. 1–7, February 1950.

[34] D. Woods, "Admittance standardization and measurement in relation to coaxial systems," *IRE Trans. Instrumentation*, vol. I-9, pp. 258–268, September 1960.

[35] J. F. Byrne, "A null method for the determination of impedance in the 100–400 Mc range," *Proc. Nat'l Electronic Conf.*, vol. 3, pp. 603–614, 1947.

[36] A. Fong, "Direct measurement of impedance in the 50–500 Mc range," *Hewlett-Packard J.*, vol. 1, April 1950.

[37] R. A. Soderman, "Improved accuracy and convenience with the type 1602-B admittance meter in the VHF and UHF bands," *Gen. Radio Exper.*, vol. 28, August 1953.

[38] A. Alford and C. B. Watts, Jr., "A wide band coaxial hybrid," *IRE Conv. Rec.*, pt. 1, pp. 171–179, 1956.

[39] J. K. Huntoon and W. B. Wholey, "The perfect load and the null shift-aids in VSWR measurements," *Hewlett-Packard J.*, vol. 3, pp. 2–4, January–February 1952.

[40] A. Weissfloch, "Ein transformationsatz verlustloser Vierpole und seine Anwendung auf die experimentelle Untersuchung von Dezimeter—und—Zentimeterwellen—Shaltungen," *Hochfrequenz und Elektroak.*, vol. 60, pp. 67–73, September 1942.

[41] A. B. Giordano, "Measurement of standing-wave ratio," in *Handbook of Microwave Measurements*, M. Sucher and J. Fox, Eds. New York: Polytechnic Press, 1963, ch. 2, pp. 73–133.
[42] R. W. Beatty, "An automatic method for obtaining data in the Weissfloch-Feenberg node-shift technique," *Proc. IEEE (Correspondence)*, vol. 53, pp. 79–80, January 1965.
[43] W. Altar, F. B. Marshall and L. P. Hunter, "Probe error in standing-wave detectors," *Proc. IRE*, vol. 34, pp. 33–44, January 1946.
[44] R. W. Beatty "Microwave impedance measurements and standards," NBS Mono. 82, Sect. 7, August 1965.
[45] , "Magnified and squared VSWR responses for microwave reflection coefficient measurements," *IRE Trans. Microwave Theory and Techniques*, vol. MTT-7, pp. 346–350, July 1959.
[46] H. E. Sorrows, W. E. Ryan, and R. C. Ellenwood, "Evaluation of coaxial slotted-line impedance measurements," *Proc. IRE*, vol. 39, pp. 162–168, February 1951.
[47] A. E. Sanderson, "A slotted-line recorder system," *Gen. Radio Exper.*, vol. 39, pp. 3–10, January 1965.
[48] A. M. Winzemer, "Methods for obtaining the voltage standing-wave ratio on transmission lines independently of the detector characteristics," *Proc. IRE*, vol. 38, pp. 275–279, March 1950.
[49] B. M. Oliver, "Time domain reflectometry," *Hewlett-Packard J.*, vol. 15, February 1964.
[50] G. F. Engen and R. W. Beatty, "Microwave reflectometer techniques," *IRE Trans. Microwave Theory and Techniques*, vol. MTT-7, pp. 351–355, July 1959.
[51] W. E. Little and J. P. Wakefield, "Measurement of precision coaxial connectors using reflectometer techniques," *IEEE Internat'l Conv. Rec.*, pt. 11, pp. 89–97, March 1965.
[52] R. E. Spinney, "A precision reflectometer method for measurements on coaxial components at ultra-high frequencies," *Proc. IEE (London)*, vol. 110, pp. 521–522, March 1963.
[53] F. Jayne, "Improved reflectometer test for coaxial connectors," *Microwaves*, vol. 4, p. 34, September 1965.
[54] P. C. Ely, Jr., "Swept-frequency techniques," this issue.
[55] C. B. Watts, Jr., and A. Alford, "An automatic impedance plotter based on a hybrid-like network with a very wide frequency range," *IRE Nat'l Conv. Rec.*, pt. 5, pp. 146–150, 1957.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967
pp. 912-923
THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

Lumped Parameter Impedance Measurements

L. E. HUNTLEY AND R. N. JONES, MEMBER, IEEE

Abstract—**This paper is intended to be tutorial in the specific area of lumped parameter immittance measurement at radio frequencies. Included is a brief background discussion with particular emphasis upon the important recent developments of precision coaxial connectors and coaxial air dielectric transmission lines as immittance standards. Special emphasis is given to precision coaxial connectors and their necessity in achieving highest accuracies. Other sections of the paper deal with standards, techniques of measurement, and instruments. The present state-of-the-art is presented in graphical form wherein the accuracies attainable by best practices are compared with the best capabilities found in specifications for commercial instruments. The state-of-the-art presentation includes two-terminal as well as three-terminal measurements. The paper concludes with some recommendations for improving the state-of-the-art in this measurement area.**

I. Introduction

THIS PAPER deals with the measurement of impedance in "lumped" circuits at radio frequencies. The frequency range of interest extends from the upper limit of the audio range, approximately 30 kHz, to the upper limit of usefulness of lumped circuit techniques, which is perhaps 100 to 300 MHz. There is no clear-cut dividing line between the areas where lumped analysis and distributed analysis are appropriate, each being applied wherever it is useful. For example, it is sometimes convenient to think of a probe extending through the wall of a coaxial line as forming a capacitive voltage divider at frequencies above 1 GHz and, on the other hand, distributed parameter analysis is applied to sections of precision coaxial line to provide the NBS standards of impedance at frequencies as low as 30 kHz [1]. Because the mathematical manipulations required are relatively simple, and the lumped parameter approach is a powerful one, the approximate equations of ordinary circuit theory are almost always used when they will yield the required accuracy over a sufficiently broad frequency range. When the effective values of circuit elements change rapidly with frequency—typically at high frequencies—or when the highest accuracy is needed at any frequency, it is usually best to use an analysis based upon the exact relations obtained from Maxwell's field equations. Ramo, Whinnery, and VanDuzer [2] and Carson [3] should be consulted for a detailed discussion of the approximations involved in lumped circuit theory.

In this discussion, impedance is considered to be a proportionality constant relating voltage and current in an electrical circuit. The relationship implied here is a linear one, and holds only for sinusoidal voltages and currents. The impedance of the two-terminal (or one-port) device of Fig. 1(a) is thus the ratio of the voltage across the terminals

Manuscript received April 16, 1967.
The authors are with the National Bureau of Standards, Boulder, Colo.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967
pp. 900-911
THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

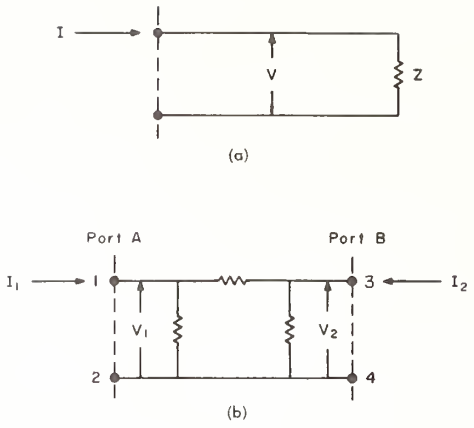


Fig. 1. (a) One-port network. (b) Two-port network.

to the current flowing through them, $Z = V/I$. The voltage-current relationships for the linear two-port of Fig. 1(b) may be expressed by the equations

$$V_1 = Z_a I_1 + Z_{ab} I_2$$
$$V_2 = Z_{ba} I_1 + Z_b I_2$$

from which are obtained the four impedances

$$\left.\begin{aligned} Z_b &= \frac{V_2}{I_2} \\ Z_{ab} &= \frac{V_1}{I_2} \end{aligned}\right\} I_1 = 0 \qquad \left.\begin{aligned} Z_a &= \frac{V_1}{I_1} \\ Z_{ba} &= \frac{V_2}{I_1} \end{aligned}\right\} I_2 = 0.$$

For reciprocal networks [4] $Z_{ab} = Z_{ba}$, and only three impedances are required to characterize the network.

It should be realized that defining impedance in this way does not necessarily imply that it is best measured in terms of voltage-current ratios. In fact, specific values of impedance may be calculated from a geometrical configuration of materials and the electromagnetic properties of the surrounding space, and impedance obtained in this way is at present much more accurately related to the basic quantities[1] than is either voltage or current.

In this article we will attempt a survey of the field of impedance measurement as applied to linear passive lumped circuits. The intent is to be general enough in scope to provide a good overview of the field, but at the same time provide enough detail to be useful in specific instances.

II. Background

The techniques of lumped parameter immittance[2] measurement at radio frequencies were primarily developed during the era between World Wars I and II. Since the 1940's, attention has been focused upon higher frequency work and the development of microwave techniques, with the result that the lower frequencies have received relatively less attention. Instruments such as RF bridges and Q-meters have undergone only minor modifications over the past two decades. Much of the instrumentation used at radio frequencies was developed through modifications to dc and audio-frequency instrumentation. An obvious example of this is the Wheatstone bridge circuit commonly used in modern RF immittance bridges. There are some notable exceptions which include the twin-T and bridged-T instruments and those employing the principles of resonance which were developed primarily for use at radio frequencies. However, all of the circuits mentioned have in common the fact that lumped parameter analysis is sufficiently exact for practical purposes. Advances in the state-of-the-art have been more toward the extension of existing techniques to higher frequencies than in the improvement of accuracies at frequencies where capabilities already existed.

A significant departure from this trend occurred in the 1950's when Woods [5], [6] of the United Kingdom introduced the concept of precision coaxial connectors and advocated using coaxial air dielectric transmission lines as immittance standards. These contributions resulted in accuracy improvements of one or even two orders of magnitude in many instances. For example, at 50 MHz a capacitance of 20 picofarads (pF) equipped with a precision coaxial connector can be measured with an uncertainty of only 0.01 percent as compared to an uncertainty of the order of one percent where a nonprecision connector is used.

III. Precision Coaxial Connectors

Connectors and their effect upon measurement accuracy is a subject of the greatest importance in immittance measurements [7]. Every immittance measurement is made in terms of some sort of electrical measuring circuit, such as that of Fig. 2. Because the unknown is an integral part of the measuring circuit, it is necessary to distinguish, and mechanically separate, the part of the circuit being measured from the rest of the circuit if the measurement is to be meaningful and useful. This is conveniently done by establishing a mechanical plane of separation which is coincident with the "reference plane" separating the two parts of the circuit. If the connection is made in a plane perpendicular to the axis of a uniform transmission line, the two parts of the circuit may be separated very precisely. If the measured immittance is to be insensitive to changes in its environment, the connection must be mechanically and thermally stable and well-shielded electrically. If the measurement is to be repeatable, the connection must be such that it may be precisely repeated.

These characteristics of a connector—well-defined reference plane, mechanical and thermal stability, good shielding, and repeatable connection—are all that are required in those practical situations where only precision is required. Any good, repeatable, shielded connector system can be used for *precise* measurements but, in addition, accurately known reference conditions are necessary for *accurate* absolute measurements. The immittance of the short circuit or open circuit commonly used as references can be calculated for coaxial structures whose inner and outer conductors terminate in a common plane perpendicular to

[1] In this paper, "basic quantities" refers to the quantities mass, length, time, and μ_0.

[2] To avoid writing "impedance and admittance" the two quantities are assigned one name, immittance.

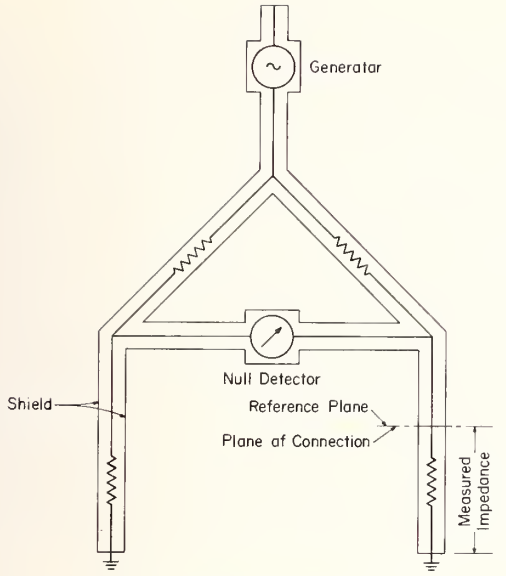


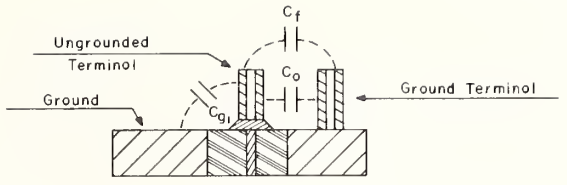
Fig. 2. Electrical measuring circuit.

their axis [8], [9]. Furthermore, both immittances may be verified by measurement if the connectors are sexless, that is, if any two connectors of the same line size are capable of being mated to each other without adaptors.

The IEEE Subcommittee on Precision Coaxial Connectors has adopted specifications [10] for precision connectors in 14-mm and 7-mm line sizes for use to 8.5 GHz and 17 GHz, respectively. Connectors in both line sizes are commercially available. While these connectors were developed primarily for use at the higher frequencies, they have the characteristics which are desirable in a connector for use in lumped circuit measurements.

The effect of connectors on immittance measurements will be demonstrated by considering a typical measurement situation. In measuring an unknown capacitor, a capacitance bridge is nulled initially with its terminals open circuited. The unknown capacitor is then connected, and the bridge standards are adjusted to restore the null. Figure 3(a) represents the "unknown" terminals of a capacitance bridge equipped with banana plugs or binding posts, and Fig. 4(a) represents the terminal of a similar bridge equipped with a precision coaxial connector. Figures 3(b) and 4(b) represent the same terminals with the unknown capacitor C_x connected for measurement.

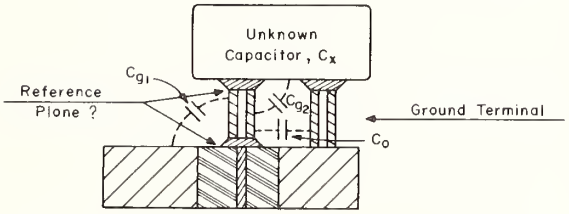
In the figures, C_0 represents the fixed component of the terminal capacitance plus the capacitance of the internal connecting leads, and is a characteristic of the bridge. In Fig. 3(a), C_{g1} is the capacitance from the ungrounded terminal to the instrument case and the surroundings. Its value varies with changes in the location of conductors in the vicinity of the connector. C_f is the fringe capacitance, defined for precision coaxial connectors to be the capacitance which is removed when two connectors are joined, forming a uniform coaxial line through the plane of connection. For banana plugs or binding posts, the fringe capacitance is not well defined. It has the same effect as a change in C_{g1}, which in the usual configuration is unknown



At Null The Bridge Reads

$$C_A = C_o + C_{g_1} + C_f$$

(a)

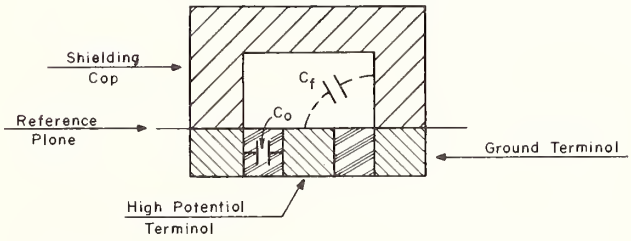


At Null The Bridge Reads

$$C_B = C_o + C_{g_1} + C_{g_2} + C_x$$
$$C_B - C_A = \Delta C = C_{g_2} + C_x - C_f$$
$$C_x = \Delta C - C_{g_2} + C_f$$

(b)

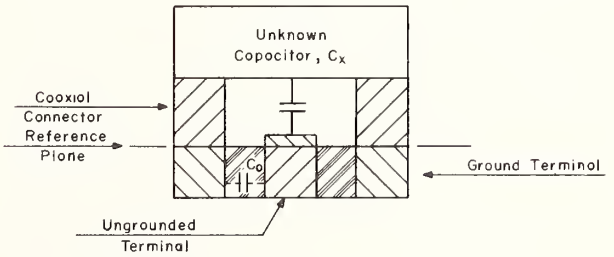
Fig. 3. (a) Open-circuit binding post connectors. (b) Binding post connector with capacitor attached.



At Null The Bridge Reads

$$C_A = C_o + C_f$$

(a)



At Null The Bridge Reads

$$C_B = C_o + C_x$$
$$C_B - C_A = \Delta C = C_x - C_f$$
$$C_x = \Delta C + C_f$$

(b)

Fig. 4. (a) Open-circuit precision coaxial connector. (b) Precision coaxial connector with capacitor attached.

and unknowable. C_{g2} is the capacitance from the ungrounded terminal to the case of the unknown capacitor.

If C_{g2} is lumped with and considered a part of C_x, and if C_{g1} is truly unchanged when C_x is connected, the equations for the measurement with binding posts are identical to those for the measurement with precision connectors. The difference lies in the precision of connection and the accuracy with which C_f can be determined. C_{g2} is affected by the length and diameter of the binding posts and by how well the capacitor is seated for measurement. Unless these are standardized, the measured value of C_x will vary from one setup to another. Attempts to measure the fringe capacitance of one particular type of binding post at 1-inch spacing have yielded a value of 0.3 pF, with an uncertainty of perhaps 50 percent. By contrast, the fringe capacitance of the shielding cap of Fig. 4(a) has been determined to be 0.2074 pF with an uncertainty of 0.3 percent. A precision connector developed and used by NBS permits capacitance measurements to be repeated with a 3-sigma[3] value of less than 0.001 pF. The practical consequence is that a 20-pF capacitor equipped with banana plugs cannot be measured with an error less than about 0.75 percent, while the same capacitor could be measured to within 0.005 percent if it were equipped with a precision connector.

A similar situation exists for inductance measurements, where the measurement is referenced to a "short" which is removed when the unknown inductor is connected. For a typical banana plug arrangement, the short might have an inductance of 0.01 μH which may be calculated with an error of about 50 percent. This means that a 1-μH inductor may be measured with an error of about 0.5 percent. The shorting disk used with the 14-mm precision connector has an inductance of less than 0.000016 μH at 1 MHz [11], which if neglected entirely would cause an error of only 0.002 percent in measuring 1 μH.

We have not discussed the advantages of precision connectors in the measurement of resistance and conductance, or in fixing residual impedances[4] in inductance and capacitance standards and in three-terminal measurements. We do not wish to belabor the point, but do want to make a convincing case for using precision connectors wherever accurate and precise immittance measurements are required. There are many immittance measurements which are not compatible with coaxial connectors, and it is necessary to compromise precision and accuracy in order to make the necessary connections. The change from a precision coaxial system can be accomplished with suitable adaptors, and this is preferably done only when it becomes necessary. The spectacular increase in accuracy of lumped immittance measurements in the past few years has resulted almost entirely from improved standards, instruments, and techniques made possible by precision connectors. This accuracy cannot be realized by the ultimate user of the measurement unless he is willing to accept the small added cost of using precision connectors.

[3] Sigma, the standard deviation, is a measure of the precision of the measurement.

[4] Residual impedance is defined in Section IV.

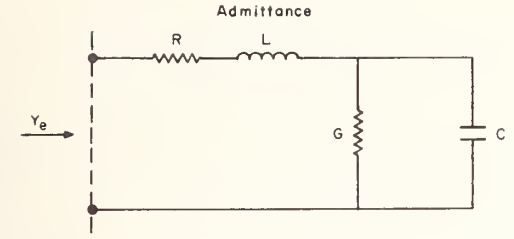
IV. Standards

When attempting to verify the accuracy of an immittance measurement in the 30 kHz to 300 MHz frequency range, an interesting dilemma arises concerning what to use for a standard. Various manufacturers produce components which are useful in determining the agreement between two instruments of the same model or manufacture, but beyond this, standards which are commonly accepted for use with all commercial instruments are nonexistent. This situation has caused a large amount of unnecessary expense and inconvenience which will continue until appropriate standards and standardization procedures are adopted on an industry-wide basis. To accomplish effective industry-wide standardization, two important initial steps must be taken. First, agreement is needed on specific values and frequencies for which standards are needed and second, precision coaxial connectors must be utilized on both instruments and standards. The following discussion pertains to standards for use as interlaboratory transfer standards.

An immittance standard for general use should have certain characteristics if it is to fulfill its purpose in maintaining uniform measurement agreement. These characteristics are long-term stability, high purity, value independent of frequency and environmental change, and convenience in use. High purity and frequency independence are related characteristics which depend upon the degree to which residual impedances are present in a capacitor, a resistor, or an inductor. A standard or a component is usually called a capacitor, resistor, or an inductor in accordance with whichever of these three parameters is predominant at low frequencies. The other parameters, which are always present in a practical situation, are called residual impedances. A capacitor of high purity has a high ratio of susceptance to conductance (high Q) and exhibits minimum change in capacitance value as the frequency is varied. Similarly, a high-purity inductor has a high ratio of reactance to resistance (high Q) and changes little in inductance value with frequency change. A high-purity resistor (or conductor) exhibits a near-zero phase angle (low Q) and only small changes in resistance or conductance with frequency.

At sufficiently low frequencies, a two-terminal capacitance or conductance may usually be represented by the equivalent circuit of Fig. 5(a) and a two-terminal inductance or resistance by the equivalent circuit of Fig. 5(b). The equations in the figures show that the effective values of a component depend upon the frequency and the residual parameters. Furthermore, precision connectors are seen to be important because they permit accurate definition and repeatability of the residuals by shielding and strict control of the circuit geometry in the vicinity of the connection. Field and Sinclair [12] have described a method for determining the residual series inductance and resistance of variable air capacitors; other methods, based upon resonance, are commonly used to evaluate the residual series inductance [13], [14].

As the frequency is increased to the vicinity of self-resonance in a standard or component, simple equivalent

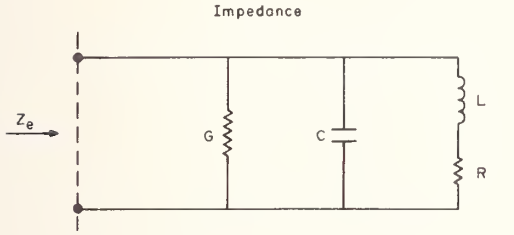


$$Y_e = G_e + j\omega C_e$$

$$G_e = \frac{G(1+RG) + R\omega^2C^2}{(1+RG-\omega^2LC)^2 + (R\omega C + G\omega L)^2}$$

$$C_e = \frac{C(1-\omega^2LC) - G^2L}{(1+RG-\omega^2LC)^2 + (R\omega C + G\omega L)^2}$$

(a)



$$Z_e = R_e + j\omega L_e$$

$$R_e = \frac{R(1+RG) + G\omega^2L^2}{(1+RG-\omega^2LC)^2 + (G\omega L + R\omega C)^2}$$

$$L_e = \frac{L(1-\omega^2LC) - R^2C}{(1+RG-\omega^2LC)^2 + (G\omega L + R\omega C)^2}$$

(b)

Fig. 5. (a) Admittance equivalent circuit. (b) Impedance equivalent circuit.

circuits such as those in Fig. 5(a) and (b) no longer adequately represent the actual distribution of the various parameters, and more complex equivalent circuits must be employed. At this point lumped parameter analysis is usually abandoned and we must rely upon calculable standards such as coaxial air dielectric transmission lines.

At the Radio Standards Laboratory of NBS, capacitance is the parameter selected as the foundation for the system of lumped parameter immittance standards [1]. This is because capacitors have been constructed which have the aforementioned desired characteristics to a higher degree than either inductors or resistors. Because an air dielectric capacitor stores energy primarily in free space, its value is essentially unaffected by the properties of the materials used in its construction, and the capacitance is determined by its geometry and the relative permittivity of air. For this reason the stability of the capacitor is largely determined by the mechanical stability of the device. Capacitors can be made which have simple geometric configurations, resulting in short current paths with correspondingly small inductive and resistive residual impedances. The principal advantage of capacitors as standards is the accuracy to which their values may be obtained from the basic quantities. Two-terminal capacitance at high frequencies is calculated [15]–[17] from the dimensions of a section of coaxial line and three-terminal capacitance is obtained from the Thompson-Lampard [18] capacitor in terms of length and the permittivity of free space. The coaxial lines have been compared to the Thompson-Lampard capacitors at audio frequencies to confirm the accuracy of calculation. Such comparisons have resulted in agreements to within 1×10^{-15} farads for values between 5 and 50 pF.

Another advantage of capacitors is that variable elements can be made which are not only highly pure but also provide essentially infinite resolution. There is often a choice of measurement circuits for a particular application wherein one circuit containing a variable air capacitor may provide the same measurement capability as another circuit containing a variable resistor or a variable inductor. Where such an alternative exists, the choice is usually made in favor of the circuit containing the variable air capacitor.

Inductors are not as desirable as capacitors for primary standards of RF impedance because a greater number of factors affect their values, making calculation in terms of basic quantities more difficult. For example, at lower frequencies current is not confined to the outer surfaces of the conductors and therefore material properties become important. The residual impedances in inductors are usually greater than in capacitors, resulting in a more pronounced frequency dependence. This is especially true of wire-wound inductors. The circuit of Fig. 5(b) indicates the residual parameters as series resistance and parallel capacitance and conductance, but there are still other factors which affect the inductance. These include the mutual inductance effect of a shield and the proximity effect of adjacent windings [19]. In inductors containing high permeability core materials, the inductance is dependent upon the current magnitude so that it is necessary to specify the current in the inductor when measured values are obtained. Good air-core inductors have all the other characteristics of desirable standards and make very good transfer standards.

High-frequency resistance is the most difficult quantity to obtain from the basic quantities because of the dependence upon material properties. Some resistance materials exhibit a long-term aging effect so that stability cannot be relied upon. This problem is largely avoided by using wire-wound resistors of stable alloys, but this creates problems at high frequencies because of large residual reactances. In recent years techniques in the manufacture of deposited metal film resistors have improved so that good transfer standards of resistance are feasible.

The NBS standards for RF immittance measurement are derived in approximately the manner represented by the chart in Fig. 6. This chain of traceability is only for a typical case and is somewhat different for other specific values and frequencies. The standards most directly related to length, time, and μ_0 are the smaller values of inductance (≤ 0.1 μH) and capacitance (≤ 50 pF). Excellent standards for small values of inductance and capacitance are provided by sections of coaxial air dielectric transmission line. For induc-

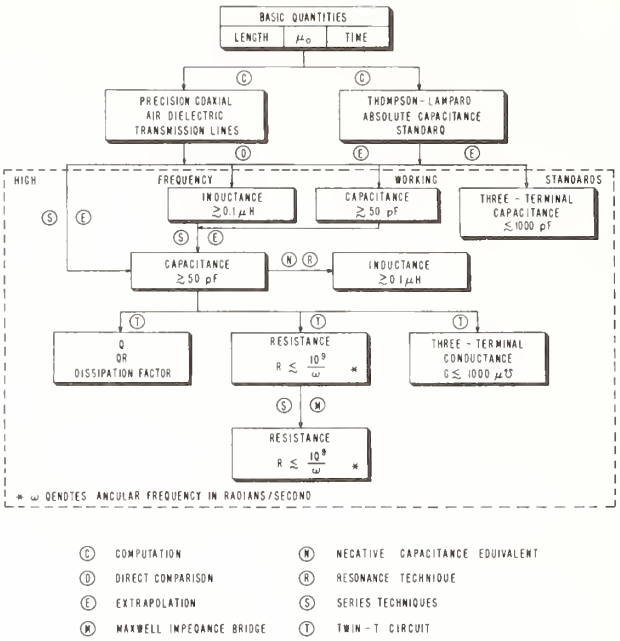


Fig. 6. Derivation of NBS immittance standards.

tance standards the lines are short circuited, and for capacitance standards the same lines are used in the open-circuit condition. At the higher frequencies (≥ 300 MHz), where a quarter wavelength is of a practical size, such coaxial lines are used for a wide range of values of inductance and capacitance.

The so-called open and short circuits play an important part as reference immittances. Since their impedance values are only nominally an open or a short, accurate knowledge of these values are required. Such immittance values can be calculated for coaxial structures whose inner and outer conductors terminate in a common plane perpendicular to their axis [8], [9].

V. Techniques

The techniques for measuring immittance have been classified in a number of ways, but we will confine the discussion to those which have been found most useful. One classification distinguishes between direct measurements—by the voltmeter/ammeter method, for example—and measurements in which the unknown immittance is compared to a standard of known value. While instruments which determine immittance from the ratio of voltage to current are available [20], they are not at present capable of the accuracy attainable by comparison to a standard and are not used at radio frequencies when high accuracy is required.

Another worthwhile distinction may be made between techniques which use the measuring instrument as the standard, and comparator techniques in which the instrument simply serves to indicate small differences between a standard and an unknown which are nearly identical immittances. The first approach is used in most commercial immittance measuring instruments. These instruments contain internal standards to which the unknown immittance is compared. The range covered by the standards typically extends from short circuit to a finite impedance, or from open circuit to a finite admittance, and the instrument usually operates over a very wide frequency range. The continuous coverage of immittance value and frequency makes these instruments extremely versatile, but also makes them extremely difficult to calibrate.

The principal advantage of the comparator approach is that even relatively inaccurate instruments can be used to make highly accurate measurements if a sufficiently accurate standard is available. For example, a one percent measurement of a difference of one percent corresponds to 0.01 percent error in comparing an unknown to a standard. The practical result is that very accurate measurements can be obtained without investing in instruments which are themselves very accurate and therefore expensive. Many commercial instruments are capable of one to three percent measurements over a broad frequency range, and would be suitable for this type of immittance measurement if they were equipped with a suitable connector.

The principal disadvantage of the comparator approach is that the necessary standards must be maintained, one near each immittance value to be measured, and the value of each standard must be known at each frequency at which it is used. This may not be as serious as it might seem, since a relatively small number of standards can give good coverage of a wide range of values. If the common 1-2-5 sequence of values is used, there are only three standards per decade. Furthermore, if these standards are used with instruments capable of 1 percent measurements, and which have sufficient resolution, the measurement uncertainty will vary from perhaps 0.001 percent for "identical" standards to at worst 0.4 percent when measuring values midway between standard values. High-quality standards vary predictably with frequency, so that it is not strictly necessary to calibrate them at every frequency at which a measurement may be made. Often sufficient accuracy may be obtained by measuring at a few standard frequencies and interpolating for intermediate frequencies.

Another useful classification of techniques distinguishes between null methods and resonance methods. The null methods use some sort of nulling circuit, such as a Wheatstone bridge or the twin-T, to indicate that a definite relationship exists between the circuit elements. Resonance techniques use inductors and capacitors to form a series or parallel resonant circuit, and a voltmeter or a current meter to determine when resonance exists [13], [21]. Instruments based on either technique may be used for direct measurement or for comparing nearly identical immittances. In general, the null condition can be more precisely determined than can the resonance condition, making null measurements most useful for precise immittance measurements.

VI. Instruments

In general, there is a class of impedance measuring instruments corresponding to each measurement technique. Which instrument is best suited for a particular measure-

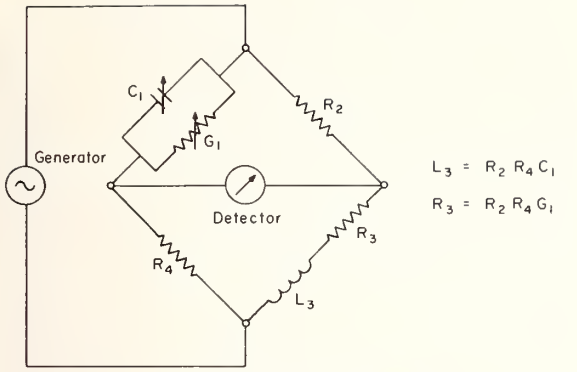


Fig. 7. Basic Maxwell bridge.

ment depends on the parameters to be measured, their magnitudes, the frequency at which they will be measured, and the accuracy required. In this section, some of the more useful instruments are discussed together with the advantages and disadvantages of each. Voltmeter/ammeter [20] or ohmmeter techniques for measuring impedance are generally not suited to accurate measurements, and will not be discussed. The null instruments which use active elements, such as amplifiers, as standards [22] are likewise not suited to accurate measurements, so the discussion will be limited to instruments using null or resonance techniques in which all the circuit elements are linear, passive devices. (The signal source, or generator, and the detector are not considered part of the instrument.)

Most null instruments use a variation of the familiar four-arm Wheatstone circuit or the twin-T null circuits described by Tuttle [23], but there are exceptions. All null instruments have in common the characteristics that one or more of the impedance elements in the circuit can be adjusted to provide zero transmission between generator and detector (null condition), and detector and generator may be interchanged without affecting the null condition.

Null instruments which have been found useful for two-terminal measurements at radio frequencies include the four-arm bridges of Figs. 7, 8, and 9, the twin-T of Fig. 10, and the transformer ratio bridge of Fig. 11. Most RF immittance measuring instruments are modifications of these basic instruments. The Maxwell bridge of Fig. 7, and the Schering bridge of Fig. 8(b) are impedance bridges, that is, they measure *impedances* between zero and a finite value and can be nulled with the terminals short circuited. The Schering bridge of Fig. 8(c), the resistance ratio bridge of Fig. 9, and the twin-T of Fig. 10 are admittance bridges which measure *admittances* between zero and a finite value, and can be nulled with the terminals open circuited. The transformer ratio bridge of Fig. 11 can be used as either an impedance bridge or an admittance bridge.

The Maxwell bridge has been used at frequencies to 5 MHz [24] but is limited in that it measures resistance in terms of an incremental conductance. The Schering bridge, which measures both resistance and reactance in terms of capacitance increments, and which measures either impedance or admittance, is the most widely used RF bridge. Most

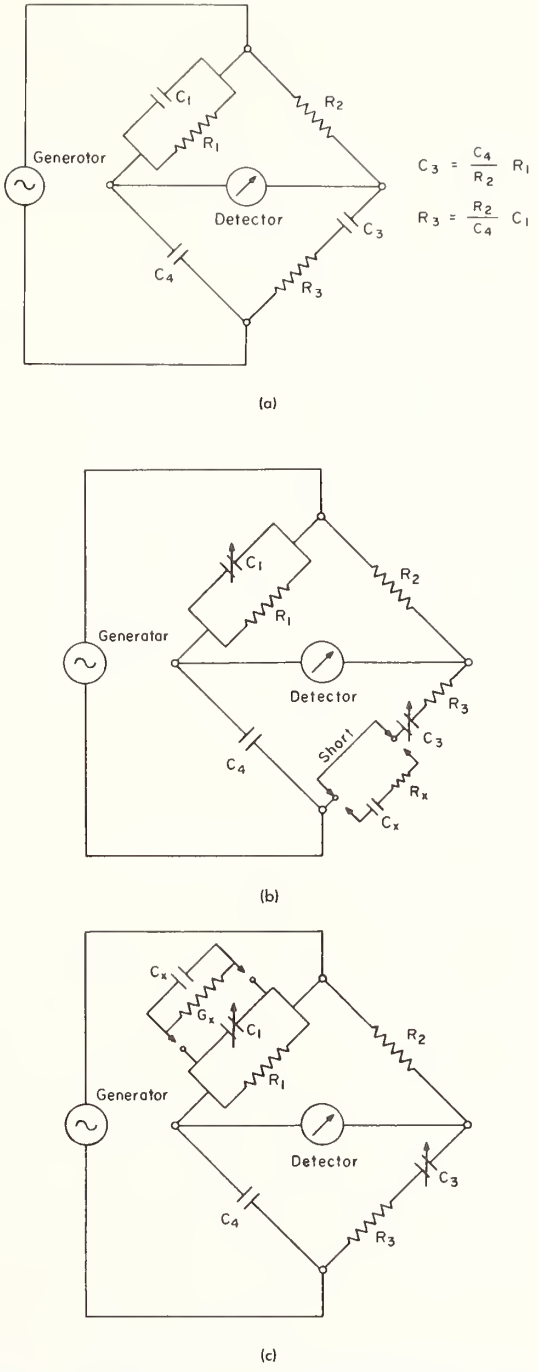
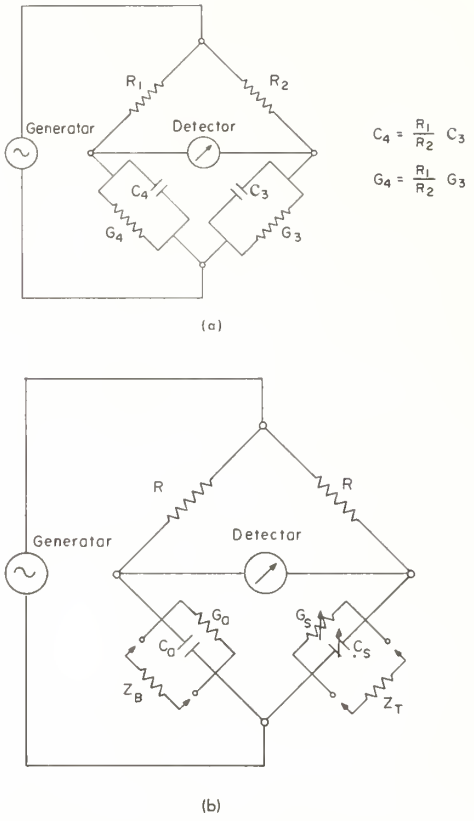


Fig. 8. (a) Basic Schering bridge. (b) Schering impedance bridge. (c) Schering admittance bridge.

commercial RF bridges use this circuit [25], and it has been successfully used to 250 MHz [26]. The ratio bridge has been used to measure admittance to 20 MHz [24]. The major weakness of this bridge is that it uses an incremental conductance standard to measure conductance. The Radio Standards Laboratory of NBS has constructed a modular [27] ratio bridge for use as an immittance comparator [Fig. 9(b)]. In the figure, $R_1 = R_2$, G_0 and C_0 balance the



$C_4 = \frac{R_1}{R_2} C_3$

$G_4 = \frac{R_1}{R_2} G_3$

Fig. 9. (a) Basic resistance ratio bridge. (b) Modular resistance ratio bridge (NBS Comparator).

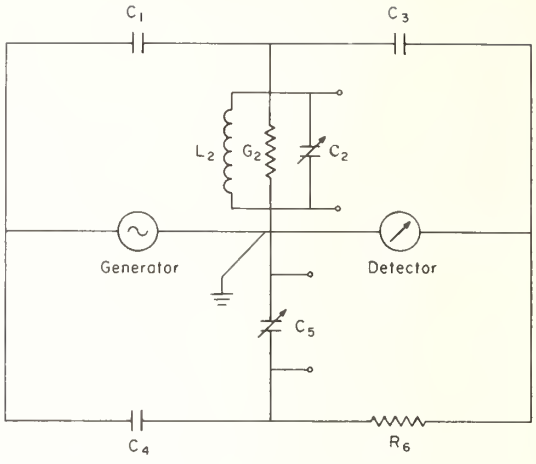


Fig. 10. Twin-T circuit.

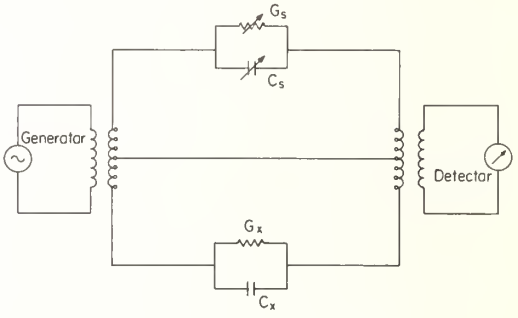


Fig. 11. Transformer ratio bridge.

initial nonzero G_s and C_s, and Z_b is nearly equal to the test impedance Z_t. The instrument measures small differences between test impedances in terms of small changes in G_s and C_s, so that G_s and C_s need not be high-quality standards. The Maxwell and Schering bridges cover an immittance range which depends upon the frequency and provide accuracies of typically 1 to 3 percent.

The commercial version of the twin-T, developed by Sinclair [28], was used to measure admittance to 40 MHz, and a dual admittance bridge developed by Woods [29] measures admittance with a basic uncertainty of 0.1 percent to 100 MHz which is degraded to 0.3 percent at 300 MHz. The Radio Standards Laboratory has constructed a twin-T useful to 15 MHz, primarily for measuring conductance in terms of capacitance increments [30]. This instrument measures conductance with errors of about 0.05 percent at 1 MHz. The advantages of the twin-T instruments include the facts that incremental capacitors are used to measure both conductance and capacitance, and the generator, the detector, the measured immittance, and the incremental standards are all connected to a common ground. The advantages are offset by the fact that the bridge balance is frequency sensitive, making twin-T instruments relatively inconvenient to use.

The transformer ratio bridge (Fig. 11) is one of the exceptions which cannot be classified as a Wheatstone, or a twin-T instrument. In this bridge, null is obtained by adjusting the input or output turns ratio, or the immittance standards, to provide zero current in the detector transformer [31]. A line of commercial instruments uses this circuit for admittance measurements at frequencies to 250 MHz.

A basic difference between admittance bridges and impedance bridges, which is not obvious from the simple circuits, is the relative complexity of the shielding required for the two types of instruments. The subject of shielding is far too complex [32], [33] to be entered into here beyond pointing out that impedance bridges usually require more complex shielding than do admittance bridges. The practical result is that impedance bridges cannot be readily improved by a simple application of precision connectors. This, plus the fact that good absolute standards of low-valued impedance are not easily obtainable, have resulted in the accuracy of measurement of small impedance lagging behind that of small admittance measurements.

While resonance methods are not at present capable of the same accuracy as null methods, they are widely enough used to justify some discussion. The simple circuits in Fig. 12 illustrate the essentials of a resonance instrument: a signal source, a capacitor and an inductor connected in series or parallel, and an ammeter or a voltmeter for detecting the resonant condition. Resonance measurements often combine the techniques of direct measurement and comparison to a standard. The resistive component is obtained directly by measuring voltage ratios, and the reactive com-

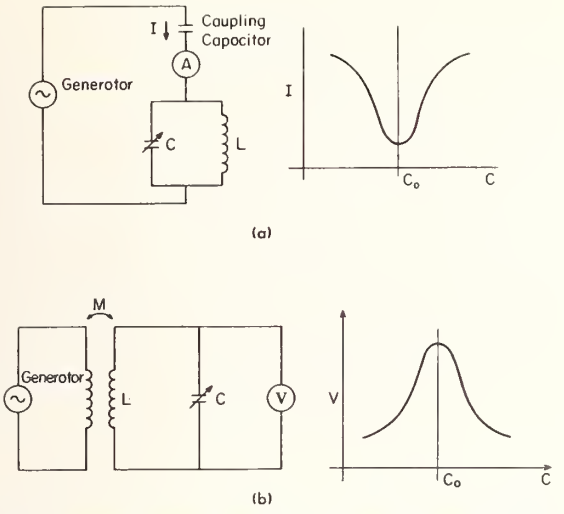


Fig. 12. (a) Series resonance method. (b) Parallel resonance method.

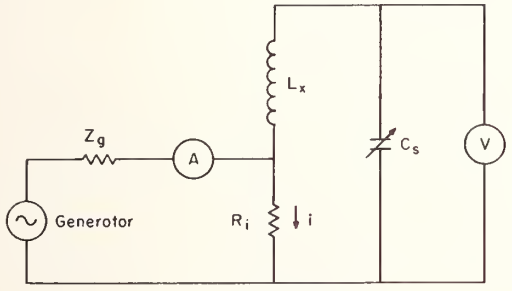


Fig. 13. Q-meter circuit.

ponent is obtained by comparison to the standard capacitor. Because resonance methods work best when the coupling to the generator is kept small, they are best suited to measurements in high-Q circuits. Resonance measurements are complementary to null measurements in that it is difficult to measure the resistive component of high-Q circuits with null instruments.

The Q meter [34], [35], represented schematically in Fig. 13, is a widely used resonance instrument. A small voltage, $e=iR_i$, appears across the series combination of L_x and C_s. The standard capacitor is tuned to the resonant value C_0, which maximizes the voltage across the capacitor as indicated by the voltmeter V. Since the voltage across either component of a series resonant circuit is Q times the voltage across the series combination, Q may be obtained from the ratio V/e. The voltage e is set to a predetermined value by adjusting the current through the known insertion resistor R_i, the current being monitored by the ammeter A. The unknown resistance R_x is obtained from the relationship

$$Q = \frac{\omega L_x}{R_x}.$$

Q meters are used at frequencies as high as 610 MHz [36].

A line of resonance instruments, which are not Q meters, measure both conductance and capacitance in terms of internal standards. A smoothly varying incremental conductance is obtained by varying the bias on a diode connected in parallel with the resonant circuit. These instruments measure admittance at frequencies between 0.1 and 100 MHz [37]. Accuracies of 1 percent for capacitance and 5 to 20 percent for resistance are claimed.

VII. Signal Generators and Detectors

In addition to the circuitry of RF impedance measuring instruments, it is necessary to give careful consideration to the signal generators and detectors used with them.

Good frequency stability, adequate power output, and low harmonic content are the main requirements for signal generators. The frequency stability required of a generator depends upon the measuring instrument with which it is used, and the item being measured and the accuracy desired. In circumstances where bridge balance equations involve frequency, or where resonant circuits are involved, stabilities of the order of 1 part in 10^6 are sufficient for accuracies of 1 part in 10^3. This is the case for the twin-T circuit and the Q meters. When a component is being measured at a frequency near self-resonance, frequency stability is especially necessary. In this situation the frequency must be accurately known, as well as stable. Bridge circuits which do not involve frequency in the balance equations do not impose critical frequency requirements, and stabilities of the order of 1 part in 10^4 are usually adequate. The required frequency stability is a criterion which must be established for specific measurement situations and can vary widely depending upon the measurement accuracies desired. This is also true for power output and harmonic content. In general, power output should be no more than that amount required to realize sufficiently sharp bridge balance conditions, and both bridge and detector can be determining factors. Increasing power output to gain sensitivity can lead to instabilities due to temperature fluctuations or to the more serious problem of damaging components. Higher generator output does not necessarily improve measurement resolution where a significant amount of leakage exists between generator and detector.

Harmonic content and frequency modulation are of serious concern in some instances and can be neglected in others. Measuring circuits whose balance equations involve frequency are subject to errors from both sources. To avoid problems arising from harmonics it is advisable to filter the generator output and use narrowband detectors.

Detectors should have the characteristics of high sensitivity and high selectivity, be well shielded to guard against leakage from the generator, and have good signal-to-noise ratio. Detector sensitivities which allow detection of signals in the range of a few microvolts are usually sufficient so that the full resolution of a measuring instrument can be utilized.

Ground loops and leakage are problems which are not solely related to either the generator or detector but are associated with the measurement setup as a whole. Ground loops are most effectively avoided by having a single low-

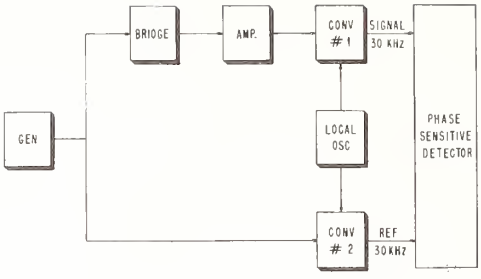


Fig. 14. Measurement setup using phase sensitive detector.

impedance ground path connecting the generator, bridge, and detector, as opposed to having a separate ground for each. Leakage can be reduced by placing added shielding around the generator and detector, and by employing semi-rigid coaxial cable and threaded connectors wherever possible.

Phase sensitive detectors are finding increasing application in measurements involving immittance bridges and offer some distinct advantages. In addition to providing greater sensitivity, they allow the operator to distinguish between unbalanced voltages due to real and imaginary immittances. An interesting result of this is that one gains an appreciation for the purity of the bridge standards because as the reactance standard is varied the variation in its resistance is displayed or vice versa. Especially when making measurements of low-Q components, balance may be difficult to locate because the reactance of the resistance standard in the bridge may vary appreciably with setting. Such "sliding null" conditions present little difficulty with phase sensitive detectors. The greater sensitivity of phase sensitive detectors comes about as a result of a much narrower effective bandwidth which results in improved signal-to-noise ratio. These detectors are not currently available for wide frequency ranges, but can be readily adapted for application at higher frequencies by using a local oscillator-mixer arrangement such as that shown in Fig. 14. The phase sensitive detector requires a reference signal which is in phase with the output from the bridge. This requirement is fulfilled by feeding the same generator signal into each converter. Under these circumstances the local oscillator is free running and need not be highly stable.

VIII. State of the Art (Two-Terminal)

The application of precision connectors to the measurement of lumped immittance has resulted in very substantial increases in measurement accuracy and precision. Better precision is a direct result of a more repeatable connection, while improved accuracy comes from two principal sources. One improvement in accuracy results from the fact that the open- and short-circuit reference conditions are accurately known. Another improvement comes from the fact that residual impedances, which cause the value of a standard to vary with frequency, are more precisely fixed by precision connectors.

Figures 15, 16, and 17 present, in graphical form, the

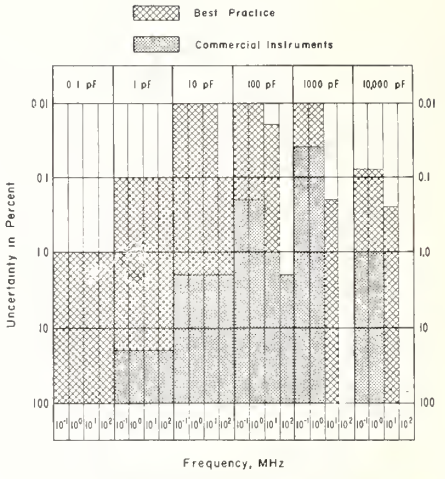


Fig. 15. State-of-the-art (two-terminal capacitance).

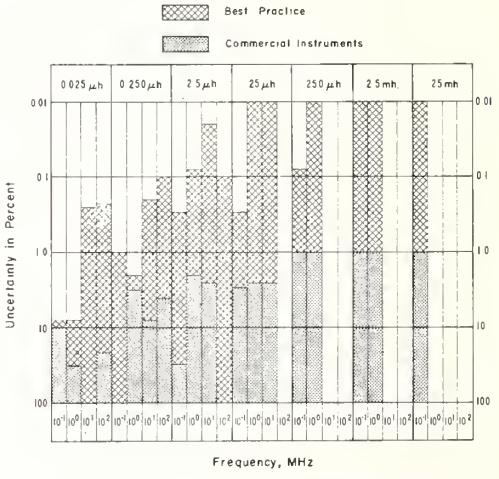


Fig. 16. State-of-the-art (two-terminal inductance).

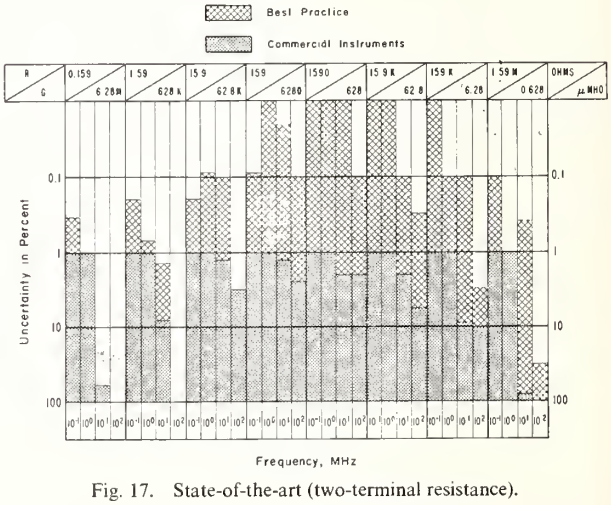


Fig. 17. State-of-the-art (two-terminal resistance).

authors' estimations of the existing state-of-the-art in two-terminal measurements of capacitance, inductance, and resistance at radio frequencies. Two classes of measurements are represented, the best possible with existing equipment and those possible with readily obtainable commercial instruments.[5] Usually, the best techniques use precision connectors. However, techniques for exploiting precision connectors in measuring low impedance have not been perfected, so that instruments equipped with banana jacks or binding posts represent the state-of-the-art in this area. One commercial instrument, the commercial version of Woods' dual admittance bridge, is equipped with a sexless, reference plane connector and represents the state-of-the-art for certain values of conductance and capacitance above 10 MHz.

IX. Three-Terminal Measurements

Present applications for three-terminal measurement are primarily confined to the measurement of small admittances, and the measurement instruments may employ either a ratio transformer or a differential capacitor [31], [38]. The technology for three-terminal measurement in the RF range is not as well developed as it is for two-terminal measurements, and at present there appears to be very little application above 5 MHz. The accuracy to which three-terminal measurements can be made is affected by such factors as the purity of the standards within the intrunent, transformer ratio uncertainties, the effect of impedances from each electrode to ground, and the inductance of the leads which connect the unknown to the instrument. Each of these factors becomes increasingly important at higher frequencies.

Figures 18 and 19 are estimates of the state-of-the-art for three-terminal capacitance and conductance measurement based upon accuracy specifications published by the manufacturers of commercial three-terminal instruments. In these figures, the accuracies claimed for commercial instruments also represent best practice primarily because there has been little effort by other than commercial laboratories to improve the state-of-the-art. The verification of accuracies is difficult because of the absence of generally accepted standards. The problems in standardization, although similar in many respects to those associated with two-terminal measurement, are much more complicated.

Three-terminal measurement at radio frequencies may be considered as a special case of the more general situation where the unknown is considered as a two-port such as that represented in Fig. 1(b). Such a two-port is completely described if the impedance values of its equivalent pi or equivalent T network are known. If the two-port is represented as a pi network, as shown in Fig. 20, the values Y_A, Y_B, and Y_C may be obtained by two-terminal measurements. These two-terminal measurements are made looking into port 1 with port 2 alternately open and short circuited and looking into port 2 with port 1 alternately open and

[5] The commercial capabilities shown are based upon manufacturer's specifications.

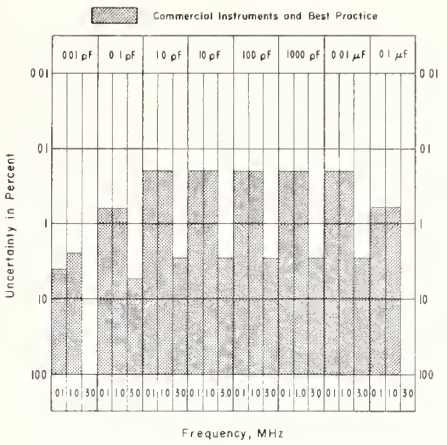


Fig. 18. State-of-the-art (three-terminal capacitance).

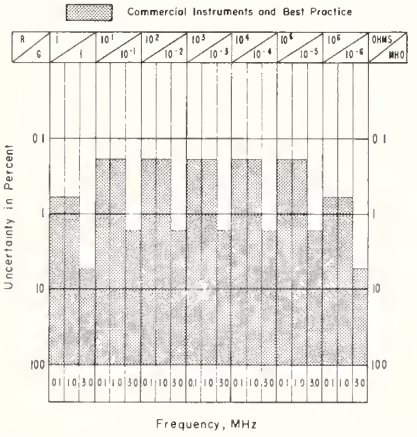


Fig. 19. State-of-the-art (three-terminal conductance).

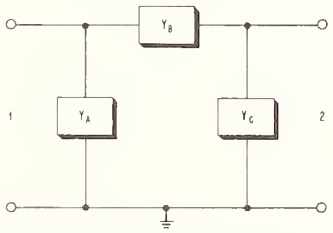


Fig. 20. Two-port (equivalent pi network).

short circuited. Any three of the resulting four two-terminal admittances may be used to solve for Y_A, Y_B, and Y_C. For best results, precision connectors are required at the ports so that the open- and short-circuited conditions are accurately known. Although two-port devices may be measured in this manner, simpler and more accurate measurements would be made possible by the development of special two-port standards.

X. Conclusions and Future Trends

The large differences in the state-of-the-art for commercial two-terminal instruments, as compared to best practices, are due almost entirely to the fact that precision coaxial connectors have not yet been utilized in commercial

instruments manufactured for the measurement of lumped parameter immittance. It is expected that this situation will change rapidly, especially where there is evidence of sufficient demand. If appropriate standards are made available along with improved instruments, the differences will be sharply reduced with very significant improvements realized over the entire frequency range of measurement for lumped immittances. Longer term benefits can be expected as well because new techniques in the construction of instruments will be made possible. For example, several different types of instruments have been constructed from one- and two-port immittance components which can be readily disconnected from one another and calibrated individually [27]. Such modular instruments offer many advantages including compactness, versatility, and ease of calibration and maintenance; however, they are not yet commercially available.

There are two areas where measurement capabilities are noticeably absent. These are for small two-terminal impedances at frequencies above 10 MHz and for three-terminal admittances above 5 MHz. It appears that there are no urgent requirements for such measurements, but it is often the case that there is no strong evidence of need for a particular measurement capability until it is developed, and then numerous applications arise.

To achieve the most effective and efficient standardization program agreement is needed regarding the specific frequencies and specific values for which standards are to be maintained. In this way, the advantages of comparator techniques can be utilized for the accurate determination of differences and measurements could be subjected to statistical control. Because of the lack of agreement in these matters such statistical control is not now possible.

References

[1] R. C. Powell, R. M. Jickling, and A. E. Hess, "High-frequency impedance standards at the National Bureau of Standards," *IRE Trans. Instrumentation*, vol. I-7, pp. 270–274, December 1958.

[2] Ramo, Whinnery, and VanDuzer, *Fields and Waves in Communication Electronics*. New York: Wiley, 1965, ch. 5.

[3] J. R. Carson, "Electromagnetic theory and the foundations of electric circuit theory," *Bell Sys. Tech. J.*, vol. VI, pp. 1–17, January 1927.

[4] W. C. Johnson, *Transmission Lines and Networks*. New York: McGraw-Hill, 1950, pp. 239 ff.

[5] D. Woods, "A coaxial connector system for precision RF measuring systems," *Proc. IEE (London)*, vol. 108B, pp. 205–213, March 1961.

[6] ——, "A precision VHF twin-*T* dual admittance bridge and associated standards," Rept. A.I.D./Rad. Elec./7, January 1950.

[7] R. N. Jones and L. E. Huntley, "Precision coaxial connectors in lumped parameter immittance measurement," *IEEE Trans. Instrumentation and Measurement*, vol. IM-15, pp. 375–380, December 1966.

[8] S. A. Schelkunoff, *Electromagnetic Waves*, 2nd. ed. Princeton, N. J.: Van Nostrand, 1943, p. 280.

[9] N. Marcuvitz, *Waveguide Handbook*, M.I.T. Radiation Lab. Ser., no. 10. New York: McGraw-Hill, 1951.

[10] D. E. Fossum, "Progress report of the IEEE instrumentation and measurement group technical subcommittee on precision coaxial connectors," *IEEE Trans. Instrumentation and Measurement*, vol. IM-13, pp. 285–291, December 1964.

[11] J. Zorzy, "Skin-effect corrections in immittance and scattering coefficient standards employing precision air-dielectric coaxial lines," *IEEE Trans. Instrumentation and Measurement*, vol. IM-15, pp. 358–364, December 1966.

[12] R. F. Field and D. B. Sinclair, "A method for determining the residual inductance and resistance of a variable air condenser at radio frequencies," *Proc. IRE*, vol. 24, pp. 255–274, February 1936.

[13] L. Hartshorn, *Radio Frequency Measurements by Bridge and Resonance Methods*. New York: Wiley, 1940.

[14] R. N. Jones, "A technique for extrapolating the 1 kc values of secondary capacitance standards to higher frequencies," Nat'l Bur. Standards, Washington, D. C., NBS Tech. Note 201, November 15, 1963.

[15] A. Russell, "The effective resistance and inductance of a concentric main and methods of computing the ber and bei and allied functions," *Phil. Mag.*, vol. 17, pp. 524–552, 1909.

[16] J. A. Stratton, *Electromagnetic Theory*. New York: McGraw-Hill, 1941, p. 553.

[17] R. E. Nelson and M. R. Coryell, "Electrical parameters of coaxial air-dielectric transmission lines," Nat'l Bur. Standards, Washington, D. C., NBS Mono. 96, June 30, 1966.

[18] A. M. Thompson and D. G. Lampard, "A new theorem in electrostatics and its application to calculable standards of capacitance," *Nature*, vol. 177, p. 888, May 1956.

[19] W. L. Everitt and G. E. Anner, *Communication Engineering*, 3rd. ed. New York: McGraw-Hill, 1956, pp. 163–164.

[20] G. J. Alonzo, R. H. Blackwell, and H. V. Marantz, "Direct-reading, fully-automatic vector impedance meters," *Hewlett-Packard J.*, January 1967.

[21] L. Hartshorn and W. H. Ward, "The measurement of the permittivity and power factor of dielectrics at frequencies from 10^4 to 10^8 cycles per second," *J. IEE (London)*, vol. 79, pp. 597–609, November 1936.

[22] J. J. Tiemann, "A note on the construction of wideband admittance bridges," *IEEE Trans. Instrumentation and Measurement*, vol. IM-15, pp. 2–5, March–June 1966.

[23] W. N. Tuttle, "Bridged-*T* and parallel-*T* null circuits for measurements at radio frequencies," *Proc. IRE*, vol. 28, pp. 23–29, January 1940.

[24] H. T. Wilhelm, "Impedance bridges for the megacycle range," *Bell Sys. Tech. J.*, vol. 31, pp. 999–1012, September 1952.

[25] R. A. Soderman, "A new bridge for the measurement of impedance between 10 and 165 Mc," *Gen. Rad. Exper.*, vol. 24, no. 19, February 1950.

[26] J. H. Mennie, "A wide range VHF *Z* meter," *BRC Notebook*, no. 2, Summer 1954.

[27] "Modular technique gives accurate HF instruments," *NBS Tech. News Bull.*, vol. 50, pp. 60–62, April 1966.

[28] D. B. Sinclair, "The twin-*T*, a new type of null instrument for measuring impedance at frequencies up to 30 megacycles," *Proc. IRE*, vol. 28, pp. 310–318, July 1940.

[29] D. Woods, "A precision dual bridge for the standardization of admittance at very high frequencies," *Proc. IEE (London)*, vol. 104C, Mono. 244R, pp. 506–521, June 1957.

[30] L. E. Huntley, "A self-calibrating instrument for measuring conductance at radio frequencies," *J. Res. NBS*, vol. 69C, pp. 115–126, April–June 1965.

[31] R. Calvert, "The transformer ratio-arm bridge," Wayne Kerr, Philadelphia, Pa., Mono. 1.

[32] L. Hartshorn, *loc. cit.*, pp. 38–66.

[33] B. Hague, *Alternating Current Bridge Methods*, 5th ed. London: Sir Isaac Pitman and Sons, 1957, pp. 522 ff.

[34] J. M. Parkyn, "Circuit magnification meter," *Marconi Instr.*, vol. 7, pp. 8–16, March 1959.

[35] L. Cook, "A versatile instrument—The *Q*-meter," *BRC Notebook*, no. 4.

[36] C. Quinn, "The BRC UHF *Q*-meter—A new and versatile tool for industry," *BRC Notebook*, no. 28, Spring 1961.

[37] *Die Kurzinformation*, Rohde & Schwarz, Passaic, N. J., no. 10, p. 11, 1964.

[38] C. H. Young, "Measuring inter-electrode capacitances," *Bell Labs. Rec.*, vol. 24, pp. 433–438, December 1946.

Other good references on the general subject of immittance measurement include:

[39] B. Hague, *op. cit.*

[40] F. K. Harris, *Electrical Measurements*. New York: Wiley, 1952.

[41] L. Hartshorn, *op. cit.*

[42] M. Wind, Ed., *Handbook of Electronic Measurements*, vol. 1. Brooklyn, N. Y.: Polytechnic Institute of Brooklyn, 1956, ch. 4, pp. 4-8 to 4-56.

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 65C, No. 4, October–December 1961

A Guide to the Use of the Modified Reflectometer Technique of VSWR Measurement *

Wilbur J. Anson

(February 10, 1961)

The theory of the modified reflectometer technique of measuring VSWR at microwave frequencies is discussed briefly, the operational procedure is outlined, and selected results are given of an unpublished error analysis. Much of the theory and procedure has been published in isolated papers. This paper unifies those details essential to the use of this technique and includes procedural suggestions that have grown out of extensive experience with the technique. The error analysis provides the means to evaluate the accuracy of any particular measurement made with this system.

1. Introduction

The modified reflectometer technique [Beatty and Kerns, 1958; Beatty, 1959; Engen and Beatty, 1959] recently developed at the Boulder Laboratory of the National Bureau of Standards is a method of making accurate VSWR measurements in rectangular waveguide. This technique is capable of measuring low VSWR's (VSWR<2) to accuracies of 1 percent quite easily. Accuracies of 0.1 percent in VSWR are obtainable with extra care in the adjustments of the modified reflectometer.

The purpose of this paper is to provide operating knowledge of the modified reflectometer technique. Much of the material presented herein has been published elsewhere but this paper gathers in one place those details essential to the use of this technique. The paper also offers suggestions on procedure which have grown out of extensive experience with this technique. Also it provides selected results of an error analysis which are important in evaluating the accuracy of any particular measurement made with this system.

2. An Outline of the Theory

The modified reflectometer measures the magnitude of the reflection coefficient of a waveguide termination at microwave frequencies. This method utilizes a directional coupler oriented to couple to the reflected wave with appropriately placed tuners adjusted so that the amplitude of the voltage wave from the side arm is directly proportional to the magnitude of the reflection coefficient.

According to references 1 and 2, the scattering equations for a three arm junction representing the combination of a directional coupler and two tuners (fig. 1) are

$$\begin{aligned} b_1&=S_{11}a_1+S_{12}a_2+S_{13}a_3\\ b_2&=S_{21}a_1+S_{22}a_2+S_{23}a_3\\ b_3&=S_{31}a_1+S_{32}a_2+S_{33}a_3, \end{aligned} \tag{1}$$

where $a_1=b_g+b_1\Gamma_G$, $a_2=b_2\Gamma_L$, and $a_3=b_3\,\Gamma_D$. These may be solved for the response in the form

$$b_3=b_g k\,\frac{(1+K\Gamma_L)}{(1-\Gamma_{2i}\Gamma_L)}, \tag{2}$$

$$k=\frac{S_{31}}{\begin{vmatrix}(1-S_{11}\Gamma_G) & S_{13}\Gamma_D\\ S_{31}\Gamma_G & (1-S_{33}\Gamma_D)\end{vmatrix}}, \tag{3}$$

$$K=\frac{\begin{vmatrix}S_{21} & S_{22}\\ S_{31} & S_{32}\end{vmatrix}}{S_{31}}, \tag{4}$$

and

$$\Gamma_{2i}=\frac{\begin{vmatrix}-(1-S_{11}\Gamma_G) & S_{12} & S_{13}\Gamma_D\\ S_{21}\Gamma_G & S_{22} & S_{23}\Gamma_D\\ S_{31}\Gamma_G & S_{32} & -(1-S_{33}\Gamma_D)\end{vmatrix}}{\begin{vmatrix}(1-S_{11}\Gamma_G) & S_{13}\Gamma_D\\ S_{31}\Gamma_G & (1-S_{33}\Gamma_D)\end{vmatrix}}. \tag{5}$$

Rearranged slightly, the equation becomes

$$b_3=\frac{b_g}{\begin{vmatrix}(1-S_{11}\Gamma_G) & S_{13}\Gamma_D\\ S_{31}\Gamma_G & (1-S_{33}\Gamma_D)\end{vmatrix}}\;\frac{S_{31}+\begin{vmatrix}S_{21} & S_{22}\\ S_{31} & S_{32}\end{vmatrix}\Gamma_L}{1-\Gamma_{2i}\Gamma_L}. \tag{6}$$

Now let $S_{31}=0$ and $\Gamma_{2i}=0$, then

$$b_3=\frac{b_g S_{21}S_{32}\Gamma_L}{(1-S_{11}\Gamma_G)(1-S_{33}\Gamma_D)}. \tag{7}$$

If b_g, Γ_G, and Γ_D are constant, then

$$b_3=C_{(\text{constant})}\Gamma_L. \tag{8}$$

For a detector not phase sensitive

$$|b_3|=|C\Gamma_L|. \tag{9}$$

Note that this last relationship is the desired response mentioned just before (1), i.e., the amplitude of the voltage wave from the sidearm is directly proportional to the magnitude of the reflection coefficient.

*Contribution from the Radio Standards Laboratory, National Bureau of Standards, Boulder, Colo.

Therefore, let $\Gamma_L=\Gamma_u$ (the unknown) so that $|b_3|_u=|C\Gamma_u|$. Then let $\Gamma_L=\Gamma_s$ (a known quantity) so that $|b_3|_s=|C\Gamma_s|$. Now form the ratio

$$\frac{|b_3|_u}{|b_3|_s}=\frac{|\Gamma_u|}{|\Gamma_s|}. \tag{10}$$

Thus, given the value of $|\Gamma_s|$, and measuring the ratio of the sidearm outputs $\frac{|b_3|_u}{|b_3|_s}$, the value of $|\Gamma_u|$ may be calculated.

3. Procedure

Any practical use of the modified reflectometer technique depends upon an understanding of the physical conditions imposed by the theory.

For $|b_3|$ to conform to (8) certain quantities as represented in (7) must remain constant. It requires a stable source (constant b_g) well isolated from load changes. It also requires apparent generator and detector impedances that are unchanging (constant Γ_G and Γ_D)—requirements which can be essentially fulfilled by isolators placed at terminal planes T_1 and T_3 in figures 1 and 2.

The remaining quantities of (8), namely S_{21}, S_{32}, S_{11}, and S_{33}, are constants of the 3-arm junction and therefore remain unchanging after adjustments are made for $S_{31}=0$ and $\Gamma_{2i}=0$.

3.1. Adjustment of $S_{31}=0$ and $\Gamma_{2i}=0$

For a directional coupler with the arms labeled as in figure 2 the statement that $S_{31}=0$ is equivalent to saying that the coupler has infinite directivity since, in terms of scattering coefficients, directivity $\equiv 20\log_{10}\left|\frac{S_{32}}{S_{31}}\right|$.

Infinite directivity can be approached by adjusting tuner A for no variation in the output $|b_3|$ when sliding a load of small reflection ($|\Gamma_L|<0.005$) in the uniform waveguide.

After making this adjustment, the condition $\Gamma_{2i}=0$ can be approached by adjusting tuner B for no variation in the output $|b_3|$ when sliding a load of large reflection (usually a shorting plunger) in the uniform waveguide.

This adjustment sequence may need to be repeated several times since the two adjustments are not completely independent. However, if the reflection from the sliding load used in the directivity adjustment is small enough, then the two adjustments are more nearly independent and one adjustment sequence will usually be enough.

NOTE: It is seldom necessary for the variation to be completely zero. The variation allowable while the loads are slid in each adjustment depends upon the desired accuracy of the measurement to be made. This allowable variation is discussed in the error analysis and is displayed in figures 5, 6, 7, and 8.

3.2. An Alternate Adjustment Procedure for $S_{31}=0$ and $\Gamma_{2i}=0$

Consider the definition for directivity in conjunction with (6) and observe that $|b_3|=0$ when $\Gamma_L=0$ if the directivity is infinite. Therefore the adjustment could be made by attaching a flat load ($\Gamma_L=0$) and adjusting tuner A for $|b_3|=0$. But there is no commercial termination flat enough for the accuracy of adjustment often needed and therefore an adjustable sliding termination such as the one described by R. W. Beatty [1957] must be adjusted for $\Gamma_L=0$.

The two adjustments for $S_{31}=0$ and $\Gamma_L=0$, using an adjustable sliding termination, are carried out alternately. The adjustable termination is slid and adjusted so that the variations in $|b_3|$ are reduced and then tuner A is adjusted so that the average level of the variation is lowered. When no variation in $|b_3|$ occurs as the adjustable termination is slid, *and* when $|b_3|$ is zero, then $S_{31}=0$ and $\Gamma_L=0$. Note that after adjustment of tuner A the variations will be greater than before. The greater variation results from increased sensitivity as infinite directivity is approached with the adjustments of tuner A.

Adjustments for $\Gamma_{2i}=0$ can be made after making the previous adjustment for infinite directivity. The symbol Γ_{2i} in (2), besides representing a group of symbols as noted in (5), actually represents the reflection coefficient "looking in" arm 2 of the junction as indicated in figure 2. Therefore the condition $\Gamma_{2i}=0$ is equivalent to saying that, at the terminal plane T_2, this input reflection coefficient is zero.

In figure 3 a scheme is illustrated for recognizing this matched condition. Power is coupled out of the main line to supply an auxiliary directional coupler

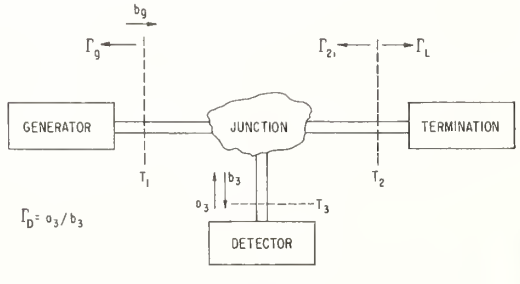


FIGURE 1. *Three arm junction representing the modified reflectometer.*

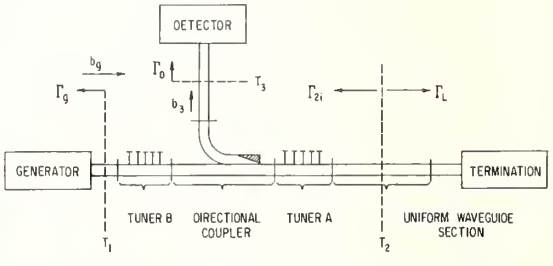


FIGURE 2. *Diagram of modified reflectometer.*

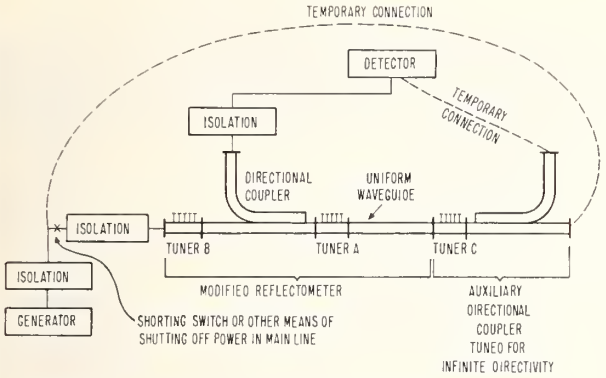


FIGURE 3. *Equipment arrangement for adjusting* $\Gamma_{2i}=0$.

tuned for infinite directivity. This coupler is temporarily attached to the uniform waveguide with power shut off in the main line supplying the modified reflectometer. Tuner B is then adjusted for no output from the sidearm of the auxiliary coupler. The condition $\Gamma_{2i}=0$ is thereby fulfilled.

There must, of course, be enough isolation on the generator side of the modified reflectometer to maintain essentially the same value of Γ_{2i} during both this matching procedure and the normal operation of the reflectometer. Also, there must be enough isolation between the generator and whatever means is used for shutting off the main line power to insure no possibility of frequency pulling.

The auxiliary directional coupler can be tuned for infinite directivity by utilizing the flat load previously discussed. This flat load is attached temporarily to the auxiliary coupler and tuner C adjusted for no output from the sidearm of the auxiliary coupler.

This adjustment sequence may also need to be repeated several times since the two adjustments are not completely independent. However, if the adjustment for infinite directivity is carefully made, one adjustment sequence will be enough.

Although this alternate procedure is sometimes easier to use, it does not provide sufficient information to evaluate the maximum errors due to imperfect adjustments of the tuners. Therefore, the adequacy of the adjustments must be checked by sliding the small and large reflection loads in the uniform waveguide and by observing whether the variations in $|b_3|$ are within the allowable limits for the measurement accuracy needed.

If the variations are not within the allowable limits it is important that the final adjustments be made with the small and large reflection sliding loads *in the uniform waveguide*. Otherwise, reflections from the waveguide joints will detune the infinite directivity adjustment.

3.3. Another Alternate Adjustment Procedure for $S_{31}=0$

By adding an auxiliary channel as illustrated in figure 4, the adjustment for infinite directivity ($S_{31}=0$) is made considerably easier and faster [Beatty, Engen, and Anson, 1960]. This auxiliary

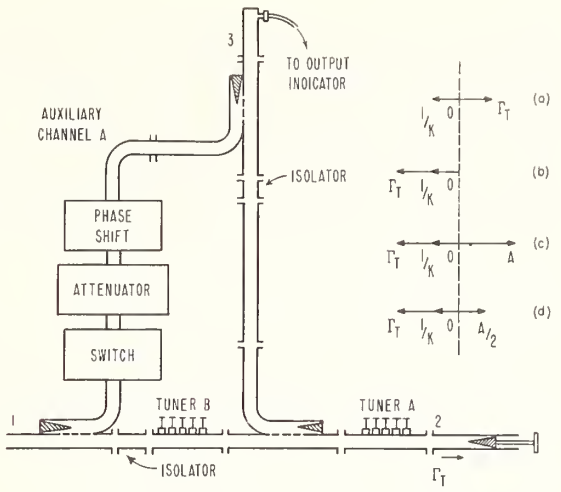


FIGURE 4. *Auxiliary channel technique.*

channel permits introduction of a signal of the proper phase and magnitude to null the signal component indicative of finite directivity.

From (2) the response from the sidearm is

$$|b_3|=|b_gK|\left|\frac{1+K\Gamma_L}{1-\Gamma_{2i}\Gamma_L}\right|.$$

If $|\Gamma_{2i}|$ is small, i.e., $|\Gamma_{2i}|<0.01$, then a very good approximation is

$$|b_3|=|b_gkK|\cdot\left|\frac{1}{K}+\Gamma_L\right|. \tag{11}$$

The following description of the procedure is best understood by referring to (11) and the vector diagrams in figure 4.

1. With the shorting switch in the auxiliary arm closed, arbitrarily position the small reflection load in the uniform waveguide and adjust tuner A for $b_3=0$. This corresponds to step (a) in figure 4 wherein $1/K$ is oppositely directed and equal in magnitude to $|\Gamma_L|$.

2. Slide the load for maximum $|b_3|$. This corresponds to step (b) in figure 4 wherein Γ_L and $1/K$ are in phase.

3. Open the waveguide switch in auxiliary channel A and adjust the phase shifter and attenuator in that channel for $|b_3|=0$ again. This corresponds to step (c) in figure 4 which illustrates a signal A having been introduced to null $\left|\frac{1}{K}\right|+|\Gamma_L|$.

4. Introduce 6.02 db of attenuation into the auxiliary arm to reduce the amplitude of signal A by one-half as illustrated in step (d) in figure 4. This A/2 cancels the $1/K$ term in (11) and results in infinite directivity. Recall that $|\Gamma_{2i}|$ is assumed small so that (11) is a good approximation.

The directivity obtained by this procedure can be checked by sliding the low reflection load and observing the variations in $|b_3|$ as noted in section 3.1. If the directivity obtained thusly is not high enough, then trial and error adjustments using the phase

shifter and the attenuator in the auxiliary arm or tuner A will allow closer adjustment. However, adjustments using the components of the auxiliary arm are in general easier because these components usually present better resolution than tuner A.

To avoid possible interaction between the two channels, two isolators should be included as shown in figure 4. Other tuning [Engen, 1960] could be employed to reduce interaction but these two isolators should suffice.

3.4. Measurement of $|\Gamma_u|$

Certain details vary from system to system and from operator to operator depending upon personal preferences, equipment available and the accuracies needed. Therefore, no attempt has been made to write this guide for a specific arrangement of equipment.

However, regardless of the arrangement of the apparatus in the system, the ratio $\left|\frac{\Gamma_u}{\Gamma_s}\right|$ can be conveniently measured by some adaptation of the following sequence:

A. Adjust $S_{31}=0$ and $\Gamma_{2i}=0$ by some combination of methods as previously outlined.

B. Check the adequacy of the adjustments to insure the achievement of the measurement accuracy required.

C. Attach the microwave termination of unknown reflection coefficient Γ_u to the uniform waveguide.

D. Set $|b_3|$ at some convenient reference level on some suitable detector and note the setting of a calibrated attenuator. This attenuator, of course, must be properly matched [Beatty, 1954; Schafer and Rumfelt, 1959] to measure accurately the voltage ratios at the detector.

E. Replace the unknown with some standard termination whose reflection coefficient is accurately known.

F. Return $|b_3|$ to the reference level by adjustment of the calibrated attenuator mentioned in step (D).

G. Record the difference in settings of the calibrated attenuator between step (D) and step (F). Calculate the value of $|\Gamma_u|$ from this difference, i.e., $R=20\log_{10}\left|\frac{\Gamma_s}{\Gamma_u}\right|$ where R is the difference in decibels. This calculation can be aided by reference to a table [Beatty and Anson, 1960] of reflection coefficient versus return loss.

4. The Error in the Measurement of $|\Gamma_u|$ and VSWR Due to $S_{31}\neq 0$ and $\Gamma_{2i}\neq 0$

As noted previously, the variation in $|b_3|$ allowable while the loads are slid in each adjustment depends upon the desired accuracy of the measurement made. This allowable variation can be computed from the following equations or taken from the graphs included for certain measurement situations. It is recommended that the adjustments be checked by this method both *before* and *after* measurements are made as an indication of changes created by accidental bumps or frequency drift.

4.1. The Error Due to $S_{31}\neq 0$

It can be shown that the maximum fractional error (assuming the worst phase combination) in the measurement of $|\Gamma_u|$, where $S_{31}\neq 0$ but all other conditions are met, is given by the expression

$$\frac{d|\Gamma_u|}{|\Gamma_u|}\leq\frac{1}{|K|}\left(\frac{|\Gamma_s|+|\Gamma_u|}{|\Gamma_u\Gamma_s|-\left|\frac{\Gamma_u}{K}\right|}\right) \tag{12}$$

where $\Gamma_s\equiv$reflection coefficient of the standard termination,

$\Gamma_u\equiv$reflection coefficient of the unknown termination,

and $$|K|=\frac{10^{R_1/20}+1}{(10^{R_1/20}-1)|\Gamma_L|},$$

in which R_1 is the db variation in $|b_3|$ as the small reflection load Γ_L is slid in the uniform waveguide.

The corresponding maximum error in the measurement of VSWR, when $S_{31}\neq 0$, is obtained from (12) applying the following relation:

$$\frac{d\sigma_u}{\sigma_u}=\frac{2|\Gamma_u|}{1-|\Gamma_u|^2}\frac{d|\Gamma_u|}{|\Gamma_u|}\text{ where }\sigma_u\equiv\text{VSWR}. \tag{13}$$

4.2. The Error Due to $\Gamma_{2i}\neq 0$

The maximum fractional error in the measurement of $|\Gamma_u|$, when $\Gamma_{2i}=0$ but all other conditions are met, is given by the expression

$$\frac{d|\Gamma_u|}{|\Gamma_u|}\leq\frac{(|\Gamma_u|+|\Gamma_s|)|\Gamma_{2i}|}{1-\Gamma_{2i}\Gamma_u} \tag{14}$$

where

$\Gamma_s\equiv$reflection coefficient of the standard termination,

$\Gamma_u\equiv$reflection coefficient of the unknown termination,

and $|\Gamma_{2i}|$ is obtained from

$$R_2=20\log_{10}\frac{1+|\Gamma_{2i}\Gamma_L|}{1-|\Gamma_{2i}\Gamma_L|}$$

in which R_2 is the db variation in $|b_3|$ observed as a large reflection load (usually a shorting plunger whose $|\Gamma_L|\cong 1$).

The corresponding maximum error in the measurement of VSWR, when $\Gamma_{2i}\neq 0$, is obtained from (14) by applying (13).

4.3. The Error Due to Both $S_{31}\neq 0$ and $\Gamma_{2i}\neq 0$

Seldom, if ever, will S_{31} or Γ_{2i} be adjusted perfectly. However, it can be shown that the maximum error due to both $S_{31}\neq 0$ and $\Gamma_{2i}\neq 0$ simultaneously is the sum of the individual maximum errors. Therefore, these individual errors may be obtained from (12) through (14), or from figures 5 through 8, and added to form the maximum error of the measurement due to $S_{31}\neq 0$ and $\Gamma_{2i}\neq 0$.

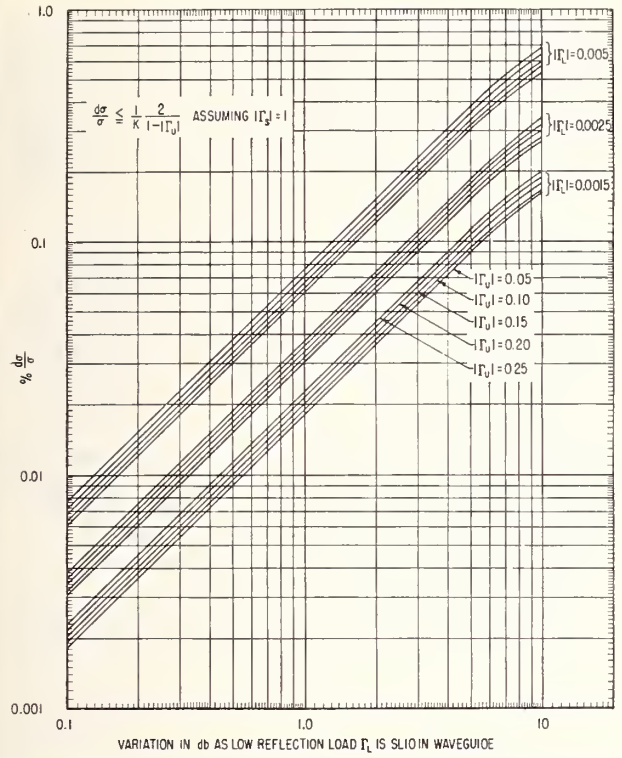


FIGURE 5. *The maximum error in measuring VSWR by the modified reflectometer technique due to* $S_{31} \neq 0$ *and assuming* $\Gamma_{2i} = 0$.

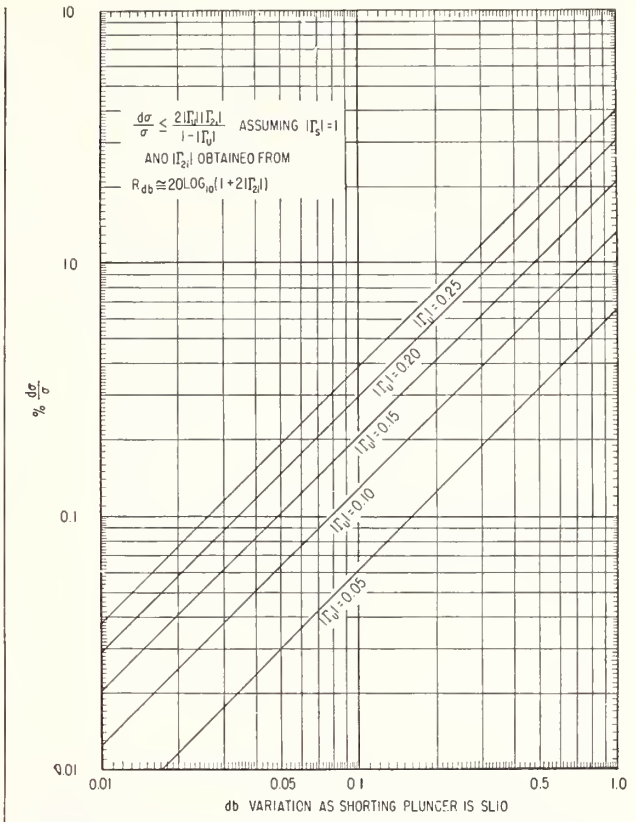


FIGURE 6. *The maximum error in measuring VSWR by the modified reflectometer technique due to* $\Gamma_{2i} \neq 0$ *and assuming* $S_{31} = 0$.

Example: Suppose there is 0.5 db variation in $|b_3|$ as a small reflection load $|\Gamma_L| < 0.005$ is slid in the uniform waveguide and there is a 0.02 db variation as a short ($|\Gamma_s| \approx 1$) is slid. What is the total limit of error in the measurement of the VSWR of a termination whose reflection coefficient is $|\Gamma_u| \approx 0.2$? From figure 5 the maximum error in measuring VSWR due to $S_{31} \neq 0$ is 0.036 percent; from figure 6 the maximum error due to $\Gamma_{2i} \neq 0$ is 0.058 percent. So the maximum error (due to both $S_{31} \neq 0$ and $\Gamma_{2i} \neq 0$) in measuring the VSWR of the termination is 0.094 percent.

5. Procedural Suggestions

1. Adjustment of the tuners for no variation in the sidearm response $|b_3|$ is easier if some systematic approach is used. One such approach is the following:

A. Slide the load for either a maximum or minimum response. In this example assume that it was slid for maximum.

B. Adjust a stub of the tuner in the direction which lowers the detector response. (Although this example uses a multiple-stub tuner the same procedure is usually adaptable to other types of tuners.)

C. Slide the load for a minimum.

D. Adjust the *same* stub in the same direction as before. If the response increases, i.e., moves toward the average of the variation, then further adjustment of that tuning stub using steps (A) through (D) repeatedly will systematically reduce the variation. If the response decreases, i.e., moves away from the average of the variation, adjustment of that stub will probably be useless and another stub should be tried.

The adjustments for no variation as the load is slid are usually quicker if several effective stubs are located by this procedure and then alternately adjusted.

If no effective stubs can be located, make some random adjustment of the tuner to raise or lower the detector response and repeat the previous procedure. This is a trial and error process. Experience with your own equipment will provide "feel" for the detector response level resulting when the sliding load variations have been reduced.

2. An approximation to the detector response level resulting when the variation in the sidearm output has been reduced for the infinite directivity can be obtained as follows:

A. Slide the low reflection termination for a minimum.

B. Adjust tuner A for a null.

C. Slide the low reflection termination for a maximum. About 6 db down from the maximum is the detector response level resulting when infinite directivity is achieved. This level can serve as a guide when making the adjustments for no variation as the load is slid.

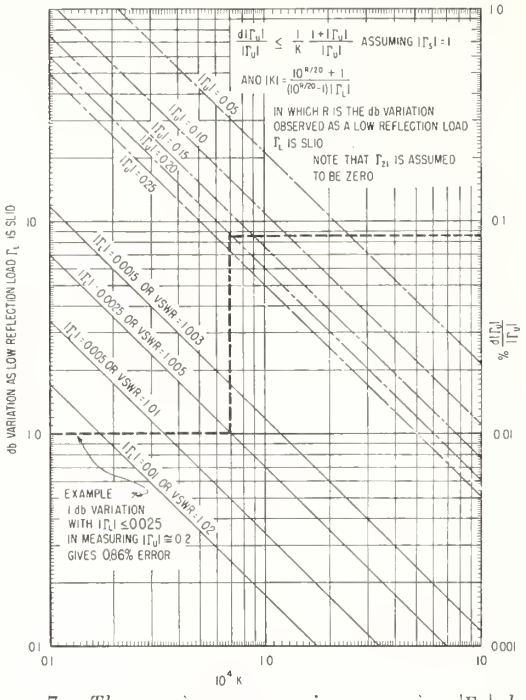


FIGURE 7. *The maximum error in measuring* $|\Gamma_u|$ *by the modified reflectometer technique due to* $S_{31} \neq 0$ *and assuming* $\Gamma_{2i}=0$.

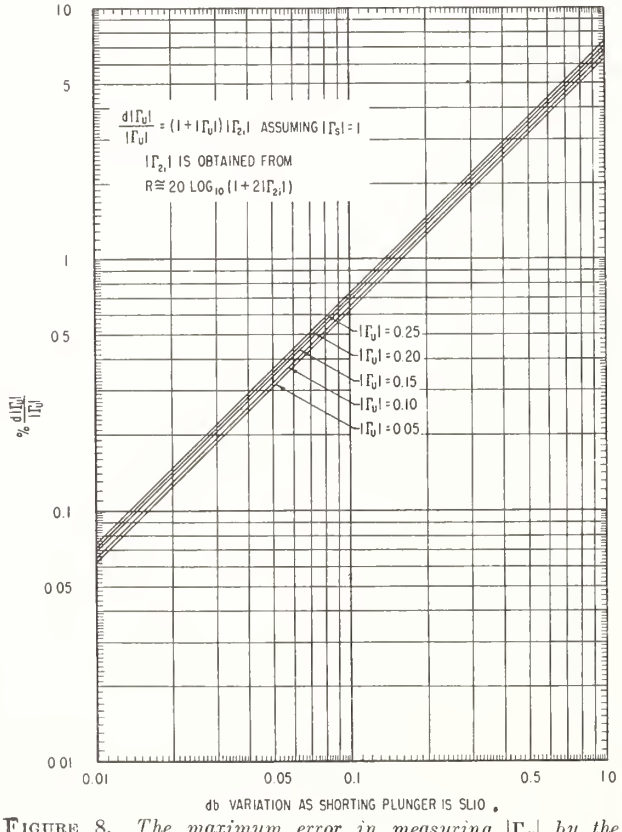


FIGURE 8. *The maximum error in measuring* $|\Gamma_u|$ *by the modified reflectometer technique due to* $\Gamma_{2i} \neq 0$ *and assuming* $S_{31}=0$.

This procedure avoids the one possible false adjustment which could result in an approximately constant output with $\left|\frac{1}{K}\right| \gg |\Gamma_T|$. This misleading adjustment is seldom encountered except when adjusting the equipment the first time. It sometimes occurs at that time because the system lacks sufficient power sensitivity for detection of the approximately constant output signifying $\left|\frac{1}{K}\right| << |\Gamma_T|$ which occurs at a low power level.

3. It is recommended that the adjustments for S_{31} and Γ_{2i} be checked both *before* and *after* a VSWR or reflection coefficient measurement is made to detect possible changes created by accidental bumps or frequency drift.

4. The reflection coefficient magnitude $|\Gamma_L|$ of the small reflection load need not be determined to any great accuracy. Computation of the maximum error requires only that the actual value of $|\Gamma_L|$ is less than the value used in the computation.

This reflection coefficient magnitude can be determined accurately enough as follows:

A. Make the normal adjustments of the system, repeating the adjustment sequence until adjustment for $\Gamma_{2i}=0$ does not require further adjustment of $S_{31}=0$.

B. Slide the small reflection load and observe the maximum and minimum $|b_3|$. There should be little, if any, detectable difference between the minimum and the maximum. Therefore, it is usually sufficient to position the sliding load for the average response and measure $|\Gamma_L|$ for this position using the magnified reflectometer technique. For a truly conservative figure, however, measure the reflection coefficient magnitude of the load in the position of maximum $|b_3|$.

If the standard termination being used is a "short" ($|\Gamma_s|=1$) the ratio $R=20 \log_{10} \frac{|\Gamma_s|}{|\Gamma_u|}$ may be too large for the calibrated attenuator of the system. Then it may be necessary to use a standard of smaller reflection coefficient in order to measure $|\Gamma_u|$.

5. The adjustment of the tuners as outlined in 5.1 is an efficient technique if effective stubs can be found. The following technique aids in finding effective stubs for the infinite directivity adjustment.

A. Adjust tuner A and the sliding load for $|b_3|=0$.

B. Slide the load for maximum $|b_3|$.

C. Now at least one of the stubs is an "effective" stub. Continue the adjustment as outlined in 5.1.

6. When using the small reflection load ($|\Gamma_L| < 0.005$) the signal level may disappear into the noise as the adjustments for infinite directivity are attempted. This difficulty indicates that the detection system does not have sufficient power sensitivity. Rather than invest in more sensitive equipment, a load with greater reflection (perhaps $|\Gamma_L| \sim 0.025$) can be used for the infinite directivity adjustment provided the adjustment sequence outlined in section 3.1 is repeated as indicated. However, this alternate procedure is considerably more time consuming.

The author is indebted to Edward Niesen for his invaluable assistance in the operation and maintenance of the equipment, to R. W. Beatty for his encouragement and helpful suggestions during the writing of this guide, and to David Wait and Michael Brady for their work on the error analysis.

6. References

Beatty, R. W., Mismatch errors in the measurement of ultra high frequency and microwave variable attenuators, J. Research NBS **52**, 7 (Jan. 1954).

Beatty, R. W., An adjustable sliding termination for rectangular waveguide, IRE Trans. on Microwave Theory and Tech. **5**, No. 3 (July 1957).

Beatty, R. W., Magnified and squared VSWR responses, IRE Trans. on Microwave Theory and Tech. **7**, No. 3 (July 1959).

Beatty, R. W., and W. J. Anson, Table of magnitude of reflection coefficient versus return loss $\left(L_R = 20 \log_{10} \frac{1}{|\Gamma|}\right)$, NBS Tech. Note 72 (Sept. 1960).

Beatty, R. W., G. F. Engen, and W. J. Anson, Measurement of reflections and losses of waveguide joints and connectors using microwave reflectometer techniques, IRE Trans. on Instrumentation **I-9**, No. 2, 219–226 (Sept. 1960).

Beatty, R. W., and D. M. Kerns, Recently developed microwave impedance standards and methods of measurement, IRE Trans. on Instrumentation, p. 319 (Dec. 1958).

Engen, G. F., A method of improving isolation in multichannel waveguide systems, IRE Trans. on Microwave Theory and Tech. **8**, p. 460 (July 1960).

Engen, Glenn F., and R. W. Beatty, Microwave reflectometer techniques, IRE Trans. on Microwave Theory and Tech. 351–355 (July 1959).

Schafer, G. E., and A. Y. Rumfelt, Mismatch errors in cascade connected variable attenuators, IRE Trans. on Microwave Theory and Tech. **7**, p. 447 (Oct. 1959).

(Paper 65C4—74)

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 70C, No. 3, July–September 1966

Precise Reflection Coefficient Measurements With an Untuned Reflectometer

W. E. Little* and D. A. Ellerbruch*

Institute for Basic Standards, National Bureau of Standards, Boulder, Colo.

(March 9, 1966)

The precision tuned reflectometer technique of reflection coefficient measurement is, at present, the most accurate technique available. Utilization of the technique requires the use of tuners and sliding terminations to eliminate or reduce the inherent errors of a directional coupler.

This paper describes a reflection coefficient measurement technique that approaches the accuracy of the tuned reflectometer. Instead of tuners, the technique uses an ideal one-quarter wavelength waveguide section to eliminate the measurement error introduced by finite directional coupler directivity and to reduce to second order the error introduced by an equivalent generator mismatch.

A comparison of the quarter-wave technique with the tuned reflectometer technique through the measurement of the reflection coefficient of three 9/16 inch coaxial waveguide terminations is also included.

Key Words: Microwave, impedance, measurements, reflectometer.

1. Introduction

The reflectometer technique for microwave reflection coefficient measurements is now widely used [1].[1] The measurement accuracy of a basic reflectometer (a simple directional coupler coupled to the reflected wave) depends on the directivity of the directional coupler and the equivalent generator mismatch. The best known method of improving the measurement accuracy of the basic reflectometer requires the use of tuners, tuning procedures, and auxiliary equipment [2, 3]. The result of this tuning is an increase in the effective directivity of the coupler and a matched equivalent generator. At present, the most accurate measurements are achieved with this tuned reflectometer.

The purpose of this paper is to present another method of improving the measurement accuracy of the basic reflectometer technique. This method utilizes a quarter-wavelength waveguide section. Proper use of the quarter-wave section during the measurement procedure will, in theory, eliminate the error caused by a finite directivity and will reduce the error caused by a mismatched equivalent generator to second order.

No detailed error analysis is included in this paper; however, a laboratory measurement comparison with the tuned reflectometer indicates that the total measurement error is small. Such an error analysis may be given at a later date, depending on results of further laboratory experiments with the method.

2. Theory

The operation of a reflectometer is discussed in detail elsewhere [3]. Therefore, only a brief review will be presented here as a starting point for the discussion of the quarter wavelength technique. The side-arm output, b_3, of a directional coupler connected to sample the reflected wave (see fig. 1) can be expressed as

$$b_3 = kb_G \frac{\frac{1}{K} + \Gamma_T}{1 - \Gamma_{2i}\Gamma_T}, \tag{1}$$

where b_G is the component of the incident wave furnished by the generator, Γ_{2i} and k are both functions of the scattering coefficients of the directional coupler and reflection coefficients of the generator and detector, and Γ_T is the reflection coefficient of the terminating element. The symbol Γ_{2i} represents the equivalent source reflection coefficient presented to the reflector under study at the reference plane. The symbol K is approximately equal to the directivity ratio of the directional coupler.

The desired performance of the directional coupler is realized when $\Gamma_{2i}=0$ and $1/K=0$. Equation (1) then reduces to

$$b_3 = kb_G\Gamma_T. \tag{2}$$

Hence, inspection of (1) shows that the terms involving $1/K$ and Γ_{2i} introduce first order errors in Γ_T, when Γ_T is determined according to (2) from a measurement of

* Radio Standards Engineering Division. National Bureau of Standards, Boulder, Colo.

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

the side-arm output, b_3, of an untuned ($1/K$ and $\Gamma_{2i} \neq 0$) directional coupler. Reference 3 discusses the minimization of these errors by proper use of tuners.

It is possible to eliminate the term $1/K$ and simultaneously reduce the error arising from a finite Γ_{2i} to second order by a measurement procedure utilizing an ideal waveguide section that is one-quarter wavelength long. The theoretical basis for the method is as follows:

If the unknown is connected directly to the reference plane of the untuned reflectometer shown in figure 1, the side-arm output, $b_{1\mu}$, is

$$b_{1\mu} = kb_G \frac{\frac{1}{K} + \Gamma_\mu}{1 - \Gamma_{2i}\Gamma_\mu} \tag{3}$$

where Γ_μ is the reflection coefficient of the unknown.

Interposing a precision quarter-wavelength waveguide section between the reflectometer and Γ_μ adds a 180° phase shift to the argument of the impedance appended to the reference plane, causing the reflection coefficient at this plane to be $-\Gamma_\mu$. With the quarter-wavelength section inserted between the reference plane and the unknown, as shown in figure 2, the side-arm output, $b_{2\mu}$, is

$$b_{2\mu} = kb_G \frac{\frac{1}{K} - \Gamma_\mu}{1 + \Gamma_{2i}\Gamma_\mu}. \tag{4}$$

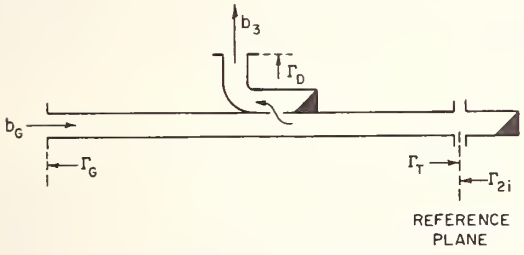


FIGURE 1. *A basic reflectometer.*

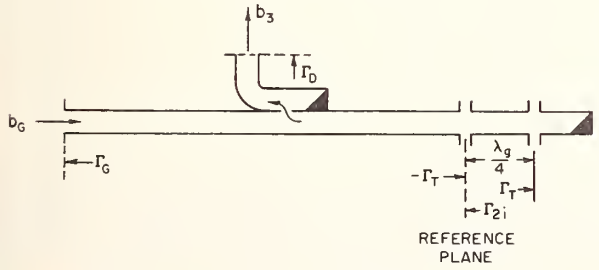


FIGURE 2. *A basic reflectometer with a quarter-wave section connected.*

The difference between $b_{1\mu}$ and $b_{2\mu}$ is then

$$b_{1\mu} - b_{2\mu} = 2kb_G\Gamma_\mu \frac{1 + \frac{1}{K}\Gamma_{2i}}{1 - (\Gamma_{2i}\Gamma_\mu)^2}. \tag{5}$$

If the unknown is now replaced with a quarter-wavelength standard short circuit[2] whose reflection coefficient is unity and the above measurement procedure repeated, formula (5), upon substitution of $\Gamma_s = 1$ for Γ_μ, reduces to

$$b_{1s} - b_{2s} = 2kb_G \frac{1 + \frac{1}{K}\Gamma_{2i}}{1 - (\Gamma_{2i})^2}, \tag{6}$$

where b_{1s} is the sidearm output when the standard short circuit is connected directly to the reference plane, and b_{2s} is the sidearm output when the ideal quarter-wavelength section is placed between the reference plane and the short circuit.

Dividing (6) by (5) results in

$$\frac{b_{1s} - b_{2s}}{b_{1\mu} - b_{2\mu}} = \frac{1}{\Gamma_\mu} \frac{1 - (\Gamma_{2i}\Gamma_\mu)^2}{1 - \Gamma_{2i}^2}. \tag{7}$$

Note that (7) does not contain the $1/K$ factor and that Γ_{2i} occurs only to the second power. $|\Gamma_{2i}|$ is usually less than 0.05 in a practical situation; hence, $|\Gamma_{2i}|^2 \ll 1$. Therefore, the limitations of the untuned directional coupler (i.e., finite directivity and a mismatched equivalent generator) have been essentially circumvented, and the correct value of Γ_μ can be determined without tuning the coupler.

To evaluate Γ_μ by means of (7), it appears that four individual (complex) quantities have to be measured. If one is interested only in a magnitude measurement, which is usually the case, a relatively simple measurement procedure can be used to determine the ratio $|b_{1s} - b_{2s}|/|b_{1\mu} - b_{2\mu}|$ directly, in decibels.

3. Measurement Procedures for Determining the Modulus of Γ_μ

A measurement system has been constructed and used to substantiate the quarter-wave method experimentally. A block diagram of the system is shown in figure 3. The measurement procedure is as follows:

1. One end of the precision quarter-wavelength section of waveguide is connected to the reference plane and terminated with a standard short circuit. The signal from the side arm under these conditions is b_{2s} of formula (7).

[2] A quarter-wavelength standard short circuit is a quarter wavelength of precision line terminated by a shorting metal wall. If a simple shorting plate is used, $\Gamma_s = -1$ is substituted for Γ_μ in (5) and one side of (7) would be multiplied by a -1.

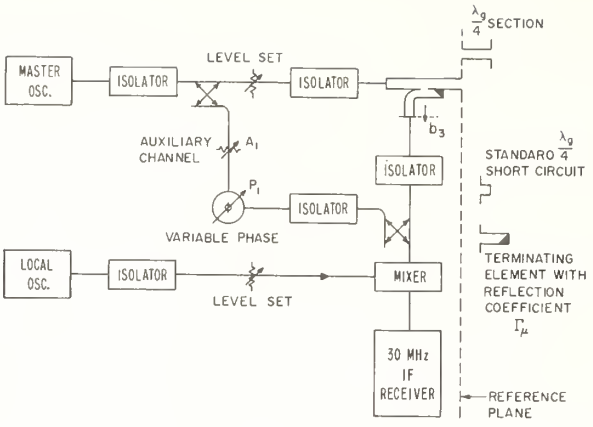


FIGURE 3. *Block diagram of quarter-wavelength measurement system.*

2. The variable phase shifter and attenuator, designated P_1 and A_1 in figure 3, are adjusted to produce a null at the 30 MHz IF receiver. This adjustment sets the signal from the auxiliary channel, b_{1A}, equal to $-b_{2s}$.

3. The ideal quarter-wavelength section is removed and the standard short circuit is connected directly to the reference plane. Under these conditions the signal from the side arm is b_{1s} of (7). Because the signal from the auxiliary channel of figure 3 was adjusted in the preceding step to produce the signal $-b_{2s}$, the receiver, which responds to the sum of the outputs of the side-arm and auxiliary channel, registers $|b_{1s}-b_{2s}|$.

4. The standard 30 MHz attenuator in the IF receiver is adjusted to produce an arbitrary reference level at the indicator.

5. The short circuit is removed from the reference plane, and the quarter-wavelength section is connected and terminated in the unknown reflection coefficient. Under these conditions the signal out of the side-arm is equal to $b_{2\mu}$ of (7).

6. The phase shifter and attenuator (P_1 and A_1) are now readjusted to again produce a null at the 30 MHz receiver. This adjustment provides that $b_{A2}=-b_{2\mu}$, where b_{A2} is the signal from the auxiliary channel.

7. The quarter-wavelength section is removed, and the unknown is connected directly to the reference plane. The signal into the 30 MHz IF receiver is again the sum of the signals from the side-arm and auxiliary channel, i.e., $|b_{1\mu}-b_{2\mu}|$. A record is made of the change (in decibels) in the setting of the standard 30 MHz attenuator in the receiver required to return the indicator to the reference level chosen in step four. This difference (in decibels) between the final attenuator setting and the setting at step four is a measure of the ratio $|b_{1s}-b_{2s}|/|b_{1\mu}-b_{2\mu}|$. Thus,

$$\mathrm{dB(measured)}=20\log\left|\frac{b_{1s}-b_{2s}}{b_{1\mu}-b_{2\mu}}\right|$$

$$=20\log\frac{1}{|\Gamma_\mu|}\frac{|1-(\Gamma_{2i}\Gamma_\mu)^2|}{|1-\Gamma_{2i}^2|}.$$

Assuming that $|\Gamma_{2i}^2|\leq 1$, this equation reduces to approximately

$$\mathrm{dB(measured)}=20\log\frac{1}{|\Gamma_\mu|},$$

whence the unknown reflection coefficient magnitude is given by

$$|\Gamma_\mu|=10^{-\frac{\mathrm{dB(measured)}}{20}}.$$

4. Measurement Results

The reflection coefficient magnitude of three 9/16-inch coaxial terminations were measured at 4 GHz using the quarter-wave technique. These measurements were compared with those obtained with a tuned reflectometer and those obtained with an untuned reflectometer that did not utilize a quarter-wavelength section. Table 1 summarizes the results.

The first column lists the values obtained for the three terminations when measured with a tuned reflectometer. The tolerances associated with these values include the measurement error as well as the non-repeatability obtained by rotating the termination. In the second column are listed the VSWR's corresponding to the reflection coefficients in the first column.

TABLE 1. *Tabulated measurement results*

TUNED REFLECTOMETER MEASUREMENT VALUES		VALUES OBTAINED WITH IMPERFECT COUPLER		VALUES OBTAINED WITH IMPERFECT COUPLER AND 1/4λ SECTION	
$\|\Gamma\|$	VSWR	$\|\Gamma\|$	VSWR	$\|\Gamma\|$	VSWR
0.0217 ±0.0006	1.0443 ±0.0012	0.042	1.088	0.0224	1.0458
0.0902 ±0.0015	1.1983 ±0.003	0.108	1.242	0.0918	1.2022
0.3334 ±0.0028	2.000 ±0.006	0.351	2.082	0.3353	2.0089

In the third and fourth columns are listed the reflection coefficients and VSWR's obtained with an untuned directional coupler but no quarter-wavelength section. These values were obtained as follows: The standard short circuit was connected to the reference plane, and an indicator reference point

was obtained. The short circuit was then replaced by the unknown impedance, and the reflection coefficient was computed from the measured change in the signal level at the indicator. These values are subject to all the errors due to imperfections in the directional coupler. The directional coupler employed in the three systems with which the data of table 1 were obtained (including that of fig. 3), had a directivity of approximately 30 dB (this corresponds to a $1/|K|$ value of approximately 0.032) and a $|\Gamma_{2i}|$ of less than 0.02.

In the fifth and sixth columns of table 1 are listed the reflection coefficient magnitude and VSWR values obtained by using the quarter-wavelength measurement technique described in this report.

The values obtained using the quarter-wave technique are in near agreement with the values obtained using the tuned reflectometer. The values obtained with the untuned reflectometer when the quarter-wave section was not used are considerably in error.

5. Conclusions

Precise reflection coefficient measurements are possible with an untuned reflectometer if the quarter-wave measurement technique is used. The theory was verified experimentally by measuring the reflector coefficient magnitudes of several terminations with a tuned reflectometer and then remeasuring them using the quarter-wave technique with an untuned reflectometer. These same terminations were also measured directly with an untuned reflectometer without application of the quarter-wavelength technique, to demonstrate the improvement in accuracy provided by the quarter-wave method.

Relative phase data were neglected during this evaluation. No serious difficulties are anticipated in modifying the measurement system to obtain phase data so that the quarter-wave technique will then provide the complex reflection coefficient.

Use of the quarter-wave technique instead of a tuning procedure eliminates the need for variable-phase high- and low-reflection loads and shortens the required length of precision line. The most important potential application of the quarter-wave technique is in the measurement of impedance at frequencies below approximately 1 GHz. With decreasing frequency the tuned reflectometer becomes increasingly cumbersome.

6. References

[1] Ginzton, E. L., Microwave Measurements, p. 300 (McGraw-Hill Book Co., Inc., New York, N.Y., 1957).

[2] Engen, G. F., and R. W. Beatty, Microwave reflectometer techniques, IRE Trans. MTT **MTT-7**, No. 3, 351–355 (July 1959).

[3] Anson, Wilbur J., A Guide to the use of the modified reflecmeter technique of VSWR measurement, J. Res. NBS **65C** (Engr. and Instr.) No. 4, 217 (Oct.-Dec. 1961).

(Paper 70C3–227)

Measurement of Reflections and Losses of Waveguide Joints and Connectors Using Microwave Reflectometer Techniques*

R. W. BEATTY†, SENIOR MEMBER, IRE, G. F. ENGEN†, AND W. J. ANSON†

Introduction

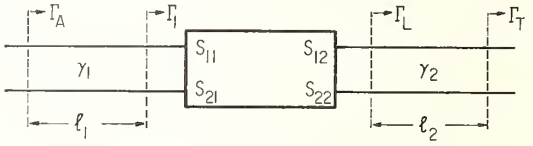
A KNOWLEDGE of the reflections and losses of waveguide joints or connectors is important in evaluating certain errors occurring in nearly all types of microwave measurements. In some cases, these errors are the limiting ones, and further improvement in the state of the art depends upon improvements in joints or connectors. Sensitive and accurate techniques to measure the characteristics of joints and connectors are vital to such an improvement.

The improvement and refinement of microwave reflectometer techniques[1–3] has led to sensitive means for determining the small losses and reflections normally associated with good waveguide joints and connectors.

The purpose of this paper is to describe and discuss these techniques which should provide a powerful tool for the investigation of the properties of waveguide joints or connectors.

Preliminary Considerations

As shown in Fig. 1, a waveguide joint or connector may be represented by a 2-arm waveguide junction characterized by four scattering coefficients, S_{11}, S_{12}, S_{21}, and S_{22}. One seldom needs to know all four in order to predict the behavior of the junction. For example, if



$$\Gamma_A = \Gamma_1 e^{-2\gamma_1 l_1}, \qquad \Gamma_1 = S_{11} + \frac{S_{12}S_{21}}{\frac{1}{\Gamma_L} - S_{22}}, \qquad \Gamma_L = \Gamma_T e^{-2\gamma_2 l_2}$$

Fig. 1—2-arm waveguide junction representation of waveguide joint.

* Received by the PGI, July 18, 1960. Presented at the 1960 Conference on Standards and Electronic Measurements as paper 4-6.

† National Bureau of Standards, Boulder, Colo.

[1] G. F. Engen and R. W. Beatty, "Microwave reflectometer techniques," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-7, pp. 351–355; July, 1959.

[2] J. K. Hunton, "Analysis of microwave measurement techniques by means of signal flow graphs," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-8; pp. 206–212; March, 1960.

[3] R. W. Beatty and D. M. Kerns, "Recently developed (Aug., 1958) microwave impedance standards and methods of measurement," IRE TRANS. ON INSTRUMENTATION, vol. I-7, pp. 319–321; December, 1958.

Reprinted from IRE TRANSACTIONS
ON INSTRUMENTATION
Volume I-9, Number 2, September, 1960

nonreciprocal behavior is excluded, $S_{12}=S_{21}$, and only three are needed. If, in addition, the junction is symmetrical, $S_{11}=S_{22}$, and only two coefficients are required. If the junction is lossless or nearly so, then for practical purposes $|S_{12}|=|S_{21}|$ and $|S_{11}|=|S_{22}|$. It is sufficient for many purposes to determine only the VSWR corresponding to $|S_{11}|$ or $|S_{22}|$ and/or the efficiency η (for energy flowing into arm 1 and out at arm 2,

$$\eta_{21}=\frac{|S_{21}|^2}{1-|S_{11}|^2},$$

and for the reverse direction,

$$\eta_{12}=\frac{|S_{12}|^2}{1-|S_{22}|^2}),$$

of a 2-arm junction terminated in a nonreflecting load. In practice, the direction of energy flow will make little difference in the efficiency of a low-loss connector (even if it is not physically symmetrical), for the reasons mentioned above. Similarly, the VSWR will be essentially independent of the direction of energy flow.

Brief Description of Method

The measurement techniques used to obtain the VSWR and efficiency are illustrated in Figs. 2 and 3, respectively. In both cases, one uses a single directional coupler reflectometer employing two auxiliary tuners, X and Y, which are adjusted in turn in the following way. Tuner X is first adjusted so that the cyclical variations in the sidearm (arm 3) output as one slides a low-reflection termination in waveguide section A (Fig. 2) are essentially eliminated. Then tuner Y is adjusted to achieve the same condition, as one slides a highly reflecting termination in waveguide section A (Fig. 3).

The joint or connector under investigation is between the identical waveguide sections A and B. It is intuitively evident that if the joint or connector were perfect ($S_{11}=S_{22}=0$, and $|S_{12}|=|S_{21}|=1$), there would be no cyclical variations in the sidearm output as one slid either termination from waveguide section A into section B. However, if the joint or connector were not perfect, it seems reasonable to expect that the adjustments made with terminations sliding in section A would not hold when these terminations were transferred to section B

This is actually the case, and the VSWR and efficiency of the joint can be obtained from the observed data in such an experiment. The presence of very small reflections is sensitively determined by the arrangement of Fig. 2 and the presence of very small losses by the arrangement of Fig. 3.

Review of Reflectometer Techniques

In order to interpret the above experiments and obtain quantitative results, it is necessary to briefly review some of the reflectometer theory as described in the literature.[1–3]

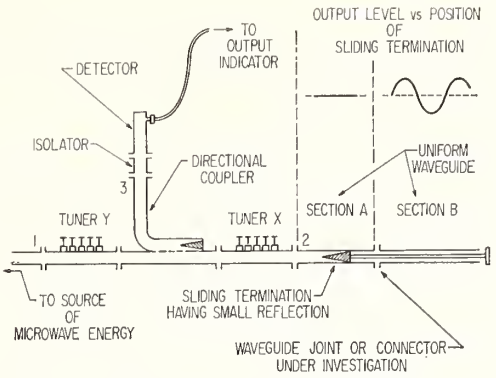


Fig. 2—Reflectometer arrangement for measuring VSWR of waveguide joint.

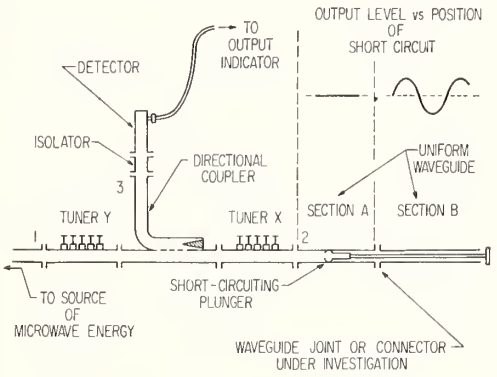


Fig. 3—Reflectometer arrangement for measuring efficiency of waveguide joint.

The amplitude b_3 of the wave emerging from the sidearm (arm 3 in Figs. 2 and 3) is related to the reflection coefficient Γ_A by the equation

$$b_3=kb_G\frac{\frac{1}{K}+\Gamma_A}{1-\Gamma_{2i}\Gamma_A}, \tag{1}$$

where b_G is the component of the incident wave amplitude furnished by the generator, K is a function of the scattering coefficients of the reflectometer, Γ_{2i} and k are also functions of these scattering coefficients and the reflection coefficients of the generator and detector, and finally Γ_A is the reflection coefficient terminating arm 2 of the reflectometer. The terminal plane of arm 2 is located in waveguide section A and may have any arbitrary position sufficiently removed from the joints so as to avoid higher modes. The term Γ_{2i} is the reflection coefficient of the equivalent generator at this reference plane, while K is practically equal to the directivity ratio of the reflectometer.

It is possible to adjust tuners X and Y to make both $1/K$ and Γ_{2i} vanish. Under these conditions, (1) becomes

$$b_3=c\Gamma_A, \tag{2}$$

where $c=kb_G$. This factor will remain constant if the

generator is stable and well isolated, the detector impedance terminating arm 3 is constant, and the tuner adjustments are stable. The detector output power P_D is proportional to $|b_3|^2$ or

$$P_D = p|\Gamma_A|^2. \tag{3}$$

The constant of proportionality can be determined if a reflection standard, such as a quarter wavelength short circuit, is connected to waveguide section A. However, it can be eliminated from consideration by measuring only power ratios in which it cancels out.

VSWR Determination

The cyclical variations obtained as one slides the low-reflection termination in waveguide section B (Fig. 2) occur as the reflection from the termination goes in and out of phase with the reflection from the joint. Assuming that the reflectometer has been adjusted so that (2) applies, the behavior of Γ_A is of interest. As indicated in Fig. 1, it may be written

$$\Gamma_A = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}, \tag{4}$$

where the $S_{m,n}$ are the scattering coefficients of the 2-arm junction representing the waveguide joint and short sections of waveguide on either side of the joint. Eq. (4) may in practice be simplified to

$$\Gamma_A \approx S_{11} + \Gamma_T e^{-2\gamma_2 l_2}, \tag{5}$$

since one uses a termination with small $|\Gamma_T|$ ($|\Gamma_T| < 0.005$) and the loss and reflection of the joint are small. Excluding nonreciprocal behavior of the joint, these considerations lead to the conditions $|S_{22}\Gamma_L| \ll 1$, and $|S_{12}S_{21}| \approx 1$, which are necessary for the above simplification of (4).

An illustration of (5), neglecting attenuation of the waveguide, is given by the diagrams of Fig. 4. The upper diagrams show the circular loci of Γ_A as Γ_T varies in phase, while the lower graphs show the corresponding variations in the output level of the sidearm. In each case, it is seen that

$$|b_3|_{\max} = |c|(|S_{11}| + |\Gamma_T|), \tag{6}$$

where c is the same as in (2).

The determination of $|S_{11}|$ proceeds as follows. Let $|b_3|_{sc} \approx |c|$ represent the output level of arm 3, when a low-loss short circuit terminates waveguide section A, and let $|b_3|_T \approx |c\Gamma_T|$ represent the corresponding output level, when the short-circuit is removed and the sliding termination inserted in waveguide section A. First $|\Gamma_T|$ may be determined by measuring the ratio of $|b_3|_T$ to $|b_3|_{sc}$ with a calibrated variable attenuator, maintaining the detector output at a fixed reference level so as to make unnecessary any knowledge of the detector law. Second, the quantity $|S_{11}| + |\Gamma_T|$ is similarly determined from the ratio of $|b_3|_{\max}$ to $|b_3|_{sc}$. Finally, $|S_{11}|$ is determined by subtraction. There are many refinements and variations of this basic technique, some of which are discussed later.

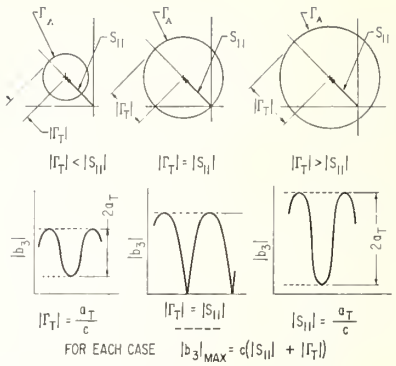


Fig. 4—Reflection coefficient diagrams and detector response curves arising from (5).

Efficiency Determination

In the 2-arm junction representation of a waveguide joint or connector, it is necessary to select terminal surfaces on either side of the joint to avoid appreciable higher modes at these surfaces. The length l_T of waveguide between the two terminal surfaces will introduce some wall loss or attenuation in addition to the loss in the joint. The measured efficiency may be written

$$\eta = \eta_A \eta_J, \tag{7}$$

where $\eta_A = e^{-2\alpha l_T}$, the efficiency of a section of waveguide of length l_T having an attenuation constant α, and η_J is the efficiency of the joint alone. (This separation of the effect of losses is not rigorous, but should hold closely in practice.)

The determination of η follows from previous work[4,5] in which one measures the radius of the circular locus of the input reflection coefficient as one slides a short circuit in the output waveguide. The situation may be represented by Fig. 1, in which Γ_T now denotes the reflection coefficient of the sliding short circuit. In a uniform waveguide section having finite attenuation, the locus of Γ_L is a logarithmic spiral converging toward the origin. However, this is transformed by the 2-arm junction, and the correspondings locus of Γ_1 is, in general, a distorted logarithmic spiral converging toward S_{11}. It is distorted because the spiral $1/\Gamma_L$ is displaced or translated an amount S_{22} before being inverted. The output level $|b_3|$ of the sidearm of the reflectometer in this general case exhibits variations as shown in Fig. 5. Analysis of the data is somewhat complicated in this case and is not within the scope of this paper.

Fortunately, in many cases the reflection coefficient S_{22} is small and has negligible effect in distorting the spiral locus of Γ_A. In this event, one obtains data similar to that shown in Fig. 6. On each side of the transition length l_T, it is appropriate to consider the waveguide

[4] A. L. Cullen, "Measurement of microwave-transmission efficiency," *Wireless Engrg.*, vol. 26, pp. 255–258; August, 1949.

[5] R. W. Beatty, "Determination of attenuation from impedance measurements," Proc. IRE, vol. 38, pp. 895–897; August, 1950.

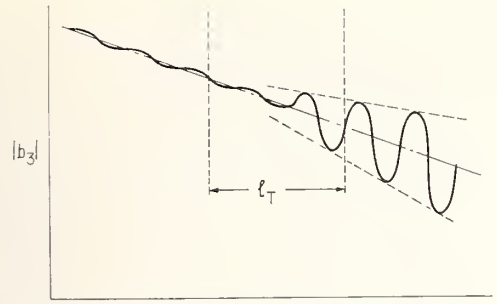


Fig. 5—Detector output vs position of sliding short when reflection from joint is moderately large.

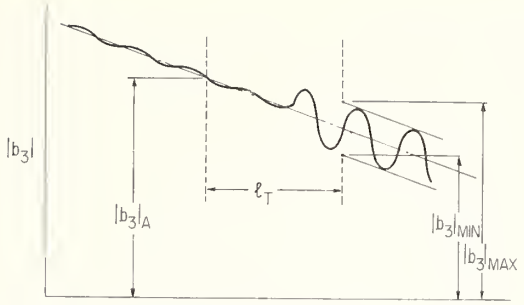


Fig. 6—Detector output vs position of sliding short when reflection from joint is small.

lossless and apply the theory developed under this assumption. This theory is briefly as follows.

It can be shown that the radius of the Γ_A-circle as the phase of Γ_T varies is

$$R_{21} = \frac{|S_{12}S_{21}\Gamma_T|}{1 - |S_{22}\Gamma_T|^2}. \tag{8}$$

The efficiency of a 2-arm junction with energy flowing into arm 2 and with arm 1 terminated in a nonreflecting load is

$$\eta_{12} = \frac{|S_{12}|^2}{1 - |S_{22}|^2}. \tag{9}$$

It is evident that to the extent that $1-|S_{22}|^2$ equals $1-|S_{22}\Gamma_T|^2$, and $|S_{12}S_{21}| = |S_{12}|^2$, $R_{21} \approx \eta_{12}|\Gamma_T|$. This approximation will be quite good if a low-loss short circuit is used. If the connector has low loss, it will very nearly be true that $|S_{12}| = |S_{21}|$, even if nonreciprocal behavior were permitted.

Referring to Fig. 7, it is apparent that

$$\frac{1}{2}\,\frac{|b_3|_{\max} + |b_3|_{\min}}{|b_3|_A} = \frac{\eta|\Gamma_{sc}|}{|\Gamma_{sc}|} = \eta, \tag{10}$$

since $\eta|\Gamma_{sc}|$ is the radius of the Γ_A-circle and the constant $|c|$ cancels. It is noted that $|\Gamma_T|$ has been replaced by $|\Gamma_{sc}|$.

It is convenient to use instrumentation developed[6–8] to measure power differences directly, when observing the changes in the sidearm output as one slides the short-circuit. In order to obtain data as shown in Fig. 6, it would be necessary to plot the square root of the observed relative power. However, it is somewhat easier to analyze the data as obtained and take square roots of only the three points needed in the calculation of efficiency.

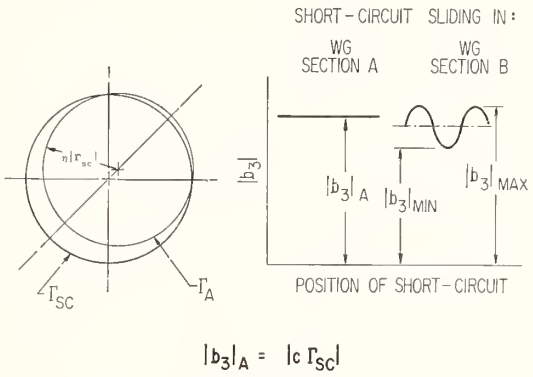


Fig. 7—Reflection coefficient diagrams and detector response curves for a sliding short in lossless waveguide.

The attenuation of the waveguide sections is obtained from the slope of the curve to the left of l_T in Fig. 6. If P_1 and P_2 are the side arm powers corresponding to two positions of the short circuit spaced a distance l, the attenuation constant α may be calculated from the expression

$$e^{-2\alpha l} = \frac{P_2}{P_1}, \tag{11}$$

where P_2 is smaller than P_1.

Supplemental Techniques in VSWR Determination

In the technique described for measuring the VSWR corresponding to $|S_{11}|$ of the 2-arm junction representing the waveguide joint or connector, it was assumed that $|1/K|$ was much smaller than $|S_{11}|$. When dealing with very small reflections, the reduction of $|1/K|$ to even smaller values may become difficult, and a number of supplemental techniques have been developed for the solution to this problem.

The behavior of the detector output as one slides a termination having a small reflection ($|\Gamma_T| \ll 1$) in waveguide section B may be described by

$$b_3 \approx kb_G\left(\frac{1}{K} + S_{11} + |\Gamma_T|\,e^{j\psi}\right). \tag{12}$$

In general, the previous technique for obtaining $|S_{11}|$ will yield $|(1/K)+S_{11}|$ instead. If $|1/K| \ll |S_{11}|$, there

[6] G. F. Engen, "Amplitude stabilization of a microwave signal source," IRE Trans. on Microwave Theory and Techniques, vol. MTT-6, pp. 202–206; April, 1958.

[7] G. F. Engen and R. W. Beatty, "Microwave attenuation measurements with accuracies from 0.0001 to 0.06 decibel over a range of 0.01 to 50 decibels," *J. Res. NBS*, vol. 64C, pp. 134–135; April–June, 1960.

[8] G. F. Engen, "A self balancing direct-current bridge for accurate bolometric power measurements," *J. Res. NBS*, vol. 59, pp. 101–105; August, 1957.

is no difficulty, but if this is not true, we cannot determine $|S_{11}|$ even if we knew $|1/K|$, because the phase difference between the two terms is unknown. It then becomes important either to reduce $|1/K|$ until it is much less than $|S_{11}|$ and to know when this is the case, or to employ a technique in which the relative phase of $1/K$ and S_{11} can be varied so as to separate their magnitudes.

In adjusting tuner X to reduce $|1/K|$ to a small value, it is necessary to avoid the possible false adjustment which could result in approximately constant output if $|1/K| \gg |\Gamma_T|$. One way to do this is to begin the adjustment of tuner X with all stubs out of the waveguide, and then adjust for a detector null with the sliding termination in an arbitrary fixed position in waveguide section A. Under this condition, $1/K = -\Gamma_T$, and if the termination is then displaced for maximum detector output, the level will be $|b_3| = 2|c\Gamma_T|$. The adjustment of tuner X is then continued until an essentially constant output level is obtained as one slides the termination, for which the detector level should be approximately half the above level, or 6 db down.

One should know when to stop trying to improve the adjustment of tuner X so as to avoid needless tedium. This may be done with the aid of Fig. 8, as illustrated in the following example. Suppose that $|S_{11}|$ approximately equals 0.00025 and that it is considered sufficient to reduce $|1/K|$ to one tenth of this value or 0.000025. (The corresponding error in determining $|S_{11}|$ by the previously described technique would then be less than 10 per cent.) If the sliding termination has a VSWR of 1.003, the graph shows that the adjustment of tuner X can cease when the total variation of the detector output is within 0.3 db. This example is representative of what can be done with commercially available components at the present time.

The supplemental techniques to be described are included as alternate ways to reduce $|1/K|$ or to prevent error in measuring $|S_{11}|$ because of finite $|1/K|$.

1) The method which is potentially the most powerful in reducing $|1/K|$ is similar to that described above, except that in place of a sliding termination of constant $|\Gamma_T|$, an adjustable sliding termination[9] is used. One alternately adjusts both tuner X and the termination, so that the sidearm output is reduced to a lower and lower constant value as the termination is slid. By doing this, both $|1/K|$ and $|\Gamma_T|$ are reduced together. It is often found that the adjustment is limited, not by the sensitivity of the detector to respond to the small reflected signal, but by the ultimate failure of $|\Gamma_T|$ to remain constant as the termination is slid. However, a return loss of 100 db has been obtained with suitable adjustable sliding terminations. This corresponds to $|1/K|$ certainly less than 0.00001. It could be considerably less, depending upon the observed variation in detector output.

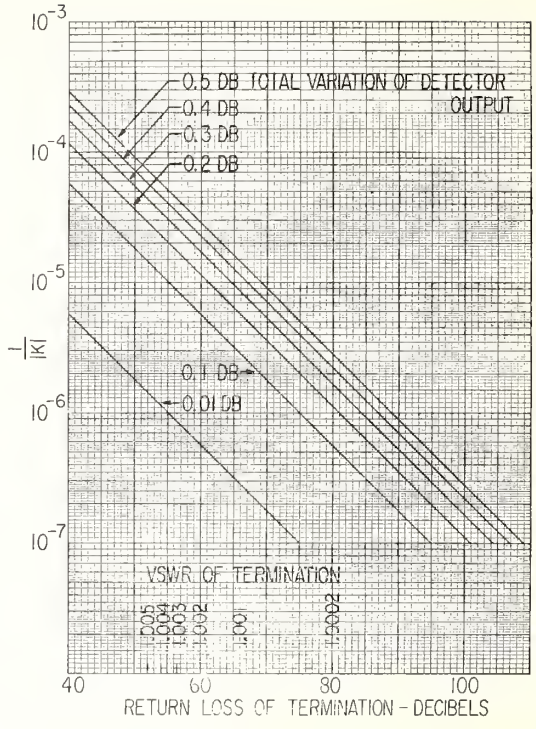


Fig. 8—Graph for determining $|1/K|$ from variation in detector output as a low-reflection termination is slid in the waveguide section A.

Apart from the reduction of $|1/K|$, an adjustable sliding termination may be used in the measurement of $|S_{11}|$ as follows. Referring to Fig. 4, it is apparent that the detector output will vanish if $\Gamma_T = -S_{11}$. This condition can be easily achieved and recognized if an adjustable sliding termination is used. Subsequently, without changing the adjustment, the termination is slid until the detector output is a maximum and proportional to the sum of $|S_{11}|$ and $|\Gamma_T|$ or $2|S_{11}|$. Comparing this output level to that obtained when the termination is removed and replaced by a short circuit will eliminate the constant of proportionality and permit determination of $|S_{11}|$.

2) An auxiliary waveguide channel[10] arranged as in Fig. 9 permits introduction of a signal to the detector of such a phase and magnitude so as to cancel the signal component due to finite $1/K$. As is indicated by the vector diagrams, the procedure begins with the adjustment of tuner X for a detector null, whereupon $1/K = -\Gamma_T$. The sliding termination is then moved to a position where the detector output is maximum, changing the phase of Γ_T by 180°, so that the signal components from Γ_T and $1/K$ add. The switch in the auxiliary arm is now opened, introducing a signal component "A" to the detector. The amplitude and phase of this component are adjusted using the phase shifter

[9] R. W. Beatty, "An adjustable sliding termination for rectangular waveguide," IRE Trans. on Microwave Theory and Techniques, vol. MTT-5, pp. 192–194; July, 1957.

[10] The circuit is quite similar to one used in the determination of barretter mount efficiencies by an impedance technique. A paper by G. F. Engen describing the use of this circuit in that technique is in preparation.

and variable attenuator to null the detector output. It is apparent from the diagram that "A" is in phase opposition to the $1/K$ signal component and will cancel it if reduced in amplitude by one half without changing its phase. This is done by adding 6.02-db attenuation in the auxiliary arm. The measurement of $|S_{11}|$ can now proceed in the manner discussed previously.

Additional tuning (not shown) could be employed to prevent possible interaction between the two channels; however, the isolator shown in the figure should prove adequate for this purpose. The resulting adjustment can be checked by means of the procedure associated with Fig. 8 and it is possible that $|1/K|$ will not be small enough. This could easily be the case if the attenuator did not accurately produce the 6.02-db change required, or produced some phase shift. Further adjustment of either tuner X or the phase shifter and attenuator would then be necessary to reduce $|1/K|$ to the desired value. In case of difficulties in obtaining fine adjustments of the tuner, the latter procedure is quite convenient and can provide good resolution.

3) An extension of the previous technique in which a second auxiliary channel is employed permits cancellation of the $1/K$ signal component without the use of a calibrated attenuator. The arrangement is similar to that of Fig. 9, except for the addition of another similar channel. Vectors representing detector signal components from $1/K$, Γ_T and the two channels A and B are shown in Fig. 10, corresponding to the steps in the procedure mentioned in the figure. Although a calibrated attenuator is not required, it is necessary to vary the phase of Γ_T by prescribed amounts, so that a micrometer drive for the sliding termination is convenient.

4) The relative phase of $1/K$ and S_{11} may be varied by means of the line stretcher arrangement shown in Fig. 11. Alternate manipulation of the line stretcher and the sliding termination to obtain maximum detector output results in alignment of the vectors representing $1/K$, S_{11}, and Γ_T, as shown in Fig. 12. The remaining steps in the procedure and the corresponding vector diagrams and pertinent equations are shown in the figure. It is apparent that the ratio of $|S_{11}|$ to $|\Gamma_T|$ equals the ratio of $|b_3|_{min}$ to $|b_3|_{max}$. If $|\Gamma_T|$ is known or determined independently, then $|S_{11}|$ may be calculated. In order to deal with conveniently measured ratios, one should use a termination having a $|\Gamma_T|$ not greatly different from $|S_{11}|$.

5) Another technique employing the same arrangement as in Fig. 11 is as follows. Tuner X is adjusted for minimum variations in detector output as a low-reflection termination slides in the waveguide section in which terminal surface 2 is located. This reduces $|1/K|$ to a small value; then, the three vectors $1/K$, S_{11}, and Γ_T are lined up as described in the previous technique. One adjusts tuner Y until the detector output variations

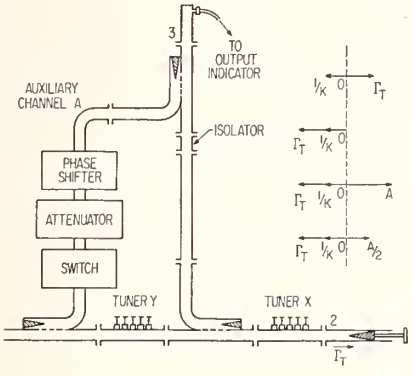


Fig. 9—Auxiliary channel technique.

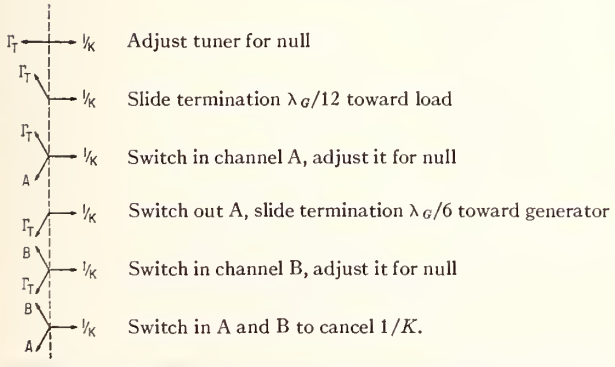


Fig. 10—Steps in a technique employing two auxiliary channels.

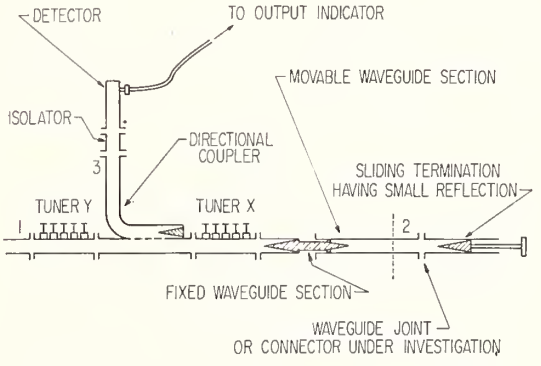


Fig. 11—Arrangement for varying relative phase of $1/K$, S_{11}, and Γ_T.

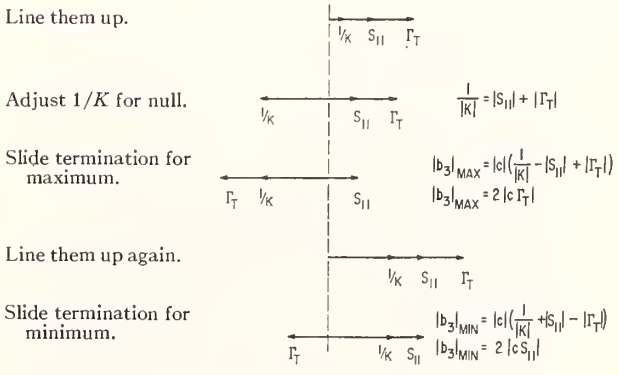


Fig. 12—Steps in a technique for determining $|S_{11}|$ in terms of $|\Gamma_T|$.

are minimized as a short circuit is slid in the waveguide section. The detector output level $|b_3|$ will now be proportional to the sum of $|1/K|$, $|S_{11}|$, and $|\Gamma_T|$. The constant of proportionality may be eliminated, in the usual way, by taking the ratio of this output to that obtained when the waveguide section is terminated in a high-quality short circuit. One then determines $|\Gamma_T|$ and $|1/K|$ independently by methods previously described and finally calculates $|S_{11}|$.

Results

The techniques described above are applicable in principle to waveguide systems employing waveguide of rectangular, coaxial, or any kind of cross section. However, experimental results have been obtained in a WR-90 (X-band) rectangular waveguide system operating at a frequency of approximately 9.39 Gc.

The effects of lateral displacement upon the reflection and efficiency of a plane butt joint in rectangular waveguide were investigated; a few measurements were performed on other types of joints, and the attenuation constants of some short sections of waveguide were determined.

Fig. 13 is a photograph of the arrangement used to obtain prescribed repeatable lateral displacements of the waveguide at a simple butt joint. Clamps were used for alignment and to insure repeatability, and strips of shim stock of various thicknesses were used to obtain the prescribed displacements. The heavy brass flanges shown were originally one piece. It was soldered in the center of a section of uniform waveguide and the edges were machined flat and square; then it was cut so as to separate into two sections of waveguide, each with its own flange. The mating surfaces were then ground so as to be flat and square.

The initial alignment was checked by visual inspection. A flashlight was used to illuminate the interior of the waveguide, and reflections of light from any visible edges at the joint proved to be a sensitive indication of misalignment or mechanical imperfections, such as burred edges. In spite of the care taken to obtain good alignment, the reflection coefficient of the joint was never below 0.00015, and in the results shown, was approximately 0.00071.

A comparison of experimental and calculated results is shown in Fig. 14. The calculated curve is based upon the equation shown, which differs from that given in the reference[11] by a factor of two, but agrees (when corrected for the different ratio of f to f_c) with the appropriate curve given in their Fig. 5. The agreement is quite good over a limited range, but it is apparent that the residual reflection that one obtains at zero displacement prevents agreement at the low end. It is probable that the approximations made in deriving the equation contribute to the disagreement at the other end.

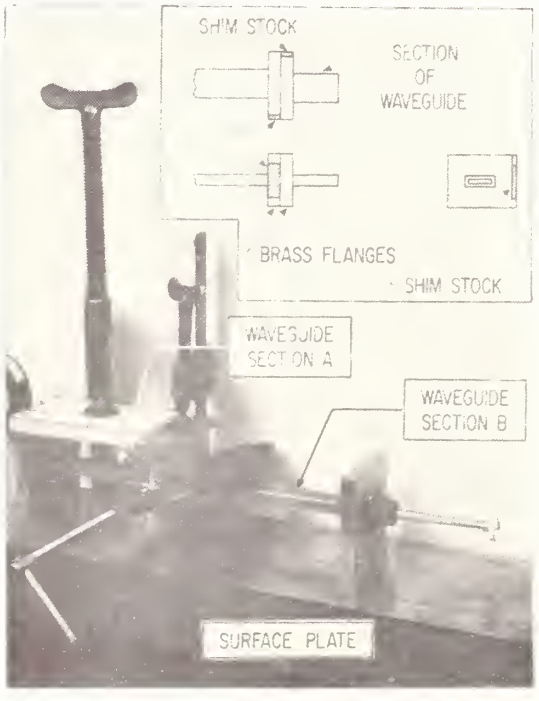


Fig. 13—Photograph of alignment and clamping apparatus for waveguide joint.

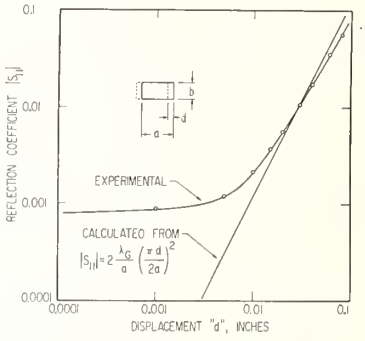


Fig. 14—Measured and calculated reflections from junction of displaced waveguide sections.

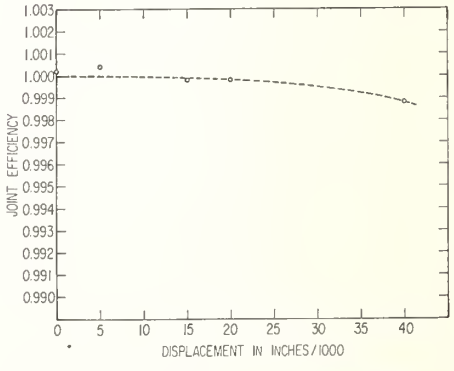


Fig. 15—Measured efficiency of junction of displaced waveguide sections.

[11] U. V. Kienlin and A. Kürzl, "Reflexionen an Hohlleiter-Flanschverbindungen," *Nachr. tech. Z.*, heft. 11; 1958.

A plot of observed data obtained in an efficiency measurement is shown in Fig. 6. The attenuation constant of the brass waveguide as determined from the slope of the curve is 0.056 db per foot. Measurements were made of the efficiency of the same butt joint described above for the same lateral displacements and the results are shown in Fig. 15. The results are not necessarily representative of this type of joint, since the loss would be expected to depend on the surface finish and cleanliness of the metal at the contacting surfaces, as well as other factors. However, it is an indication of what can be obtained.

The results of additional measurements on commonly used types of joints are given in Table I. Again, the results are probably not representative, but show what can be obtained with careful alignment of the waveguides and cleaning of the joint surfaces. It should be noted that the waveguide sections united by a joint were originally a single section of waveguide which was sawed in half. Thus, there is very little, if any, change in the waveguide cross section at any of the joints.

TABLE I

Reflections and Losses of Some Joints, in WR-90 (*X*-band) Rectangular Waveguide Measured at 9.39 Gc

Type of Flanges Used with Butt Joint	Reflection Coefficient $\lvert S_{11} \rvert$	Efficiency
Choke-Cover	0.00064	0.9996
Cover-Cover	0.0015	0.9993

Acknowledgment

The following contributions were vital to the work described. C. Hyder contributed to the design of the arrangement for displacing the waveguide by known amounts and fabricated special apparatus. R. Williams and R. Ridge assisted in the measurements and calculations.

Measuring Impedance Through an Adapter Without Introducing Additional Error

In making impedance measurements one may be faced with the problem of nonmating connectors on the one-port device whose impedance is to be measured and the measuring instrument. It is then customary to use an adapter. There may also be a transition section such as between a slotted coaxial line having convenient internal diameters and its output connector, which may have smaller internal diameters.

One can either evaluate the adapter and transition section and correct for the effect [1], or tune them out with a lossless tuner [2]. In the first case, additional errors are introduced because of errors in measuring the characteristics of the adapter and transition section. In the second case, some error is introduced by limited resolution of the apparatus used to recognize when the correct tuning adjustment has been made. In addition to this error, there is no way to evaluate the additional error that results from violation of the assumption of losslessness.

The purpose of this correspondence is to call attention to the capability of the tuned reflectometer in avoiding errors such as just described. In the process of adjusting the tuners of the reflectometer in the usual way, the reflections introduced by the addition of the adapter will be tuned out. In addition, this tuning process will compensate for the effect of dissipative losses in the adapter. This capability will be illustrated by an example in which the main part of the tuned reflectometer is composed of rectangular waveguide components, but the one-port device to be measured has a type-N coaxial connector. Such an arrangement has been used [3] in order to take advantage of the superior properties of rectangular waveguide components. The reverse of the arrangement is also useful in taking advantage of the broadband capabilities of coaxial components. This hybrid concept is also useful if access to the one-port device is by means of unconventional waveguide such as ridged waveguide, strip line, or reduced-height waveguide.

In the foregoing example, a waveguide to coaxial adapter is connected to the output end of the rectangular waveguide directional coupler as shown in Fig. 1. A uniform section of coaxial line is attached to the adapter, and provision is made to slide loads inside of this line. In making the tuning adjustments of the reflectometer [4], one must slide each load a distance of at least half wavelength. In this example, one could not introduce the loads in the waveguide past the connector because of the bead support, so provision is made to break the outer conductor. One can then insert the sliding loads, tune the reflectometer, and then replace the portion of the outer conductor containing the bead. Any discontinuity from this joint in the outer conductor can cause error, but one can reduce this error to negligible amounts by careful construction of the joint.

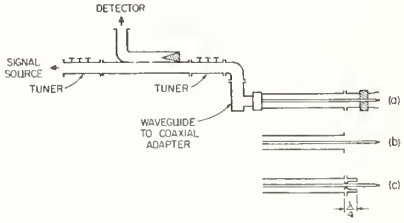


Fig. 1. Terminating arrangements for a tuned reflectometer mainly composed of rectangular waveguide components, but having a coaxial output. (a) Outer conductor in place. (b) Outer conductor removed so that sliding loads may be inserted. (c) Quarter-wavelength short-circuit reference standard in place.

The tuning procedure is exactly the same as if the adapter were not used, and the measurement procedure is the same. No assumption need be made that the adapter is lossless, and one does not need to separately tune it or evaluate it since it is considered as part of the reflectometer

The terminal surface at which the impedance is measured is a plane surface perpendicular to the axis of the coaxial line and its location is determined by the short-circuit impedance standard used. It is located at the joint in the outer conductor as shown in Fig. 1. One can then refer this impedance to other planes by well-known techniques. It should be noted that the effects of the reflections from the outer connector pair are attributed to the one-port device under test, even though a portion of this connector pair belongs to the reflectometer. If phase as well as amplitude measurements are made, no phase error is introduced by the adapter, even though the standard phase shifter may be in rectangular waveguide, and the one-port device has a coaxial connector.

The foregoing example represents a particularly awkward case, and less difficulty would be experienced if the one-port device under test had reduced-height rectangular waveguide rather than coaxial line.

In comparing this procedure to evaluating an adapter by usual techn ques, one sees that, in addition to providing a sliding short circuit, one also needs to provide a sliding load having low reflection. The difficulty of doing this is balanced against the necessity for making additional measurements in evaluating an adapter and the additional errors that would be introduced.

In single frequency, high-precision measurements, the technique described in the foregoing would seem preferable, whereas evaluation of an adapter might be worthwhile if swept-frequency measurements over a broadband are required.

R. W. Beatty
National Bureau of Standards
Boulder, Colo.

References

[1] Ginzton, E. L., *Microwave Measurements*, New York: McGraw-Hill, 1957, sec 5.8, p 275.
[2] Mathis, H. F., Measurement of reflection coefficients through a lossless network, *IRE Trans. on Microwave Theory and Techniques (Correspondence)*, vol MTT-3, Oct 1955, p 58.
[3] Beatty, R. W., and W. J. Anson, Applications of reflectometer techniques to accurate reflection measurements in coaxial systems, *Proc. IEE (London)*, vol 109, pt B, Jul 1962, p 345.
[4] Anson, W. J., A guide to the use of the modified reflectometer technique of VSWR measurement, *J. Res. NBS*, vol 65C, Oct-Dec 1961, p 217.

Manuscript received April 28, 1965.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 53, NO. 6, JUNE, 1965
pp. 656-657

An Automatic Method for Obtaining Data in the Weissfloch-Feenberg Node-Shift Technique

The purpose of this correspondence is to describe a method which has been used to give automatic readout of the node shift curve obtained in the Weissfloch-Feenberg [1], [2] technique. The method was developed in 1951, but was not published, since at that time, it was felt to be of rather limited interest. However the recent increase in activity in the measurements field has increased the importance of automatic techniques. The well-known [3] node-shift technique is used in the evaluation of properties of 2-ports such as those representing discontinuities in coaxial lines, for example.

In the application of this technique, one locates a sliding short-circuit in the line or waveguide on one side of the 2-port under test, and connects a slotted line to the other side. The data to be obtained consists of the displacements d of a node in the standing wave pattern in the slotted line, corresponding to given displacements s of the short circuit. Plots of d vs s or $(d-s)$ vs s show a cyclical variation repeating every half wavelength. The amplitude of this cyclical variation is a measure of the amplitude of the reflection from the discontinuity in the section of line represented by the 2-port.

The node shift curve is ordinarily obtained by point-by-point measurements. If many measurements are to be taken, the semiautomatic method of Tischer [4], [5] may be used, in which the slotted line probe is mechanically coupled to the drive mechanism used to displace the sliding short-circuit. In Tischer's method, provision is made for small adjustments of the length of the coupling rod, so that the probe can be set on the nodal position at any time. A dial gauge is used to read these small changes, and the node shift curve is then obtained by a point-by-point semiautomatic method. A further refinement used by Tischer is a vibrating probe which can be set to sweep across the node in the standing-wave pattern. The probe output displayed on an oscilloscope gives a picture of the node. Synchronization of the scope sweep with the probe vibration permits one to measure the node shift on the screen of the oscilloscope, which may be calibrated in terms of distance. A further use of a mechanically coupled sliding probe and load, in a different measuring technique has been recently [6] reported.

In the automatic method, the probe is mechanically coupled to the sliding short-circuit, as shown in Fig. 1, and the probe is purposely displaced a small amount from the nodal position. Any further relative displacement between the probe and the nodal position, such as that due to the node shift, will cause a change in the probe output. If the recorder drive is mechanically coupled to the short circuit and probe drive, a node-shift curve such as that shown in Fig 2 is obtained automatically.

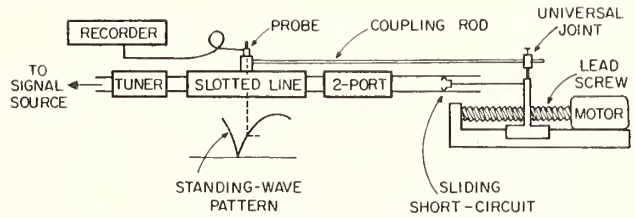


Fig. 1—Simplified arrangement of apparatus for automatic readout of data.

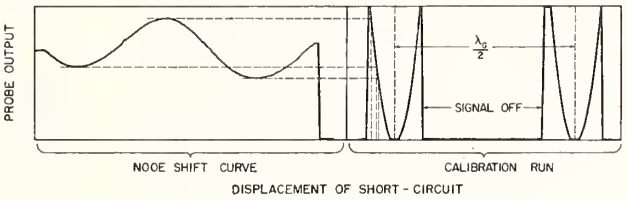


Fig. 2—Sample data and calibration curve obtained automatically at 2970 MHz.

It remains to calibrate the probe's output in terms of its relative displacement from the node. This is conveniently done with automatic readout by uncoupling the probe from the short circuit and keeping the probe stationary while the short circuit is driven. The calibration curve, as shown in Fig. 2, can then be used to obtain the amplitude of the node shift curve.

Since the probe output will also vary if the incident wave amplitude in the slotted line varies, the signal source must be 1) stable, 2) well padded or isolated to prevent pulling as the short circuit is slid, (30 db of padding or 60 db of isolation is usually sufficient), and 3) nonreflecting.

Only the third requirement is peculiar to the automatic method, but it is not difficult to satisfy. One can place a tuner between the signal source and the slotted line and adjust it for this condition. The nonreflecting condition may be recognized in various ways, such as elimination of cyclical variations in probe output as the short circuit and probe are driven, but with effectively no discontinuity present between the slotted line and the short circuit.* The effect of a discontinuity can be eliminated [7] by an auxiliary slide-screw tuner installed near the load end of the slotted line.

It is immediately apparent that many modifications of this basic method are possible to fit individual requirements regarding available instrumentation and convenience. For example, the recorder and the short circuit drive need not be mechanically coupled if they are both driven by synchronous motors. On the other hand, a linear potentiometer may be used to convert displacement to an electrical current so that an $X-Y$ recorder may be used.

The speed of obtaining data by this method makes possible more extensive and thorough investigations within a given time. This will justify, in many cases, the additional complexity of the instrumentation required and the calibration and tuning procedures which are essential to good accuracy.

Thanks are given to A. Roa, who took the data shown in Fig. 2 while at NBS in December, 1951.

ROBERT W. BEATTY
Engrg. Division
Radio Standards Lab.
National Bureau of Standards
Boulder, Colo.

REFERENCES

[1] A. Weissfloch, "Ein Transformationsatz verlustloser Vierpole und seine Anwendung auf die experimentelle Untersuchung von Dezimeter- und -Zentimeterwellen-Shaltungen," *Hochfrequenz. und Elektroak.*, vol. 60, pp. 67–73; September, 1942.
[2] E. Feenberg, "The relation between nodal positions and standing wave ratio in a composite transmission system," *J. Appl. Phys.*, vol. 17, pp. 530–532; June, 1946, see also Sperry Gyroscope Rept. 5220–120; Garden City, N. Y., June 5, 1943, of same title.
[3] E. L. Ginzton, "Microwave Measurements," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 278–285; 1957.
[4] F. J. Tischer, "Mikrowellen-Messtechnik," Springer-Verlag, Berlin, Germany, pp. 158–160; 1958.
[5] F. R. Huber and H. Neubauer, "Measurement techniques for the determination of the major characteristics of coaxial components," *Microwave J.*, vol. 5, pp. 196–203; September, 1962.
[6] B. Weinschel, G. Sorger, S. Raff, and J. Ebert, "Precision Coaxial Impedance Measurements By Coupled Sliding Load Technique," presented at 1964 Conf. on Precision Electromagnetic Measurements NBS, Boulder, Colorado; June, 1964; to be published in the December, 1964 issue of the IEEE TRANS. ON INSTRUMENTATION AND MEASUREMENT.
[7] A. E. Sanderson, "A new high-precision method for the measurement of the VSWR of coaxial connectors," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-9, pp. 524–528; November, 1961.

Reprinted from the
PROCEEDINGS OF THE IEEE
VOL. 53, NO. 1, JANUARY, 1965
Pp. 79-80

* If the short-circuit slides inside the slotted line, and the mechanically coupled probe is set at a voltage maximum position, one can adjust the equivalent generator (as seen at some reference plane in the slotted section) to be nonreflecting. However, if the short-circuit slides in a line or waveguide section connected to the slotted line, the adjustment will be in error by the amount of reflection from the discontinuity.

7.a. **The measurement of arbitrary linear microwave two-ports,** H. M. Altschuler, Proc. IEE 109, Part B, Suppl. 23, 704-712 (May 1962).

The measurement of the scattering parameters of linear but otherwise arbitrary two-ports (active, lossless or dissipative, and reciprocal or non-reciprocal) by means of an interference bridge is described. Since the measured data form circular loci from which the parameters are then derived, known precision curve-fitting techniques are applicable. The various methods available for distinguishing the proper locus from its inverse are discussed. Various methods of balancing the bridge and many of the errors associated with bridge balancing are given. Finally, the deliberate use of bridge imbalance for increasing measurement accuracy is considered.

7.b. **Application of reflectometer techniques to accurate reflection measurements in coaxial systems,** R. W. Beatty and W. J. Anson, Proc. IEE 109, Part B, No. 46, 345-348 (July 1962).

The theory of a single directional-coupler reflectometer having two auxiliary tuners is reviewed, giving emphasis to precision-measurement possibilities.

Problems in setting up this system for measurements in coaxial line are discussed, and a system for measurements in the frequency range 3.95-5.85 Gc/s is described.

The development of $\lambda/4$ short-circuit reflection standards and sliding terminations is presented.

The results of measurements of the reflection from connectors and various loads at 4 Gc/s are given.

*7.c. **Measuring the directivity of a directional coupler using a sliding short-circuit and an adjustable sliding termination,** R. W. Beatty, IEEE Trans. Microwave Theory & Tech. MTT-12, No. 3 (May 1964).

7.d. **Microwave impedance measurements and standards,** Robert W. Beatty, NBS Monograph 82, (August 1965).

A survey and discussion of well-known microwave impedance measurement techniques is presented. The discussion includes an introduction which emphasizes basic concepts and reflection coefficient-VSWR relationships. Sources of error in the various measurement techniques are discussed and methods to reduce errors are presented. The discussion of errors in slotted line and reflectometer techniques is most thorough. Methods using rotating loops and resonance lines are included and a brief discussion of microwave impedance standards is given.

7.e. **Microwave standards and measurements in the U.S.A., 1963-1966,** R. W. Beatty, *Progress in Radio Science 1963-1966*, Part I, Proc. XVth General Assembly of URSI (September 1966).

*7.f. **Second-harmonic effects in tuned reflectometers,** M. Michael Brady, IRE Trans. Microwave Theory & Tech. MTT-10, No. 1 (January 1962).

7.g. **Inductance and characteristic impedance of a strip-transmission line,** R. L. Brooke, C. A. Hoer, and C. H. Love, J. Res. NBS 71C (Eng. & Instr.), No. 1, 59-67 (Jan.-Mar. 1967).

A general method is developed for determining the inductance and characteristic impedance of uniform transmission lines. A non-uniform current distribution is allowed in the transverse plane. The system is represented by a matrix equation which can be programmed for computer solution. The correct inductance and impedance are obtained as the result of a simple limiting process. The method is applied to one particular geometry, a four-tape stripline system. Results are given for the inductance, resistance, and current distribution as functions of frequency and resistivity for a particular geometry. A method for extending the results to strip lines with proportional dimensions is developed. An accuracy of one part in 10^5 was found to be feasible for the determination of the inductance per unit length.

7.h. **Current distribution and impedance of lossless conductor systems,** R. L. Brooke and J. E. Cruz, IEEE Trans. Microwave Theory & Tech. MTT-15, No. 6, 358-364 (June 1967).

A general method for determining the characteristic impedance of uniform, lossless transmission systems is developed. The current distribution within the system is determined by means of a matrix equation programmed for computer solution. Once the current distribution is known, the inductance per unit length and characteristic impedance are determined. The results obtained by applying this method to several rectangular coaxial systems are compared with the predictions of an approximate analytic expression. The reflection coefficient of a variable characteristic impedance coaxial line is measured on a time-domain reflectometer (TDR), and the results are compared with both the matrix method and the approximate analytic expression.

*7.j. **A variable characteristic impedance coaxial line,** J. E. Cruz and R. L. Brooke, IEEE Trans. Microwave Theory & Tech. MTT-13, No. 4, 477-478 (July 1965).

The development of Time Domain Reflectometry (TDR) for UHF and microwave impedance measurements in coaxial systems has created a problem of calibrating the TDR system for accurate measurements of small reflections. A need for coaxial impedance standards has been resolved by the development of a variable impedance device which can be calibrated by the use of fixed coaxial standards.

This correspondence deals with the design and analysis of a variable impedance line. This line is described and its performance characteristics are discussed. Its mea-

* Private communication.

sured characteristic impedance is compared at discrete points with the impedance obtained from empirical and approximate theoretical formulas.

7.k. **Exact inductance equations for rectangular conductors with applications to more complicated geometries,** Cletus Hoer and Carl Love, J. Res. NBS 69C (Eng. & Instr.), No. 2, 127-137 (Apr.-June 1965).

Exact equations are given for the calculation of the self-inductance of rectangular conductors and of the mutual inductance between combinations of parallel filaments, thin tapes and rectangular conductors. A general procedure is also given for calculating the self-inductance of complicated geometries by dividing the geometry into simple elements whose inductances can be calculated. This general procedure is valid for conductors having nonuniform, as well as uniform current densities.

7.l. **A self-calibrating instrument for measuring conductance at radio frequencies,** Leslie E. Huntley, J. Res. NBS 69C (Eng. & Instr.), No. 2, 115-126 (Apr.-June 1965).

Accuracies obtainable in measuring resistance (conductance) at radio frequencies have lagged behind those obtainable in measuring reactance (susceptance), because frequency-dependent changes in the values of resistors are not as well known as are such changes in capacitors, and because no satisfactory method has existed for comparing resistors to capacitors at radio frequencies. An instrument based on the twin T null circuit can be made self-calibrating at a given frequency, allowing conductances in the proper range to be measured directly in terms of a change in capacitance, without any need for accurate knowledge of the values of the circuit elements. Analysis of a practical circuit shows that such an instrument is potentially capable of measuring conductance to within a few parts in 10^6 at frequencies below about 15 MHz, and that accuracies actually obtained will be limited only by the accuracy to which changes in effective capacitance are known. Conductances between 0.01 and 0.0001 mho were measured at 10^7 radians/sec with an estimated error of 0.05 percent, which is the error associated with the measurement of changes in effective capacitance at that frequency.

7.m. **Standards for the calibration of Q-meters 50 kHz to 45 MHz,** R. N. Jones, J. Res. NBS 68C (Eng. & Instr.), No. 4, 243-248 (Oct.-Dec. 1964).

The National Bureau of Standards is now equipped to provide improved calibration services for Q-standards in the frequency range extending from 50 kHz to 45 MHz. As a result of recent development work, calibration uncertainties have been reduced to magnitudes which are comparable to the resolution of Q-meters in both resonating capacitance and Q. The uncertainties in the values of the NBS standards are consistent with the best two-terminal impedance measurements currently obtainable, but beyond this the values of the NBS standards have been statistically adjusted to provide a higher degree of standardization. Included in the paper is a discussion of the differences between the Q-meter indicated values for a standard and the effective values of the standard as given by NBS. These differences are largely due to residual immittances in the Q-meter circuit and methods for evaluating these residuals are presented.

7.n. **Precision coaxial connectors in lumped parameter immittance measurement,** R. N. Jones and L. E. Huntley, IEEE Trans. Instr. & Meas. IM-15 No. 4, 375-380 (December 1966).

Amid the concern surrounding the design and application of precision coaxial connectors for use in precise electrical measurements there is one area which has received very little attention. This has to do with the application of such connectors to measurements in the lumped parameter region. This paper points out significant sources of error in lumped parameter measurements which are attributable to connector uncertainties and shows how they can be greatly reduced through the application of precision coaxial connectors. Where these connectors are utilized in lumped parameter immittance measurements, extensions of frequency, magnitude, and accuracy are realized. Specific examples are discussed.

7.o. **Perturbation theorems for waveguide junctions, with applications,** D. M. Kerns and W. T. Grandy, Jr., IEEE Trans. Microwave Theory & Tech. MTT-14, No. 2, 85-92 (February 1966).

Perturbation theorems are derived in the context of a theory of waveguide junctions. These theorems express changes in impedance or admittance matrix elements, due to changes in a waveguide junction, in terms of integrals over products of perturbed and unperturbed basis fields associated with the junction and with its adjoint. Media involved are required only to be linear. Concepts of first-order perturbation theory are discussed briefly, and the term "correct to the lowest order" is precisely defined. The need of explicit theorems telling when one may expect results actually correct to the lowest order is noted.

Two problems are solved approximately by the perturbation approach: (1) reflection at the junction of rectangular waveguide with filleted waveguide of the same main dimensions; and (2) the effect of finite conductivity of both obstacle and waveguide wall for half-round inductive obstacles in rectangular waveguide.

*7.p. **A coaxial adjustable sliding termination,** W. E. Little and J. P. Wakefield, IEEE Trans. Microwave Theory & Tech. MTT-12, No. 2 (March 1964).

See also 1.b, 1.e, 3.a, 6.a, 10.d.

* Private communication.

8. Radio Frequency Materials

Papers

Abstracts

Measurement of RF Properties of Materials A Survey

H. E. BUSSEY, SENIOR MEMBER, IEEE

Abstract—**Methods for radio and microwave measurements of dielectric and magnetic properties of materials are discussed and referenced. The reference period extends to mid-1966. Measurement errors are analyzed where appropriate.**

I. Introduction

MAGNETIC PERMEABILITY and electric permittivity measurements are reviewed from the viewpoint of radio and communications applications. A knowledge of these parameters may be applied in circuit design and wave transmission calculations. The actual measurements usually involve three steps: 1) solution of the boundary value problem, 2) dimensional gauging of the system, and 3) electrical measurements on the system as, for example, the complex impedance or the complex resonant frequency of the system. This paper gives references for the standard configurations that are widely used for dielectric and magnetic measurements. Additional references on some aspects are given in another review [59]. When it is feasible and the state of development warrants it, a discussion of errors is given. Of special interest is how these measurements depend upon the state of development of basic RF measurements—attenuation, impedance, etc. However, it is true that dimensional gauging of the system and small departures of the system from the form assumed in the ideal boundary value problem are often the limiting factors, rather than the electrical observations.

II. Dielectric Constant Measurements

A. Reviews and Bibliographies

The book by von Hippel [1] covers the field of dielectric and, to some extent, magnetic measurements; it is a book of broad scope, including some practical examples. The chapter by Redheffer [2] on dielectrics compactly covers most of the microwave arrangements that may be used. A recent chapter by Altschuler [3] explains useful microwave methods in some detail. A *Handbuch der Physik* article [4] treats dielectric measurements rather completely. Section D-150 of a book by the American Society of Testing and Materials (ASTM) [5] gives a detailed treatment of measurements in capacitor-type dielectric holders. Shestopalov and Yatsuk [6] thoroughly review microwave dielectric measurements. There is an annual *Digest of Literature on Dielectrics* by the National Academy of Sciences—National Research Council [7] which contains a chapter on measurements. The same source issues an *Annual Report—Conference on Electrical Insulation* (the current issue is NRC Publication 1356) [8]. Two recent books give brief descriptions of capacitor dielectric measurements [9], [10].

Manuscript received January 13, 1967; revised March 3, 1967.
The author is with the National Bureau of Standards, Boulder, Colo.

B. Definitions

An electric field $Ee^{j\omega t}$ in a dielectric material produces in general a displacement current $j\omega\varepsilon_0\varepsilon' E$ and an ohmic (heating) current σE. MKS units are used, and ε_0 is the permittivity of free space, 8.85 pF/m. The conduction current σE may be formally defined to be $j\omega\varepsilon_0(-j\varepsilon'')E$. With this notation the total current is $j\omega\varepsilon_0\varepsilon^* Ee^{j\omega t}$, where $\varepsilon^* = \varepsilon' - j\varepsilon''$ is the complex relative permittivity and $\sigma = \omega\varepsilon_0\varepsilon''$. The loss tangent, tan δ, is $\varepsilon''/\varepsilon'$. The capacity between parallel plates of area S and separation s is $C = \varepsilon_0\varepsilon^* S/s$. The admittance of a capacitor is $Y = j\omega C$. The complex propagation constant in an infinite dielectric is $(-\varepsilon^*\omega^2/c^2)^{\frac{1}{2}}$, c being the speed of light, and the relative permeability μ^* being unity.

C. Capacitor Measurements of ε^*

A parallel-plate capacitor is commonly used as the test cell or sample holder for dielectric disk specimens [11]. Both a two-terminal-type holder [Fig. 1(a)] for roughly 10^{-2} to 10^9 Hz [11], [12] and a three-terminal-type [Fig. 1(c)] for roughly less than 100 kHz are used. Commercial two-terminal holders are available.

Customarily at NBS, Boulder, the specimen is smaller in diameter than the capacitor electrodes by at least 2s, so that the specimen will be in a uniform field. Some workers use larger samples, thus incurring fringing field corrections. A sample extending beyond the electrodes may, however, be used (with corrections) if an adsorbed foreign material on the sample surface furnishes a surface conductance which significantly augments the intrinsic loss tangent. At low frequencies, the three-terminal configuration may reduce the loss contribution due to the foreign-material conductance of the edge of the sample.

The complex capacitance, which gives the dielectric constant, may be observed in several ways. Bridges at various frequencies may measure the change of capacitance ΔC and of conductance ΔG due to putting the sample in and out, which yields

$$\varepsilon' = 1 + \Delta C/C_0 \tag{1}$$

$$\varepsilon'' = \Delta G/\omega C_0 \tag{2}$$

where $C_0 = \varepsilon_0 S/s$ is the vacuum capacitance of the sample

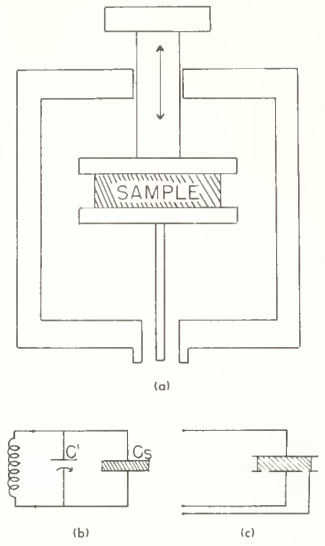


Fig. 1. (a) Hartshorn-Ward capacitive holder. Upper electrode is micrometer driven. (b) Representation of the holder (a) coupled to a coil; C_s is the capacity of the sample volume only, and C' includes all other capacitance. (c) Three-terminal guarded electrode holder used with Wagner-earthed bridge.

volume. Correction terms for the inductance and resistance of the connecting leads, and nonideal geometry are omitted above; see the literature [1], [4], [5], [11], [12].

A symmetrically disposed bridge of capacitors has been used for dielectric measurements over the range 8 to 500 MHz [13].

To attain accuracy, it is necessary to cope with the air gaps between the sample and the electrodes. Two methods of doing this are either to eliminate the gap by applying secondary electrodes such as rolled-on tinfoil or evaporated metal, or to work with a definite air gap. For the latter, see the ASTM publications [5] and the work of Morris in Bussey et al. [14]. The fractional error in ε' due to a small air gap is approximately

$$\Delta\varepsilon'/\varepsilon' = (1 - \varepsilon')g/s, \tag{3}$$

where g is the air gap and s is the sample thickness. This error increases with ε'. The exact form of (3) is obtained [5] by adding the sample and gap capacitances in series, $C^{-1} = g/\varepsilon_0 S + s/\varepsilon^* \varepsilon_0 S$.

The Hartshorn-Ward method [11] is an important contribution in capacitive dielectric measurements because corrections for lead inductance and resistance are usually negligible, and the micrometer of the holder may be the only high-precision component required. Yet the accuracy for ε' may be tenths of a percent, if air gaps are eliminated as will now be assumed. In this method, the total capacity $C(i)$ of the empty holder [Fig. 1(a)] is calibrated versus micrometer setting i by an accurate low-frequency method. This calibration may hold with negligible error up to several hundred megacycles [25], [26], [11], [12]. Resonance on a Q-meter or balance on a bridge is determined with the sample in. The sample is removed and the capacitance restored by means of the calibrated micrometer; then $\varepsilon' = 1 + \Delta C/C_0$, as in (1), where ΔC is the restored capacitance by the calibrated micrometer motion and C_0 is the vacuum capacitance of the sample volume $\varepsilon_0 S/s$. The loss is obtained from the change in Q or by the change of conductance ΔG. Lead inductance-resistance errors drop out.

The ε'' may be obtained from the change in Q when the sample is inserted. The analysis may be based on the representation in Fig. 1(b). The capacitance is divided into two parts: one, representing the sample region only, is $C_s = \varepsilon^* \varepsilon_0 S/s$ exactly (S = sample area), and the other, C', is that of the electrodes outside the sample plus any other shunting capacity of the leads, Q meter, and coil. The inverse Q contributed by the sample gives $\tan\delta$ as

$$\tan\delta = Q_s^{-1} \cdot C_s/(C_s + C'), \tag{4}$$

where $1/Q_s = 1/Q_{\text{with sample}}$ minus $1/Q_{\text{empty}}$. To avoid lead errors [11] Q should be measured as $2(C_s + C')/\Delta C$, where ΔC is the bandpass width measured by an auxiliary small shunt capacitor at or within the test cell. [See also (10).]

D. Immersion for High Accuracy

Liquid immersion methods [5], [15], [16] can measure $\tan\delta = 10^{-5}$ to 10^{-6} for certain materials, e.g., polyethylene.

The great accuracy of the ratio or inductive voltage divider bridge has been used to bypass the dimensional gauging limitation on accuracy [17]. Four measured capacitances, i.e., with and without the sample, using two immersion fluids (one being air), are sufficient to give ε' of a disk of unknown dimensions to about 0.01 percent accuracy.

E. Cavity and Transmission-Line Methods

The fields in a uniform transmission line, including open space, and in a cavity, vary as $\gamma^{j\omega t - \gamma z}$, where

$$\gamma^2 = k_c^2 - \varepsilon^* \omega^2/c^2, \tag{5}$$

in which z is distance along the line, c is the vacuum speed of light, and k_c is the transverse or cutoff wavenumber of the line; $k_c = 0$ in open space. There are many papers on VHF and microwave dielectric measurements. The principles are well known [1], [2], [18], [19].

1) Short-Circuited Line: The short-circuited transmission-line method [Fig. 2(a)] is simply explained and historically referenced by Dakin and Works [20]. The method is especially easy for nonbrittle low-ε' materials such as polymers that may be squeezed into the line tightly to avoid air-gap errors, which are of the same order in most lines [21] as the capacitive-gap error (3). However, the air gap may be filled by a conducting epoxy and measurements made of ε' up to 1500 [22]. The calculations utilize impedance methods. The impedance of the terminating short circuit is transformed through the section of line containing the sample and is

$$Z = (j\omega\mu_0/\gamma) \tanh(\gamma d) \tag{6}$$

at the input face of the sample, where d is the length of the sample and γ^2 is defined in (5). The input impedance at the sample face is determined by the traveling probe and VSWR, γ is found [1], [20] from (6), and ε^* is calculated from (5).

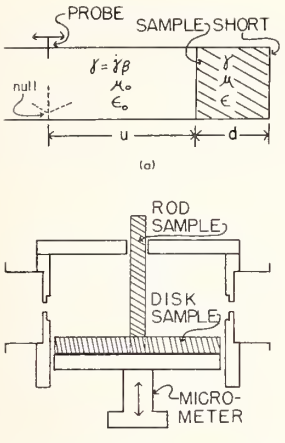


Fig. 2. (a) Sample in shorted line with traveling probe. Impedance at sample face, $(j\omega\mu/\gamma)$ tanh γd (see text), is measured by probe as $(\omega\mu_0/\beta)(1-jr\tan\beta u)/(r-j\tan\beta u)$, where β is phase-constant in air line and r is the VSWR. (b) Tunable TE_{01n}-mode cavity resonator, either one of the two alternative sample forms is used; groove in cylinder at irises lifts degeneracy with TM_{11n}-mode.

2) Resonators: The cavity resonator, at least in the TE_{01} mode, is very convenient for dielectric measurements over a wide range—ε' from 1 to 100 or more and loss tangents from 10^{-5} to 1.0. The main difficulty is that commerical cavities are not available; also, any one cavity is rather narrow-banded. The TE_{01}-mode cavity is being used in the national standards laboratories of the U.S.A., U. K., and U.S.S.R., and in some industrial laboratories. Either a disk sample or a coaxial rod sample may be used, as illustrated in Fig. 2(b). An advantage of this method is that loss tangents around 10^{-5} may be measured, e.g., on sapphire [23]. Another advantage is that the fit error, either of a gap around a disk sample or of a top hole for inserting a rod sample, is small and easily calculated. Finally, such difficult cases as titania ceramic with $\varepsilon^* = 95 - j1$, and salt water with $\varepsilon^* = 70 - j40$ are easily handled using a small rod and a tubular container, respectively. The analysis with a disk sample is essentially the same as in the transmission-line case; (6) is used and the length of the cavity measures Z at the top face of the disk. If the loss tangent is less than 0.1 it is usually valid to separate the real part and loss calculations. The latter utilizes the cavity Q. The above statements also apply to the rod sample case except that the propagation is looked upon as radial and the impedance at the sample boundary involves a Bessel function, e.g.,

$$Z = (\omega\mu_0/k)J_1(ka)/J_0(ka) \tag{7}$$

at $r=a$, the sample radius, where $k^2 = \omega^2\mu_0\varepsilon_0\varepsilon^* - k_g^2$, in which k_g is the guide wavenumber in the axial direction; k is the radial wavenumber in the sample. The observed data are the dimensions of the cavity at resonance, particularly the change in resonant length with the sample in place, the Q, and/or the cavity transmission coefficient. These data are sufficient to measure the impedance Z of (7), thus giving (ka) and the unknown ε^*.

3) Perturbation Method: The cavity perturbation method pioneered by Birnbaum and Franeau [19] is often simple, convenient, and reasonably accurate, especially for low values of ε' and moderate losses. A simple rectangular waveguide resonator in a TE_{103} or TE_{105} mode may be constructed by squeezing thin diaphragms with irises between waveguide flanges that are suitably separated; see ASTM publication C-525 [24]. Observed data are the volumes V and v of the cavity and the small sample rod (1- to 2-mm diameter at 9 GHz), and the frequency and Q, without the sample (f_1 and Q_1), and with the sample (f_2 and Q_2).

$$\varepsilon' = 1 + (f_1 - f_2)V/2vf_2,$$
$$\varepsilon'' = (Q_2^{-1} - Q_1^{-1})V/4v. \tag{8}$$

Perturbation theory is thoroughly discussed in most books on microwave ferrites, e.g., [46], [47].

4) Re-entrant Cavities: Re-entrant coaxial (hybrid) cavity resonators [25], [26] are used in the range 0.1 to 1 GHz. A parallel-plate capacitor region for the sample is formed by a gap in the center conductor. The capacitance of this variable gap versus separation is calibrated at a low frequency just as in the Hartshorn-Ward method [11]. Theoretically there is an inductance correction associated with changing the gap [25]. However, Beardsley [26] (and further unpublished NBS work) has shown experimentally that it is best to omit the inductance term in his cavity. In this case, the calculation of ε' is identical with that described for the Hartshorn-Ward method. The ε'' is obtained either from the loaded Q with the sample, $Q_{L,s}$, or from the change of the cavity transmission coefficient, i.e., by the generalized voltage ratio method.

F. Voltage Ratio to Measure Q Change

In the resonators mentioned, namely, the lumped elements on the Q meter, the microwave cavities, and the hybrid cavity, there is power coupled in from the generator, and an output power to a detector. In each case, the observed Q is the loaded Q_L summed up as

$$Q_L^{-1} = Q_u^{-1} + Q_g^{-1} + Q_d^{-1} \tag{9}$$

where u, g, d denote unloaded losses, generator coupling losses, and detector coupling losses, respectively. The voltage output to the detector is a measure of the amplitude in the resonator and this voltage changes with sample loss.

When the sample is put in and resonance restored, Q_L^{-1} includes the Q contribution of sample losses Q_s^{-1}. If $Q_{u,g,d}$ do not change during retuning as is assumed, for example, in the Hartshorn-Q meter method, and V_e, V_s denote the detector voltage empty and with sample, then it may be shown that

$$Q_s^{-1} = Q_{L,e}^{-1}(V_e/V_s - 1) \tag{10}$$

where e denotes empty, and s denotes with sample. In a lumped circuit $Q_{L,e}^{-1} = \Delta C/2C_{\text{total}}$. This Q_s^{-1} would be used in (4). Baker [10] discusses the voltage ratio method. In general, especially in microwave cavities, the Q's all change

during retuning, i.e., $Q_{u,g,d}^{-1}$ change to $K_u Q_u^{-1}$, $K_g Q_g^{-1}$, and $K_d Q_d^{-1}$, where the K's may be calculated from theory [23]. Equation (10) then becomes

$$Q_s^{-1} = K_g^{\frac{1}{2}} K_d^{\frac{1}{4}} Q_{L,e}^{-1} V_e / V_s - K_u Q_u^{-1} - K_g Q_g^{-1} - K_d Q_d^{-1}. \quad (11)$$

With (11) the voltage ratio method may be exact for any reasonator. Confirming microwave results [14], [23] are available. The advantage of the voltage ratio method is that it is often quicker and easier, and sometimes more precise to measure a small change in the voltage ratio than to measure the small difference $Q_{L,s}^{-1} - Q_{L,e}^{-1}$, the direct Q change.

G. Material Resonators

Dielectric and magnetic resonators have gained importance as suitable materials have developed. Such resonators may be used for measuring the material constants. There are two main types of dielectric resonators: those of large ferroelectric permittivity (ε' of the order of 1000) and small size which do not radiate very much, even with no metallic walls, and those of arbitrary permittivity with partial metallic walls such that the waves in the air space are evanescent. Those of the latter type are useful for ε' from 2 to 10 or more, especially for millimeter-wave dielectric measurements [27], [28]. A typical reference for the high-ε' type is Yee [29] who refers to Okaya and Barash; see also an early paper [30]. Dielectric resonators formed by fully enclosing a sample in platinum foil have been used by Westphal [31] for good high-temperature (1400°C) dielectric measurements.

H. Specialized Measurements

The literature on dielectric measurements is responding to the recent emphasis on thin films, semiconductors, optical communications, etc. Permittivities may be measured by laying down a strip line on a film and measuring the propagation constant [32]. Thin films of known thickness may be measured in a cavity resonator by usual methods (cf. Conklin [33]) which will usually reduce to perturbation methods because of the small sample volume. Ellipsometry methods [34] (reflections with both perpendicular and parallel electric vector) can be used in the millimeter-wave and optical region to determine both the thickness and the complex permittivity of a film.

Eddy-current methods may be useful in the audio through the VHF frequency range in order to circumvent contact potential problems in dielectric measurements on electrolytes, semiconductors, and some ferroelectric and ferrite materials. A useful configuration for frequencies from 0.5 to 100 MHz has been analyzed [35]. The holder is a pot core, and the sample is a toroid.

Dielectric measurements on liquids using capacitive holders in the kHz and the 100-MHz range have been carefully analyzed [36], [37], Liquids may be measured at microwave frequencies by cavity methods [18], [19], [38] and by millimeter-wave ellipsometry and interferometry methods.

Permittivities of gases are measured by filling a capacitor or a cavity resonator with the gas [39]. In a cavity, $(\omega/c)^2 = (k_c^2 + k^2)/\varepsilon'$, where k_c and k are axial and radial wave numbers that depend on the fixed dimensions. Thus ε' is measured by the frequency shift when the gas is let in.

The complex permittivity of semiconductors is being measured in waveguide in order to study the conductivity [40], [41].

I. Standard Samples and Comparison Measurements

It becomes feasible to issue standard reference samples when a stable material is available, and the properties in question can be measured with adequate, specified accuracy. A recent comparison of dielectric measurements by the national standards laboratories of the U.S.A., England, and Canada [14] demonstrated that these laboratories agreed to within ± 0.2 to 0.3 percent on ε' over the range 10^3 to 10^{10} Hz. The agreement on low loss tangents was usually ± 0.0001, but larger errors occurred. In the capacitive range foil, mercury, and air-gap electrodes were used. In the microwave range, cavity methods, essentially those of Horner et al. [18], and slotted lines were used. Cavities seem to be superior for measuring low losses.

One of the easiest ways to verify the accuracy of a measuring system is to measure the properties of a standard sample, i.e., a sample of known properties. Standard samples of silica and glass, $\varepsilon' - j\varepsilon'' = 3.82 - j\ 0.0005$ and $6.20 - j\ 0.033$, respectively, are available from NBS [42], [43].

J. Typical Measurement Errors

Errors in a precision standards laboratory may be smaller than in a typical engineering or industrial laboratory. International comparison [14] shows that ε' errors of ± 0.3 percent are expected on ordinary good dielectrics in a precision laboratory at RF and microwave frequencies, and recently (unpublished NBS work) at 140 GHz on liquids. The errors increase with frequency; errors on liquids at low frequencies are ± 0.1 percent.

For less critical requirements, the ε' errors for ordinary solid materials, $\varepsilon' = 2$ to 10, are 1 to 2 percent below 10 GHz, and increase to 5 percent or more at 140 GHz.

The tan δ errors on ordinary electronic materials (low-loss plastics, silica, alumina) are of the order of ± 0.00005 or ± 3 percent, whichever is larger, in a precision laboratory under a carefully controlled dry or vacuum environment. For less critical requirements, the usual tan δ error statement is ± 0.0001 or ± 5 to 10 percent, whichever is larger.

Absolute error determinations are not available for unusual materials, e.g., ferroelectrics, very lossy microwave absorbers, etc. However, in any specific problem the errors usually may be made rather small if a real need exists.

The errors originate in small part from the RF measurement errors, and in large part from the nonideal or uncertain geometry of the holder and sample. Air-gap or fit errors both in a capacitor and in most microwave structures contribute generally an error of 0.1 percent at least, and with a little carelessness and $\varepsilon' > 10$, the error may be 5 to 10 per-

cent. When small samples are used, as may be necessary for high-ε^* materials, an error in gauging the small dimension of the sample may be doubled or even quadrupled as it contributes to ε' errors.

K. Dependence on RF Measurements

The accuracy of dielectric measurements may depend upon the accuracy of the electrical measurements as well as on the gauging and conformity with the assumed boundary value problem. At low frequencies, the measurement of Q depends on the calibration of a voltmeter or power meter which may limit the accuracy to a few percent. In the range where good bridges are available, the loss accuracy is usually better than where Q is used; cf. White and Wilhelm [16]. The precision of a good bridge for loss measurement is equivalent to $1/Q$ of the order of 10^{-6}. Where high-accuracy rotating-vane attenuators are available in the microwave range, the Q may be measured to about 1 percent accuracy and 0.1 percent precision, which permits a precision and accuracy on low loss tangents of approximately ± 0.00001 [14], [23]. In coaxial systems, the direct attenuators are not sufficiently accurate, but heterodyne techniques permit very high accuracy [44], using the below-cutoff attenuator. In fact, since the measurement of bandpass width only requires relative insertion-loss measurements, most questions about the linearity or calibration accuracy of the detector and any UHF attenuator may be resolved with the thousandths-of-a-decibel precision of below-cutoff attenuators. Likewise, in the transmission-line impedance method of measurements where VSWR or reflection coefficient is to be measured, the highest accuracy is obtained using the below-cutoff attenuator to measure the standing wave profile, or the return loss.

III. Magnetic Measurements

The important radio magnetic materials are mainly thin films and nonconducting forms—powdered-iron suspensions and ferrites. These materials have many useful properties, are versatile, and require many measurements for their characterization. Some of the common permeability measurements will be described here, but switching and information-storage characteristics will be omitted. In some cases, the accuracy of measurement has not been of great concern. The interest has been more to explore the properties and to find what measurements are meaningful.

A. References

Magnetic activities are summarized annually in the *Magnetic Materials Digest* (latest issue by White and Wickersheim [45]). The papers of the Annual Conference on Magnetism are published each year in a special issue of the *Journal of Applied Physics*. Another international annual conference, INTERMAG, emphasizing engineering aspects, also issues proceedings. Two of five recent books on microwave ferrites are by Soohoo [46] and Lax and Button [47].

B. RF Scalar Permeability

The relative complex scalar permeability μ^* may be defined by accepting the constitutive equation $B=\mu_0\mu^*H$. Radio frequency measurements of $\mu^*=(\mu'-\mu'')$ are best done with toroidal cores. The adjectives describing this permeability, e.g., *initial*, *incremental*, etc., depend upon the field strengths, previous history, etc., and will not concern us here. The magnetic toroid in a coaxial holder presents a boundary value problem having essentially exact solutions.

A series of conferences on magnetic cores has been held; the latest conference gives a review of core measurements [48], which points out that the techniques for these measurements are now highly developed. The so-called standard method, used for calibrating the four methods to be enumerated, uses a short-circuited coaxial line of variable length and a bridge which acts as a high-sensitivity indicator. For example, the μ' of the sample is given by the change of length of the line [49], i.e., $\mu'-1=\Delta L/L$, where L is the sample length and ΔL is change in length of the line to restore the bridge balance. A coaxial bridge at 1 kHz to 100 kHz was investigated [48]. Mulhall [50] gives a comprehensive report of the Philips work, emphasizing bridge techniques that are useful up to 30 MHz. A second type of instrument, called the NBS permeameter, uses two toroidal cores. One core, of suitable properties, is wound with several turns of wire. The unknown core is loop-coupled to the driven core in a very symmetrical way by enclosing both in a small coaxial metal container. The change of the input impedance of the winding when the unknown sample core is taken in or out gives μ^* of the unknown core. The permeameter has been refined by evaluating a general T-network representation of the instrument [51]. A third method uses a split demountable coil winding for toroids. A fourth method uses a suitable length of coaxial line as a cavity resonator. These methods overlap in frequency and have been carefully cross-checked [52].

C. Magnetization

This quantity, important in the design of ferrite devices, is measured by vibrating samples and by the force balance. Case and Harrington [53] review the absolute calibration of the vibrating magnetometer, and Feldman and Hunt [54] discuss the instrumentation.

D. Effective Susceptibility and Intrinsic Permeability

More than one representation for ferrimagnetic interactions with electromagnetic waves has been used. Some microwave ferrite devices may be analyzed by what amounts to perturbation theory. For example, the perturbed propagation constant of a line partially filled with ferrite may be calculated. This procedure naturally makes use of an external or an effective susceptibility tensor, $[\chi_e]$, defined by the relation $\boldsymbol{m}=[\chi_e]\boldsymbol{h}_e$, where $\boldsymbol{m}$ is the RF magnetization and $\boldsymbol{h}_e$ is the external RF magnetic field. However, the internal RF field may be found from the external RF field by

using demagnetizing factors. Then the RF induction is $\boldsymbol{b} = [\mu]\boldsymbol{h}$ where $[\mu]$ is the intrinsic tensor and $\boldsymbol{h}$ is the internal (demagnetized) RF field. Steinert [55] gives convenient relations between $[\chi_e]$ and $[\mu]$.

Measurements on a small ferrimagnetic sphere in a microwave cavity are convenient and may be analyzed by means of cavity perturbation theory. The simplest and most natural result of the analysis is χ_{eff}, a certain element of the effective susceptibility tensor. The results are as follows:

$$\text{Real part of } \chi_{\text{eff}} \propto (f_1 - f_2)V/vf_2$$

$$\text{Imaginary part of } \chi_{\text{eff}} \propto (Q_2^{-1} - Q_1^{-1})V/v$$

where the symbols are defined as in (8). The recent books on microwave ferrites develop the theory of cavity perturbations by both ferrimagnetic and dielectric samples [46]. A perturbation theory for the impedance of a waveguide junction, containing a material of very general properties, is available [56].

E. Measurements of ΔH and χ''

The magnetized spin system of a ferrite is a resonant system. For a given shape of specimen, there is a resonance of $\chi_{\text{eff}} = \chi' - j\chi''$, occurring for a spherical sample at angular frequency $\omega = \gamma H_0$, where γ is the gyromagnetic ratio and H_0 is the external dc field. As H_0 is varied, the effective absorption χ'' goes through a maximum at resonance. The profile of χ'' versus H_0 is similar to that of any Lorentzian resonator. The width of this resonance curve at half-maximum is called the linewidth, ΔH. The Q or loss of the FMR motion is measured by χ''_{res} at resonance and by ΔH. These two measures of loss differ, and generally χ''_{res} is superior as a measure [65].

Measurements of χ''_{res} and ΔH of polycrystalline, i.e., wide-linewidth, materials are usually made on a small perturbing sphere in a TE_{10n}-mode rectangular cavity. ASTM specifications for ΔH, and also for saturation magnetization, and for permittivity are available [24]. The ASTM method for ΔH is explained by Preston and Case [57].

The linewidth of single crystal garnet (YIG) is of the order of 0.5 Oe. Such a ferrimagnetic resonant YIG sphere in a waveguide or transmission line acts itself like a cavity resonator coupled to the line. It is a material resonator of the first type as discussed in Section II-G. It has been shown that the loaded Q of such material resonators may be observed and separated into an external Q and an unloaded Q just as for any resonator [58]. For other remarks on material resonators and their use as delay lines, see Bussey [59].

F. Measurements of Tensor Permeability $[\mu]$

$[\mu]$ is discussed in the microwave ferrite books. Note that generally $[\mu]$ and $[\chi_e]$ or χ_{eff} resonate at different values of H_0. Measurements of $[\mu]$ are probably best made by successively applying left and right circularly polarized radiation. Either a small sphere may be used [47] or a centered cylindrical specimen in a TM_{ij0}-mode cavity. The latter case possesses exact solutions [60], [61]. Recent work [62], [63] makes use of $[\mu]$ in designing devices.

G. Spin-Wave Coupling

Spin waves, i.e., Fourier-like wave components representing nonuniform magnetization, are especially significant in polycrystalline ferrites where voids produce nonuniformity. The so-called uniform precessional mode (the mode usually excited in microwave devices) of the FMR is coupled to the spin waves. This coupling is an important loss mechanism for the FMR. Furthermore, as the shape of the piece of ferrite material is changed, or as the dc field H_0 is changed, the loss engendered by the spin-wave coupling changes. This change of loss arises because the dispersion relations for the spin waves and uniform mode depend differently on H_0; different regions of the spin-wave spectrum become degenerate with the uniform FMR mode. Thus, ΔH and χ''_{res} of a simple spherical sample are not adequate in general for characterizing either other ferrite shapes or devices with H_0 not at resonance. New measurements, [64], [65], where the effective sample shape is changed by tilting a spheroid, demonstrate these effects. Figure 3 illustrates, for a certain ferrite, how the loss varies with internal field, varied by tilting. The ASTM sphere measurement [24] would only give the loss at one place near 2200 Oe on the abscissa for this ferrite.

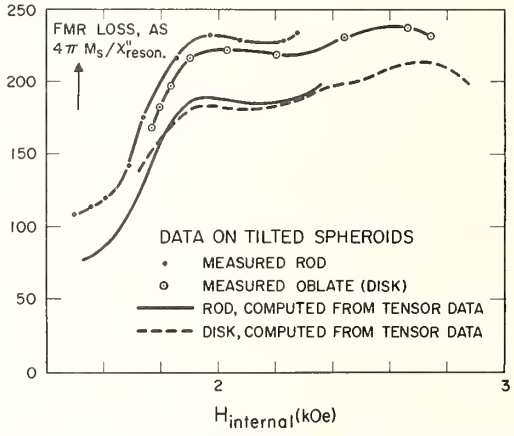


Fig. 3. Ordinate gives the peak loss at ferromagnetic resonance, measured as $4\pi M_s/\chi''_{\text{eff,resonant}}$ versus internal biasing field as the abscissa. $4\pi M_r/\chi''_{\text{eff,resonant}}$ is approximately equal to ΔH.

In addition to the directly measured data, Fig. 3 shows data computed [65] from intrinsic tensor permeability measurements [60]. The general agreement of the two methods demonstrates that intrinsic data from one shape of sample contain the correct information for other shapes. Any general program of device design should utilize either measured intrinsic data or measured loss variations as in Fig. 3. This is necessary because the variation of loss with field illustrated in Fig. 3 is not predicted by available theories. It must be measured.

References

[1] *Dielectric Materials and Applications*, A. R. von Hippel, Ed. Cambridge, Mass.: M.I.T. Press, 1954.

[2] R. M. Redheffer, "The measurement of dielectric constant," in *Techniques of Microwave Measurements*, C. G. Montgomery, Ed. New York: McGraw-Hill, 1947, ch. 10, pp. 561–676.

[3] H. Altschuler, "Dielectric constant," in *Handbook of Microwave Measurements*, 3rd ed., M. Sucher and J. Fox, Eds. Brooklyn, N. Y.: Polytechnic Press, 1963, ch. 9, pp. 495–545.

[4] L. Hartshorn and J. A. Saxton, "The dispersion and absorption of electromagnetic waves," in *Handbuch der Physik*, vol. 16, S. Flugge, Ed. Berlin: Springer-Verlag, 1958, pp. 640–725.

[5] "ASTM standards on electrical insulating materials," Committee D-9; "Capacitor, hybrid cavity, and air gap methods," D-150; and "Immersion method," D-1531, American Society for Testing and Materials, Philadelphia, Pa.

[6] V. P. Shestopalov and K. P. Yatsuk, "Methods of measuring dielectric constants at microwave frequencies," *Soviet Uspekhi*, vol. 4, no. 4, pp. 617–636, 1962.

[7] NAS-NRC Committee, *1965 Digest of Literature on Dielectrics*, vol. 29, National Academy of Sciences, Washington, D. C., NRC Publ. 1461, 1965.

[8] NAS-NRC Committee, *Annual Rept.—Conf. on Electrical Insulation*, National Academy of Sciences, Washington, D. C., NRC Publ. 1239, 1964.

[9] J. C. Anderson, *Dielectric*. New York: Reinhold, 1964.

[10] W. P. Baker, *Electrical Insulation Measurements*. New York: Chemical Publishing, 1965.

[11] L. Hartshorn and W. H. Ward, "The measurement of permittivity and power factor of dielectrics at frequencies from 10^4 to 10^8 cps," *J. IEE* (*London*), vol. 79, pp. 567–609, November 1936.

[12] M. G. Broadhurst and A. J. Bur, "Two-terminal dielectric measurements up to 6×10^8 Hz," *J. Res. NBS*, vol. 69C, pp. 165–172, July 1965.

[13] S. L. Epshtein, *Izmeritel'naya Tekhnika*, pp. 49–51, June 1963; English transl.: "Measuring the parameters of insulating materials in the meter and decimeter wavelength range," *Meas. Tech.*, pp. 516–517, June 1963.

[14] H. E. Bussey, J. E. Gray, E. C. Bamberger, E. Rushton, G. Russell, B. W. Petley, and D. Morris, "International comparison of dielectric measurements," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-13, pp. 305–311, December 1964.

[15] T. Hazen, "Improved methods for measuring dielectric constant and loss characteristics of low loss polyethylenes," *Insulation*, vol. 11, pp. 45–53, April 1965.

[16] L. D. White and W. T. Wilhelm, "A 0.1 to 10 MHz dielectric specimen bridge with dissipation factor accuracy of 10^{-6}," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-15, pp. 293–298, December 1966.

[17] W. P. Harris and A. H. Scott, "Precise measurement of dielectric constant by the two-fluid technique," in *NAS-NRC Annual Rept., 1962 Conf. on Electrical Insulation*, National Academy of Sciences, Washington, D. C., NRC Publ. 1080, 1963.

[18] F. Horner, T. A. Taylor, R. Dunsmuir, J. Lamb, and W. Jackson, "Resonance methods of dielectric measurement at centimetre wavelengths," *J. IEE* (*London*), vol. 89, pt. III, pp. 53–68, 1946.

[19] G. Birnbaum and J. Franeau, "Measurement of the dielectric constant and loss of solids and liquids by a cavity perturbation method," *J. Appl. Phys.*, vol. 20, pp. 817–818, 1949.

[20] T. W. Dakin and C. N. Works, "Microwave dielectric measurements," *J. Appl. Phys.*, vol. 18, pp. 789–796, September 1947.

[21] H. E. Bussey and J. E. Gray, "Measurement and standardization of dielectric samples," *IRE Trans. on Instrumentation*, vol. I-11, pp. 162–165, December 1962.

[22] G. A. Burdick, T. J. Lyon, and J. E. Pippin, "Measurements of large dielectric constants and loss tangents at 35 Gc/s," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-13, pp. 318–323, December 1964.

[23] H. E. Bussey, "Cavity resonator dielectric measurements on rod samples," *Annual Rept. of 1959 Conf. on Electrical Insulation*, National Academy of Sciences, Washington, D. C., NRC Publ. 756, pp. 15–20, 1959.

[24] "Tentative methods of test for ferrimagnetic resonance linewidth and gyromagnetic ratio," C524-63T; "Complex dielectric constant (microwave)," C525-63T; and "Saturation magnetization or saturation induction (by a vibrating sample magnetometer)," C527-63T, American Society for Testing and Materials, Philadelphia, Pa.

[25] C. N. Works, "Resonant cavities for dielectric measurements," *J. Appl. Phys.*, vol. 18, pp. 605–612, 1947.

[26] J. H. Beardsley, "A variable length re-entrant cavity for dielectric measurements from 100 to 400 Mc," *Rev. Sci. Instr.*, vol. 24, no. 2, pp. 180–181, 1953.

[27] B. W. Hakki and P. D. Coleman, "A dielectric resonator method of measuring inductive capacities in the millimeter range," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-8, pp. 402–410, July 1960.

[28] G. B. Walker and S. P. Luthra, "Use of ghost modes to determine dielectric constant and loss tangent," *Proc. IEE* (*London*), vol. 109B, suppl. 23, pp. 853–857, 1962.

[29] H. Y. Yee, "Natural resonant frequencies of microwave dielectric resonators," *IEEE Trans. on Microwave Theory and Techniques* (*Correspondence*), vol. MTT-13, p. 256, March 1965.

[30] R. O. Bell and G. Rupprecht, "Measurement of small dielectric losses in material with a large dielectric constant at microwave frequencies," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-9, pp. 239–242, May 1961.

[31] W. B. Westphal, "Dielectric constant and loss measurements on high temperature materials," Lab. for Insulation Research, M.I.T., Cambridge, Mass., Tech. Rept. 182, 1963.

[32] H. A. Wheeler, "Transmission-line properties of parallel strips separated by a dielectric sheet," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-13, pp. 172–185, March 1965.

[33] G. E. Conklin, "Measurement of the dielectric constant and loss tangent of isotropic films at millimeter wavelengths," *Rev. Sci. Instr.*, vol. 36, pp. 1347–1349, September 1965.

[34] E. Passaglia, R. R. Stromberg, and J. Kruger, "Ellipsometry in the measurement of surfaces and thin films," NBS Misc. Publ. 256, 1964.

[35] R. C. Powell and A. L. Rasmussen, "A radio-frequency permittimeter," *IRE Trans. on Instrumentation*, vol. I-9, pp. 179–184, September 1960.

[36] L. Hartshorn, J. V. L. Parry, and L. Essen, "The measurement of the dielectric constant of standard liquids," *Proc. Phys. Soc.* (*London*), sec. B, vol. 68, pp. 422–446, April 1955.

[37] J. V. L. Parry, "The measurement of permittivity and power factor of dielectrics at frequencies from 300 to 600 MHz," *Proc. IEE* (*London*), vol. 98, pt. III, pp. 303–311, July 1951.

[38] G. Roussy and M. Felden, "A sensitive method for measuring complex permittivity with a microwave resonator," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-14, pp. 171–175, April 1966.

[39] A. C. Newell and R. C. Baird, "Absolute determination of refractive indices of gases at 47.7 GHz," *J. Appl. Phys.*, vol. 36, pp. 3751–3759, December 1965.

[40] J. N. Bhar, "Microwave techniques in the study of semiconductors," *Proc. IEEE*, vol. 51, pp. 1623–1631, November 1963.

[41] K. S. Champlin, D. B. Armstrong, and P. D. Gunderson, "Charge carrier inertia in semiconductors," *Proc. IEEE*, vol. 52, pp. 677–685, June 1964.

[42] *Federal Register*. Washington D. C.: Gov't Printing Office, pt. 230, p. 1201, January 29, 1966.

[43] "Standard reference materials: Catalog and price list of standard materials issued by NBS," NBS Misc. Publ. 260, Section 4.23 of Quarterly Insert Sheets, April 1, 1966.

[44] W. E. Case, R. D. Harrington, and L. B. Schmidt, "Ferrimagnetic resonance measurements using IF substitution techniques," *J. Res. NBS*, vol. 68C, pp. 255–259, October–December 1964.

[45] *Magnetism and Magnetic Materials: 1965 Digest*, R. L. White and K. A. Wickersheim, Eds. New York: Academic, 1965.

[46] R. Soohoo, *Theory and Application of Ferrites*. Englewood Cliffs, N. J.: Prentice Hall, 1960.

[47] B. Lax and K. J. Button, *Microwave Ferrites and Ferrimagnetics*. New York: McGraw-Hill, 1962.

[48] R. D. Harrington and A. L. Rasmussen, "Magnetic core permeability measurement techniques," *Proc. Magnetic Core Conf.*, vol. 7, pp. 11–24, 1965.

[49] R. D. Harrington, "Cavity techniques for permeability measurement in the VHF region," *Proc. Electronic Components Conf.*, pp. 27–29, May 1955.

[50] B. E. Mulhall, "The measurement of magnetic permeability at radio frequencies," *Philips Res. Rept.*, vol. 19, pp. 78–102, May 1955.

[51] C. A. Hoer and A. L. Rasmussen, "Equations for the radio-frequency magnetic permeameter," *J. Res. NBS*, vol. 67C, no. 1, pp. 69–76, 1963.

[52] C. A. Hoer and R. D. Harrington, "Parallel reversible permeability

measurement techniques from 50 kc/s to 3 Gc/s," *J. Res. NBS*, vol. 67C, no. 3, pp. 259–265, 1963.
[53] W. E. Case and R. D. Harrington, "Calibration of vibrating-sample magnetometers," *J. Res. NBS*, vol. 70C, pp. 255–262, October–December 1966.
[54] D. Feldman and R. P. Hunt, "Circuit techniques for vibrating sample magnetometers" (in German), *Zeits für Imstrumkde*, vol. 73, no. 3, pp. 62–65, 1965.
[55] L. A. Steinert, "Geometrical anisotropy of magnetic materials in waveguides and cavities," *J. Appl. Phys.*, vol. 30, no. 7, pp. 2106–2107, 1959.
[56] D. M. Kerns and W. T. Grandy, Jr., "Perturbation theorems for waveguide junctions, with applications," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-14, pp. 85–92, February 1966.
[57] C. C. Preston and W. E. Case, "Tables to facilitate the determination of FMR linewidth of non-metallic magnetic materials," NBS Tech. Note 173, pp. 1–30, 1963.
[58] A. S. Risley and H. E. Bussey, "Interpretation of ferromagnetic resonance measurements made in a nonresonant system," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-15, pp. 393–396, December 1966.
[59] H. E. Bussey, "Progress in measurement of electromagnetic properties of materials, 1963 to 1965," in *Progress in Radio Science*, S. Silver, Ed. Berkeley, Calif.: Space Sciences Lab., University of California, 1967.
[60] H. E. Bussey and L. A. Steinert, "Exact solution for a gyromagnetic sample and measurements on a ferrite," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-6, pp. 72–76, January 1958.
[61] L. B. Schmidt, R. D. Harrington, and W. E. Case, "Tensor permeability measurements at *L*-band frequencies using a degenerate mode cavity," *J. Res. NBS*, 1967, to be published.
[62] W. H. von Aulock, "Theory of linear ferrite devices for microwave applications," *J. Appl. Phys.*, vol. 37, no. 3, pp. 939–946, 1966.
[63] J. L. Allen, D. R. Taft, and F. K. Hurd, "Computer-aided design of ferrite devices using intrinsic material parameters," presented as Paper S5 at the 1966 12th Annual Conf. on Magnetism and Magnetic Materials, Washington, D. C.
[64] A. S. Risley and H. E. Bussey, "Ferromagnetic resonance relaxation, wide spin-wave coverage by ellipsoids," *J. Appl. Phys.*, vol. 35, pt. 2, no. 3, pp. 896–897, 1964.
[65] A. S. Risley, E. G. Johnson, Jr., and H. E. Bussey, "Polycrystalline spin wave theory of ferromagnetic resonance compared with the tilting experiment," *J. Appl. Phys.*, vol. 37, pp. 656–668, February 1966.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 55, NO. 6, JUNE, 1967
pp. 1046-1053
THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

Measurement of RF Properties of Materials, A Survey[1]

Equation (4), page 1047, should have read

$$\tan \delta = Q_s^{-1}(C_s + C')/C_s. \qquad (4)$$

The term in $C_s + C'$ will cancel in the Hartshorn-Ward method using (10), and C', the auxiliary capacitance shunting that of the sample C_s, may remain unknown.

Manuscript received January 22, 1968.
[1] H. E. Bussey, *Proc. IEEE*, vol. 55, pp. 1046–1053, June 1967.

Reprinted from the PROCEEDINGS OF THE IEEE
VOL. 56, NO. 4, APRIL, 1968
p. 729
THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 67C, No. 1, January–March 1963

Equations for the Radiofrequency Magnetic Permeameter

Cletus A. Hoer and Alvin L. Rasmussen

(October 25, 1962)

The rf permeameter is an impedance transformer for measuring toroidal-shaped magnetic materials at radiofrequencies. Several equations already exist for calculating permeability (μ') and dissipation factor (tan δ) from measured input impedances of the permeameter. The results from these equations do not agree with each other except over a small range of tan δ. In this paper the permeameter is represented by its equivalent T-network and an exact solution for μ' and tan δ is obtained in terms of the input impedances and the impedance of a calibration core. The resulting equations are valid for all values of μ' and tan δ for the conditions discussed in the paper. Values of μ' and tan δ calculated from these exact equations are compared with those calculated from previous equations, and these results, differing as much as several orders of magnitude, are shown graphically. A frequency range from 0.1 to 50 Mc/s and a tan δ range from 0.0005 to 5 are covered.

1. Introduction

Since the development of the radiofrequency permeameter [1] [1, 2, 3, 4] [2] several equations have been derived for calculating the permeability (μ') and dissipation factor (tan $\delta=\frac{\mu''}{\mu'}$ where μ' and μ'' are related by the complex permeability, $\mu^*=\mu'-j\mu''$) of toroidal-shaped magnetic materials. In all of these derivations various assumptions have been made about the transformer characteristics of the permeameter. Although useful in the design and application of the rf permeameter, these assumptions have led to a number of expressions each of which gives correct values of μ' and tan δ for only a limited range of tan δ. The equations here derived are valid for all values of μ' and tan δ for the conditions discussed in section 2. Problems arising from assuming certain characteristics of the transformer are circumvented by analyzing the equivalent T-network corresponding to the permeameter. It is assumed that the arms of the equivalent network are not a function of the load impedance. A low loss toroidal-shaped magnetic material is used as an impedance standard in evaluating the T-network. The T-network yields a set of exact equations which have been used in determining values of μ' and tan δ for tan δ ranging from very low loss (0.0005) to very high loss (5.0). These values agree within experimental error with those obtained by other methods not using a permeameter. An evaluation of all permeameter formulas shows that the former equations for tan δ are in error as much as several orders of magnitude for low loss materials, and that one of these equations for μ' is in error as much as an order of magnitude for high loss materials.

The use of the permeameter makes the measurement of many magnetic materials possible in the frequency range of 0.1 to 50 Mc/s. These results are not readily obtained using other techniques.

[1] See appendix A for a description of the rf permeameter.
[2] Figures in brackets indicate the literature references at the end of this paper.

2. Derivation of the Exact Equation

A brief derivation of the permeameter equations for μ' and tan δ is given in eqs (1) through (11). The rf permeameter is an impedance transformer and can be represented by the T–network shown in figure 1.

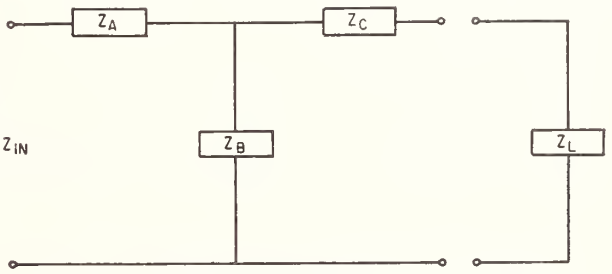


FIGURE 1. *Equivalent T-network for rf permeameter.*

The input impedance is $Z_{in}=Z_A+\frac{Z_B(Z_C+Z_L)}{Z_B+Z_C+Z_L}$, where Z_A, Z_B, and Z_C are the impedances [3] of the arms of the T-network, and Z_L is the impedance added to the secondary. If the input impedance is measured with three different known values of Z_L, an equation for any unknown Z_L may be derived eliminating Z_A, Z_B, and Z_C, the source of error in previous derivations. Let the input impedance be measured using the following definitions:

$Z_1=Z_{in}$ when the secondary is open; $Z_L=\infty$.
$Z_0=Z_{in}$ when the secondary is closed without a sample enclosed; $Z_L=0$ (this corresponds to the output terminals being shorted in fig. 1).
$Z_{1'}=Z_{in}$ when the secondary is closed with a known or standard impedance sample enclosed; $Z_L=Z_S$, where Z_S is the impedance of the standard sample minus the reactance of the space it replaces.

[3] The impedance due to the shorting plate is included in Z_C.

$Z_f = Z_{\text{in}}$ when the secondary is closed with an unknown sample enclosed; $Z_L = Z_u$, where Z_u is the impedance of the unknown sample minus the reactance of the space it replaces.

When a sample is enclosed in the secondary to measure Z_f or $Z_{f'}$, it is assumed that there is negligible direct coupling with the primary and that there is no significant change in the distributed capacitance of the secondary. If there is no measurable change in Z_1 when a sample is placed in the open secondary, then the coupling and change in distributed capacitance are negligible.

Expressing the input impedance in terms of Z_A, Z_B, Z_C, and Z_L gives

$$Z_1 = R_1 + j\omega L_1 = Z_A + Z_B, \tag{1}$$

$$Z_0 = R_0 + j\omega L_0 = Z_A + \frac{Z_B Z_C}{Z_B + Z_C}, \tag{2}$$

$$Z_{f'} = R_{f'} + j\omega L_{f'} = Z_A + \frac{Z_B(Z_C + Z_S)}{Z_B + Z_C + Z_S}, \tag{3}$$

$$Z_f = R_f + j\omega L_f = Z_A + \frac{Z_B(Z_C + Z_u)}{Z_B + Z_C + Z_u}. \tag{4}$$

From these equations the unknown impedance, Z_u, may be written

$$Z_u = Z_S \left(\frac{Z_{f'} - Z_1}{Z_{f'} - Z_0}\right)\left(\frac{Z_f - Z_0}{Z_f - Z_1}\right). \tag{5}$$

The product of the first two factors on the right is a constant for a permeameter at a given frequency. Defining this complex calibration constant as $\omega(x + jy)$ we have

$$\omega(x + jy) = Z_S \left(\frac{Z_{f'} - Z_1}{Z_{f'} - Z_0}\right). \tag{6}$$

By substituting (6) into (5) and expressing Z_u in terms of μ' and tan δ, (5) becomes

$$\omega L_a[\mu' \tan \delta + j(\mu' - 1)] = \omega(x + jy)\left(\frac{Z_f - Z_0}{Z_f - Z_1}\right), \tag{7}$$

where L_a=equivalent air inductance of the sample.[4] Solving (7) for μ' and tan δ gives [5]

$$\mu' - 1 = \frac{(x/y)B + A}{(R_{f1})^2 + (\omega L_{f1})^2}\left(\frac{y}{L_a}\right) \tag{8}$$

$$\tan \delta = \frac{(x/y)A - B}{(x/y)B + A}\left(\frac{\mu' - 1}{\mu'}\right) \tag{9}$$

where

$$A = R_{f0}R_{f1} + \omega^2 L_{f1}L_{f0}$$

$$B = \omega L_{f0}R_{f1} - \omega L_{f1}R_{f0}.$$

The calibration constants x and y may be expressed in terms of μ' and tan δ of the standard core. Z_S may be written

$$Z_S = \omega L_{as}[\mu'_s \tan \delta_s + j(\mu'_s - 1)].$$

Then from (6) we have

$$y = \frac{L_{as}[E(\mu'_s - 1) - F\mu'_s \tan \delta_s]}{(R_{f'0})^2 + (\omega L_{f'0})^2} \tag{10}$$

$$\frac{x}{y} = \frac{F(\mu'_s - 1) + E\mu'_s \tan \delta_s}{E(\mu'_s - 1) - F\mu'_s \tan \delta_s}, \tag{11}$$

where

$$E = R_{f'0}R_{f'1} + \omega^2 L_{f'1}L_{f'0}$$

$$F = \omega L_{f'0}R_{f'1} - \omega L_{f'1}R_{f'0}.$$

Primed subscripts and quantities with subscript s refer to the standard of known impedance.

Assuming that the load impedance does not change the characteristics of the T-network, these equations should be exact for all values of μ' and tan δ. The accuracy of the resulting answers appears to be limited only by the accuracy of the device used to measure the input impedances, the reproducibility of the secondary "short," and the accuracy of the standard impedance Z_S. Reducing errors due to the impedance measuring device has been discussed in the literature [1, 3].

The present secondary shorting arrangement is silver-to-silver surface contact formed under high and relatively evenly distributed pressure exerted by a cap which is screwed onto the permeameter (fig. 2). This arrangement yields a contact resistance

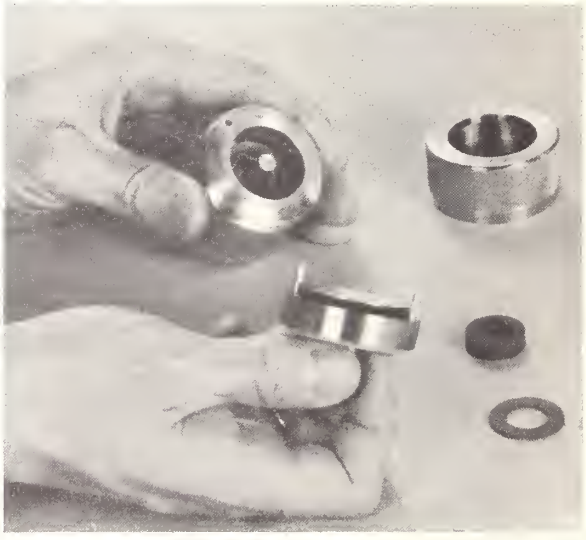


FIGURE 2. *An rf permeameter showing silver to silver contact surfaces.*

[4] $L_a = 2 \times$ (height in meters) $\times \ln$(outside diameter/inside diameter) $\times 10^{-7}$ henrys.

[5] The double subscripts denote differences, for example $R_{f1} = R_f - R_1$, $L_{f1} = L_f - L_1$.

which is reproducible to a precision approximately equal to or greater than that of the impedance measuring devices presently used. The error introduced by calibrating the permeameter with a known impedance may be minimized by choosing the proper calibration core.

3. Approximate Equations for Low Loss Samples

The working eqs (8) through (11) can be simplified considerably for low loss samples. These approximate equations are also useful in determining the desired characteristics of a standard calibration core. Equations (8) through (11) may be expressed in terms of resonating capacitance and Q by letting

$$Q_1=\omega L_1/R_1 \quad Q_0=\omega L_0/R_0 \quad Q_f=\omega L_f/R_f \quad Q_{f'}=\omega L_{f'}/R_f$$

$$C_1=1/\omega^2 L_1 \quad C_0=1/\omega^2 L_0 \quad C_f=1/\omega^2 L_f \quad C_{f'}=1/\omega^2 L_{f'}.$$

If the assumption is made that the square of each Q is much greater than one, that is

$$Q_1^2\gg 1 \quad Q_0^2\gg 1 \quad Q_f^2\gg 1 \; QQ_f^2, \quad \gg 1,$$

and that $\tan\delta\ll 1$, then (8) and (9) reduce to

$$\mu'-1=\frac{y}{L_a}\frac{L_{f0}}{L_{f1}}=\frac{y}{L_a}\frac{C_1}{C_0}\frac{C_{0f}}{C_{1f}} \tag{8a}$$

$$\tan\delta=\left(\frac{x}{y}-\frac{B}{A}\right)\left(\frac{\mu'-1}{\mu'}\right) \tag{9a}$$

where B/A reduces to

$$\frac{B}{A}=\frac{R_{f1}}{\omega L_{f1}}-\frac{R_{f0}}{\omega L_{f0}}=\frac{C_f}{C_{0f}C_{f1}}\left\{\frac{C_{0f}}{Q_1}+\frac{C_{f1}}{Q_0}-\frac{C_{01}}{Q_f}\right\}.$$

Also (10) and (11) reduce to

$$y=(\mu_s'-1)L_{as}\frac{L_{f'1}}{L_{f'0}}=(\mu_s'-1)L_{as}\frac{C_0}{C_1}\frac{C_{1f'}}{C_{0f'}} \tag{10a}$$

$$\frac{x}{y}=\frac{\mu_s'}{\mu_s'-1}\tan\delta_s+\frac{F}{E} \tag{11a}$$

where F/E reduces to

$$\frac{F}{E}=\frac{R_{f'1}}{\omega L_{f'1}}-\frac{R_{f'0}}{\omega L_{f'0}}=\frac{C_{f'}}{C_{0f'}C_{f'1}}\left\{\frac{C_{0f'}}{Q_1}+\frac{C_{f'1}}{Q_0}-\frac{C_{01}}{Q_{f'}}\right\}.$$

The results from (8a) through (11a) agree with the results from (8) through (11) to better than 1 percent for the NBS permeameters, provided that $\tan\delta<0.1$. In part 4 it is suggested that the calibration core have a low loss. The calibration constants x and y may then be calculated from (10a) and (11a).

4. Reference Sample or Calibration Core

The reference sample presently used at the National Bureau of Standards to find the calibration constants x and y is a stable low loss powered iron carbonyl SF core. The most accurate determination of y can be made if the calibration core is machined to have an L_a such that either $L_{f'1}/L_{0f'}$ or $C_{f'1}/C_{0f'}$ in (10a) approaches unity, depending on whether inductance or capacitance measurements are being made. Since y and C_0/C_1 do not change greatly over the frequency range of 0.1 to 50 Mc/s, one calibration core may be cut to have an L_a that will be near optimum for calibration over this frequency range.

The most accurate determination of x/y may be made by using a core with the lowest possible value of $\tan\delta_S$. The value of x/y for the National Bureau of Standards permeameters is of the order of 0.01. Examination of (11a) shows that if $\tan\delta_S$ is much less than 0.01, say 0.0001, x/y is approximately independent of $\tan\delta_S$ in which case $x/y\approx F/E$ and only an approximate value of $\tan\delta_S$ need be known. Therefore, the permeameter may be calibrated with a minimum of error by choosing a very low loss material having a permeability that is stable with time and temperature and having an optimum value of L_a.

The permeability μ_s' and loss tangent $\tan\delta_s$ of the SF calibration core were determined in the frequency range of 1.0 to 50 Mc/s as follows. The core was measured in demountable coils and coaxial lines below 1.0 Mc/s on a Maxwell-type bridge and in variable length re-entrant cavities above 50 Mc/s. The μ_s' remained constant from 0.1 to 100 Mc/s. To obtain a curve of $\tan\delta_S$ between 1.0 and 50 Mc/s the curve above 50 Mc/s was extended down and the curve below 1 Mc/s was extended up to make a smooth transition as shown in figure 3.

From (8a) and (10a) it can be shown that any error, $\Delta\mu_s'$, in μ_s' will produce an error, $\Delta\mu'$, in the μ' of any measured sample of an amount

$$\Delta\mu'=\Delta\mu_s'\frac{\mu'-1}{\mu_s'-1}. \tag{12}$$

The percent error in μ' is approximately equal to the percent error in μ_s'.

Likewise it can be shown from (9a) and (11a) that any error, $\Delta\tan\delta_S$, in $\tan\delta_S$ will produce an error, $\Delta\tan\delta$, in the $\tan\delta$ of any measured sample of an amount

$$\Delta\tan\delta=\Delta\tan\delta_s\frac{\mu_s'}{\mu_s'-1}\frac{\mu'-1}{\mu'}\approx\Delta\tan\delta_s. \tag{13}$$

Therefore if a more accurate determination of μ_s' and tan δ_S is made later, the μ' and tan δ of previously measured samples may easily be corrected using (12) and (13).

5. Q-Meter Measurements

Since the permeameter is often measured on a Q-meter, it is of value to note that most of the corrections to the measured Q and C due to residuals in the Q-meter cancel out when using (8a) through (11a). A Q-meter circuit including the residuals R_m, L', L_c, and R_v is shown in figure 4. The effective Q, Q_e, and the effective capacitance, C_e, can be expressed in terms of the indicated Q, Q_i, and indicated capacitance, C_i, as

$$C_e=\frac{C_i}{1-\omega^2L_mC_i} \tag{14}$$

$$Q_e=\frac{Q_i(1-\omega^2L_mC_i)}{1-\omega C_iQ_i(R_m+R_G')-\omega^2L_cC_i} \tag{15}$$

where

$$L_m=L'+L_c$$

and

$$R_G'=\frac{(\omega^2L_cC_i-1)^2}{\omega^2C_i^2R_v}.$$

In obtaining (14) and (15) it is assumed that

$$R_v^2\gg\left(\frac{1-\omega^2L_cC_i}{\omega C_i}\right)^2.$$

If each capacitance in (8a) and (10a) is replaced by the corresponding corrected expression from (14), the resulting equation for $\mu'-1$ is identical to (8a) in which indicated values of capitance are used. The corrections cancel.

If each capacitance and Q in (9a) and (11a) is replaced by the corresponding corrected expression from (14) and (15) the resulting expression for tan δ becomes

$$\tan\delta=\left(\frac{x}{y}-\frac{B}{A}+K\right)\left(\frac{\mu'-1}{\mu'}\right) \tag{16}$$

where K is a correction term which becomes significant above about 10 Mc/s.

$$K=\omega^2L_c\left[\frac{C_f}{C_{0f}C_{f1}}\left\{\frac{C_1C_{0f}}{Q_1}+\frac{C_0C_{f1}}{Q_0}-\frac{C_fC_{01}}{Q_f}\right\}\right.$$
$$\left.-\frac{C_{f'}}{C_{0f'}C_{f'1}}\left\{\frac{C_1C_{0f'}}{Q_1}+\frac{C_0C_{f'1}}{Q_0}-\frac{C_{f'}C_{01}}{Q_{f'}}\right\}\right]. \tag{17}$$

Each Q and capacitance in (16) and (17) is the indicated value as read on the Q-meter. Note that L_c is the only residual appearing in the correction term. The corrections due to L', R_m, and R_v cancel. Tan δ may be corrected either with (17) or by correcting each indicated Q with the expression

$$Q_e=\frac{Q_i}{1-\omega^2L_cC_i}$$

and using these corrected values of Q in (9a) and (11a).

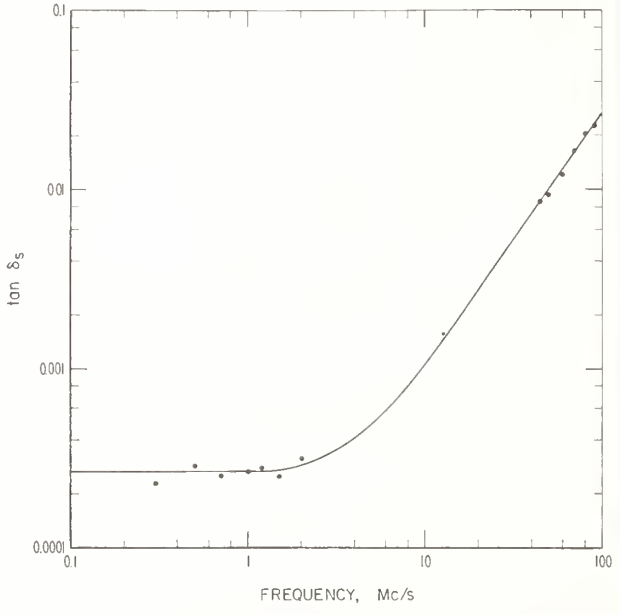


FIGURE 3. *Carbonyl SF calibration core,* tan δ_s *versus frequency showing demountable coil results (0.1 to 2.0 Mc/s) and re-entrant cavity results (45 to 100 Mc/s).*

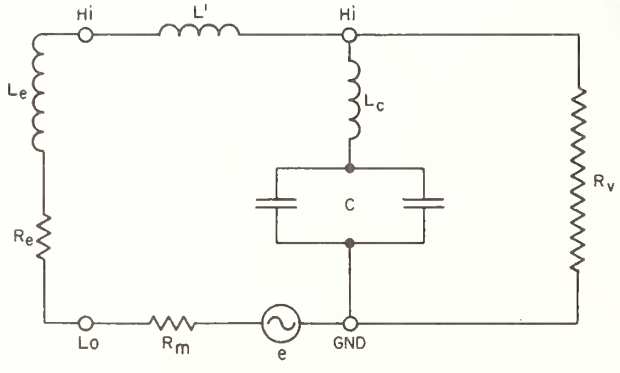


FIGURE 4. Q *meter approximate equivalent measuring circuit.*

6. Comparison With Previous Permeameter Equations

The following eqs. (18) through (24) have been used in the past to evaluate μ' and tan δ from permeameter data using the Q-meter.

From circuit (a) and assumptions 1, 2, and 3 (fig. 5) Haas [1] derived (18) and (19).

$$\mu'-1=\frac{K}{L_a}\frac{C_{0f}}{C_{f1}} \tag{18}$$

$$\tan\delta=\frac{C_{01}C_{f1}}{C_{0f}}\left[\frac{1}{C_{f1}^2}\left(\frac{C_f}{Q_f}-\frac{C_f^2C_1}{C_1^2Q_1}\right)-\frac{1}{C_{01}^2}\left(\frac{C_0}{Q_0}-\frac{C_0^2C_1}{C_1^2Q_1}\right)\right]. \tag{19}$$

Because (19) becomes negative when measuring relatively low loss materials, McKnight [2] derived (20) using circuit (b) and assumptions 1, 2, and 3

$$\tan\delta=\frac{C_{01}C_{f1}}{C_{0f}}\left[\frac{1}{C_{f1}^2}\left(\frac{C_f}{Q_f}-\frac{C_1}{Q_1}\right)-\frac{1}{C_{01}^2}\left(\frac{C_0}{Q_0}-\frac{C_1}{Q_1}\right)\right]. \tag{20a}$$

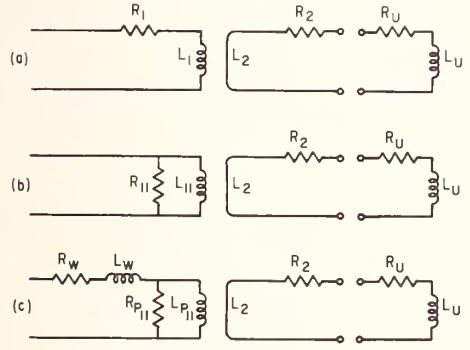
Equation (20a) reduces to (19) plus $2/Q_1$

$$\tan\delta=\tan\delta_{(19)}+2/Q_1. \tag{20b}$$ [6]

In another attempt to correct the error of (19), Rasmussen, Enfield, and Hess [3] derived eq (21) from circuit (c) and assumptions 1, 2, 3, 4, and 5. In the primary section the losses of the winding were separated from those of the material. Tan δ is here defined as μ''/μ', not as $\mu''/(\mu'-1)$ as was done in [1] and [2].

$$\tan\delta=\left(\tan\delta_{(19)}+2\frac{R_P}{X_P}\right)\frac{\mu'-1}{\mu'}. \tag{21}$$

[6] Subscripts on tan δ refer to equations giving the associated expression for tan δ.



ASSUMPTIONS

1. $R_1^2 << \omega^2 L_1^2$
2. $R_2^2 << \omega^2 L_2^2$
3. $(R_2+R_U)^2 << \omega^2(L_2+L_U)^2$
4. $(R_P/X_P)^2 << 1$
5. $2(R_P/X_P)\ \text{TAN}\ \delta_{(19)} << 1$
6. $2(R_P/X_P)\ \text{TAN}\ \delta_{(24)} << 1$

FIGURE 5. *Circuits used and assumptions made in deriving equations (18) through (24).*

R_p and X_p are the series resistance and reactance of the primary material.

Since assumption 3 is not valid when measuring relatively high loss samples, Rasmussen and Hess [4] derived (22) through (25), using circuit (c) and assumptions 1, 2, 4, and 6.

$$\mu'-1=\frac{K'}{L_a}\frac{C_0}{C_1}(D-1) \tag{22}$$

$$\tan\delta=\left(\tan\delta_{(24)}+2\frac{R_P}{X_P}\right)\frac{\mu'-1}{\mu'} \tag{23}$$

where

$$\tan\delta_{(24)}=\frac{\frac{1}{C_{f1}}\left(\frac{C_1}{Q_f}-\frac{C_f}{Q_1}\right)D-\frac{1}{C_{01}}\left(\frac{C_1}{Q_0}-\frac{C_0}{Q_1}\right)}{D-1} \tag{24}$$

and

$$D=\frac{\frac{C_{01}}{C_0}C_fC_{f1}}{C_{f1}^2+\left(\frac{C_1}{Q_f}-\frac{C_f}{Q_1}\right)^2} \tag{25}$$

A number of samples with losses ranging from 0.0005 to 5.0 were measured in the permeameter and the resulting data evaluated using eqs (8) and (9), and (18) through (23). The resulting values of μ' and tan δ for 5 of these samples are shown in figure 6. Also included are μ' and tan δ as measured in a coaxial line. Below 15 Mc/s, samples F3, P1, and P2 could no longer be measured in the coaxial line because of their low resistance.[7] To increase the resistance to a value within the limits of the measuring system, the volume of the material was increased by adding a number of cores of the same material to the coaxial line. In this way an average value of tan δ for each material was obtained at several frequencies below 15 Mc/s. These results are in close agreement with tan δ obtained by measuring individual samples in the permeameter and evaluating the data with eq (9), the exact expression for tan δ.

A study of the results from the different equations (with $\mu' \cong 4$ to 4,000 and tan $\delta \cong 0.0005$ to 5.0) shows that the equations may be used within the following limitations:

(1) Equation (18) can be used for all values of μ' provided tan δ does not exceed approximately 0.1. For tan δ greater than 0.1, (18) gives values for μ' that are too high.

(2) Equation (22) gave correct results for μ' in the range of μ' and tan δ studied. ($\mu' \cong 4$ to 4,000 and tan $\delta \cong 0.0005$ to 5.0).

(3) Equations (19), (20), (21), and (23) can be used for tan δ in the approximate range of 0.1 to 1.0, with the exception of (23) which can be used up to at least 5.0.

[7] The equivalent series resistance of the sample is $\omega\mu' L_a \tan\delta$.

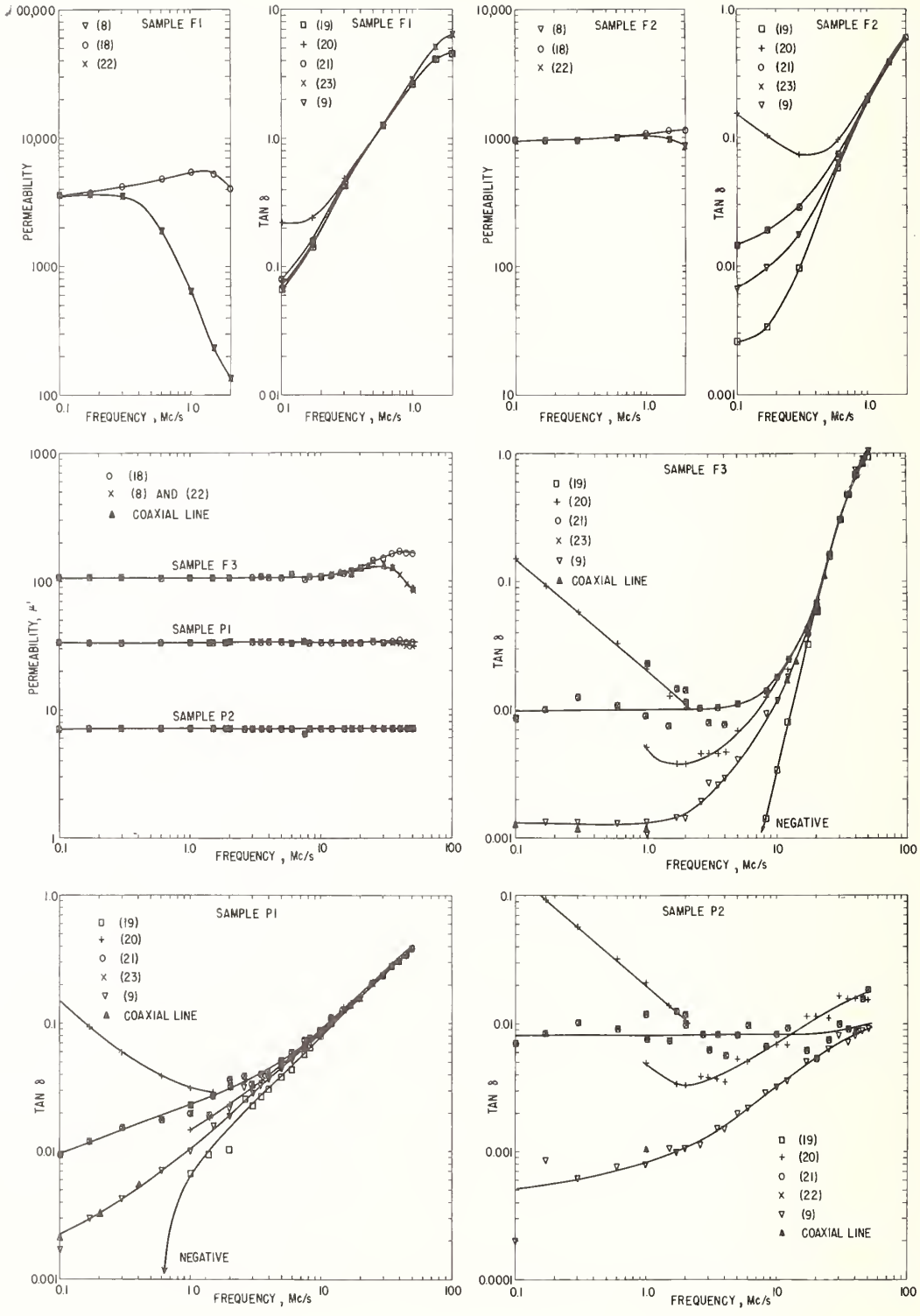


FIGURE 6. *Three ferrite samples (F1, F2, and F3) and two powdered iron samples (P1 and P2) measured in rf permeameters on a Maxwell type bridge (0.1 to 2.0 Mc/s) and on a Q meter (1.0 to 50 Mc/s).*

The data were evaluated using the different equations discussed in the text. The coaxial line values below 2.0 Mc/s were found by filling a coaxial line with a number of like samples and obtaining a value of μ' and tan δ for the group representing the average of the values for the individual cores. (Typical sample size: OD = 1.00, ID = 0.60, H = 0.16 inches.)

(4) The exact equations (8) and (9) can be used for values of μ' and tan δ at least within the range of $\mu' \cong 4$ to 4,000 and tan $\delta \cong 0.0005$ to 5.0.

(5) The approximate eqs (8a) and (9a) agree with (8) and (9) to better than 1 percent for tan $\delta < 0.1$.

The authors thank Mr. W. A. Pittman for making the many measurements and calculations.

7. Appendix A. Description of the RF Permeameter

The rf permeameter is an impedance transformer. It has a wound powered iron or stable ferrite toroidal core as a primary and a coaxial shorting enclosure as a secondary which envelops both the primary and a toroidal test sample core in such a manner that the magnetic field is concentric with the coaxial structure. Primaries, provided with connectors for terminals at one end of the enclosure, have different input impedances depending upon the frequency of interest and cover a range of at least 20 kc/s to 50 Mc/s. The core to be tested is inserted into the permeameter secondary where it is surrounded by a nearly uniform current sheet. Relatively small impedance changes in the secondary are easy to detect at the input to the primary. Becausc the form of the test core is toroidal, corrections for demagnetization effects are not necessary.

In making measurements, the primary of the permeameter is placed in an arm of an impedance bridge or across the terminals of a Q meter. The input impedance of the primary of the permeameter is measured, first with permeameter secondary open, second with permeameter secondary shorted without a sample, and third with permeameter secondary shorted with either a standard or unknown sample enclosed. From these data, the initial permeability (μ') and the dissipation factor (tan $\delta = \mu''/\mu'$), i.e., the reciprocal of the Q of the material, are derived.

8. Appendix B. Accuracy Check Using Metal Rings

At radiofrequencies a metal ring or toroid has an impedance which can be calculated from its physical dimensions and the resistivity of the metal. Such a ring would be an ideal standard of impedance for calibrating the permeameter provided that the impedance of the ring was within the impedance range of the permeameter. It can be shown that the inductance and resistance of a metal ring at radiofrequencies are

$$L = 2(H-\delta) \ln \frac{OD-\delta}{ID+\delta} 10^{-9} \text{ henry} \qquad (26)$$

$$R = \frac{\rho}{\pi\delta}\left[\ln \frac{OD-\delta}{ID+\delta} + \frac{(H-\delta)(OD+ID)}{(OD-\delta)(ID+\delta)}\right] \text{ohm} \qquad (27)$$

where

$$\delta = \frac{1}{2\pi}\sqrt{\frac{\rho 10^9}{f}}, \text{ skin depth in cm,}$$

ρ = resistivity, Ω cm,

f = frequency in cycles/second,

OD, ID, H = outside diameter, inside diameter, height in cm.

Several brass and copper rings were made and used in an attempt to calibrate the permeameter, but the results lacked precision. The impedance of the metal rings and the Q of the permeameter with a ring enclosed were too low for accurate determination of the calibration constants x and y.

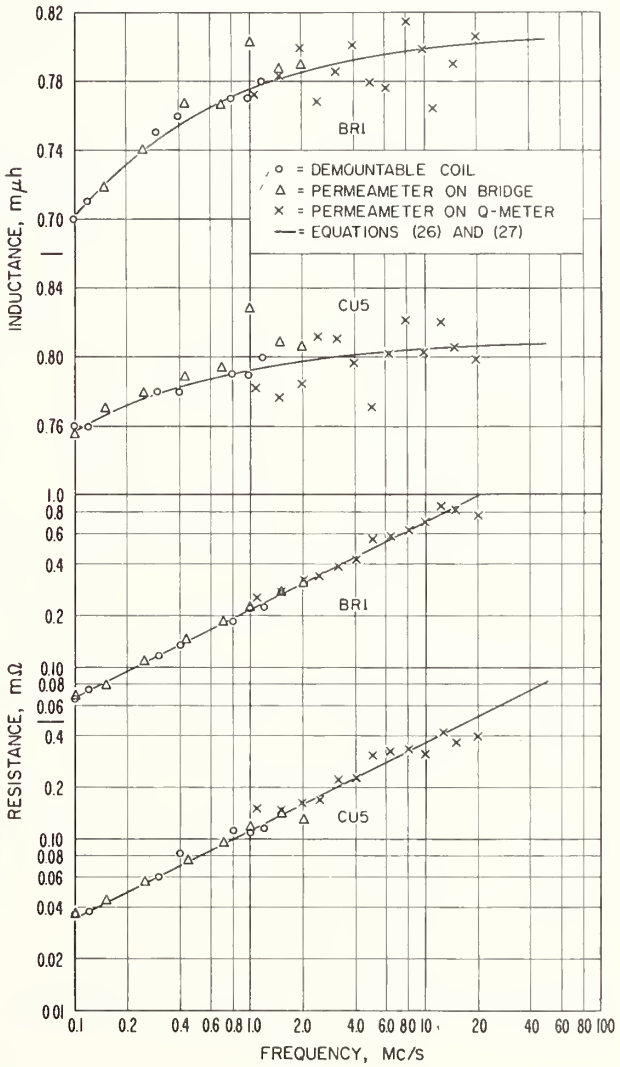


FIGURE 7. *Resistance and inductance of a brass (BR1) and a copper (CU5) ring as measured in a demountable coil and in the permeameter.*

The solid lines are calculated from equations (26) and (27).

It was decided that the metal rings would not make good calibration cores, but that they could be used to check the accuracy of the permeameter when using the exact equations. Figure 7 shows the results of the measured values of inductance and resistance versus frequency of two brass and two copper rings of different size. The rings were measured in the permeameter on a Maxwell-type bridge from 0.1 to 1.0 Mc/s and on a Boonton Q meter 260–A from 1.0 to 50 Mc/s. The SF core described in part 4 was used as the calibration core. The data were evaluated with the exact equations in the form

$$R-j\omega L=\omega(x+jy)\left(\frac{Z_f-Z_0}{Z_f-Z_1}\right). \tag{28}$$

The rings were also measured in a demountable coil [5] from 0.1 to 1.0 Mc/s on the Maxwell type bridge to compare the results of the permeameter and demountable coil with each other as well as with the results from (26) and (27).

As figure 7 shows, the results of the permeameter are in complete agreement with that from the demountable coil and from (26) and (27) within the precision of the measuring devices. The precision obtained for both the permeameter and the demountable coil is approximately 20 ph for L and 0.02 mΩ for R from 0.1 to 1.0 Mc/s on the bridge. The precision obtained from the permeameter on the Q-meter is approximately 50 ph for L and 0.1 mΩ for R.

9. References

[1] Haas, P. H., A radio frequency permeameter, J. Research NBS **51,** 221–228 (Nov. 1953) RP2454.

[2] McKnight, G. P., Permeameter for radio frequencies, Elec. Mfg. **52,** No. 4, 150 (Oct. 1953).

[3] Rasmussen, A. L., A. W. Enfield, and A. E. Hess, Advances in the design and application of the radio frequency permeameter, J. Research NBS **56,** 261–268 (May 1956) RP2673.

[4] Rasmussen, A. L., and A. E. Hess, RF permeameter techniques for testing ferrite cores, Elec. Mfg. **61,** No. 5, 86 (May 1958).

[5] Danielson, B. L., and R. D. Harrington, A technique for reducing errors in permeability measurements with coils. Proc. IRE **48,** No. 3, 365–366 (March 1960).

(Paper 67C1–121)

JOURNAL OF RESEARCH of the National Bureau of Standards—C. Engineering and Instrumentation
Vol. 67C, No. 3, July–September 1963

Parallel Reversible Permeability Measurement Techniques From 50 kc/s to 3 Gc/s*

Cletus A. Hoer and R. D. Harrington

(April 30, 1963)

New measurement techniques are described for determining the complex reversible permeability of ferrimagnetic materials from 50 kc/s to 3 Gc/s with d-c fields applied parallel to the rf fields in toroidal specimens. In the 50 kc/s to 50 Mc/s range, emphasis is directed towards recent improvements in the rf permeameter. For measurements in the 50 Mc/s to 100 Mc/s range, the feasibility of using variable length re-entrant cavities with quarter wavelength chokes is demonstrated. Two half-wave variable length cavities are described for obtaining reversible permeability data in the 100 Mc/s to 3000 Mc/s range. A technique for avoiding the use of quarter wavelines for isolating the d-c and rf circuits in these half-wave cavities is emphasized and represents a definite improvement over corresponding slotted line methods. A brief description of the d-c circuitry developed for these measurements as well as some typical spectrum data is also given.

1. Introduction

Considerable effort has been directed in recent years towards the study of the magnetic spectra of ferrite and garnet materials in the radio and microwave frequency ranges. This work is in general concerned with an evaluation of the frequency spectrum of the initial complex permeability ($\mu_i^* = \mu_i' - j\mu_i''$) in which only an rf field is applied to the sample or the measurement of the reversible complex permeability ($\mu^* = \mu' - j\mu''$) in which a d-c magnetic field is superimposed on the rf field. Several investigations of the dependence of the initial permeability spectrum on the composition of ferrites, the temperature environment, etc., have also been reported. However, the reversible permeability spectra studies with the applied d-c fields appears to be one of the more promising techniques for gaining further information concerning the resonance and relaxation phenomena observed in the magnetic spectra. Nevertheless, there has been relatively little work in this area compared to the effort that has been directed towards initial permeability spectra measurements.

Previous studies of the parallel reversible permeability spectra of ferrites as a function of frequency with applied d-c fields have been confined to measurements covering relatively limited frequency ranges such as the radiofrequency range or the microwave range [1–5].[1] These investigations have utilized primarily inductance coil measurements for obtaining data at lower frequencies or slotted line techniques for obtaining information at higher frequencies. It is the purpose of this paper to describe the use of the radiofrequency permeameter, and variable length re-entrant and half-wave length cavities for obtaining the reversible permeability of ferrimagnetic toroidal materials over the frequency range from 50 kc/s to 3 Gc/s with d-c fields applied parallel to the rf fields. Recent improvements in the permeameter as well as some of the advantages of variable length half-wave coaxial cavities over slotted lines for measurements of this type will be emphasized.

2. Radiofrequency Permeameter, 50 kc/s to 50 Mc/s

The advantages of the radiofrequency permeameter over inductance coil methods or coaxial transmission line techniques for initial permeability measurements at frequencies below 50 Mc/s have been discussed in the literature [6]. The satisfactory results obtained using the permeameter made it seem reasonable to extend its application to the measurement of reversible permeability. Therefore, the permeameter was modified to make parallel reversible permeability measurements by enclosing it in a d-c coaxial line as shown in figure 1. The d-c coaxial line may enclose the whole permeameter [7] as in figure 1a or only that part which encloses the sample, figure 1b. Both methods were used, and although the latter design is more difficult to make, it proved to be the better instrument. If the d-c coaxial line encloses the whole permeameter, the d-c field is applied to the primary core as well as to the sample being measured. The varying d-c field changes the characteristics of the primary which results in errors that are difficult to correct. The coupling between rf and d-c circuits is greater when the whole permeameter is enclosed, creating a need for filters in the d-c circuit from below 1 Mc/s to 50 Mc/s. When the primary was not enclosed by the d-c coaxial line, as in figure 1b, the primary was not affected by the d-c current, and filters were

*This work was partially supported by the Department of the Navy under a Bureau of Ships contract.
[1] Figures in brackets indicate the literature references at the end of this paper.

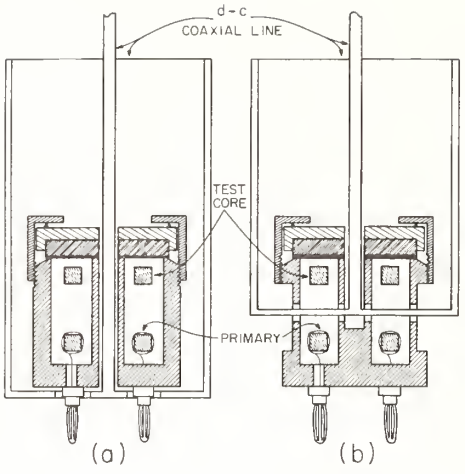


FIGURE 1. *Radiofrequency permeameter with d-c coaxial line to provide a d-c field parallel to the rf field.*

The d-c coaxial line in (b) cuts through the center of the permeameter to avoid applying the d-c field to the primary core.

needed only above 4 Mc/s. These filters are parallel L–C circuits tuned to have a maximum impedance at the frequency at which measurements are being made. The maximum impedance Z_{max} of such an L–C circuit at a frequency ω is approximately

$$Z_{max}=\frac{Q}{\omega C}$$

In addition to the requirements of high Q and low resonating capacitance, C, the characteristics of the filter must not change appreciably with greatly varying d-c currents. Good results were obtained with large powdered iron SF cores wound with bare No. 8 soft copper wire. The capacitors were small, variable mica capacitors. Up to 400 amps were applied for short periods of time with no significant change in Z_{max}. It was necessary to use a filter in both d-c leads to completely eliminate all external rf paths. Each filter was connected separately and tuned to give a maximum Q for the permeameter with the secondary closed and empty. These filters, the d-c leads, and connections for water cooling the center conductor of the d-c coaxial line are shown in figure 2. An exploded view of the permeameter and d-c coaxial line is shown in figure 3.

The measurement of reversible permeability is made in the same way as is the initial permeability, except for the application of the d-c field. A recent study of the permeameter equations for evaluating μ' and μ'' has resulted in a set of exact equations which are independent of the transformer characteristics [8]. The permeameter is represented by a T-network which is evaluated at a given frequency using a stable, low loss standard sample of known μ' and μ''. Using a standard sample in calibrating the permeameter, the procedure in measuring the reversible permeability of a sample is to measure the

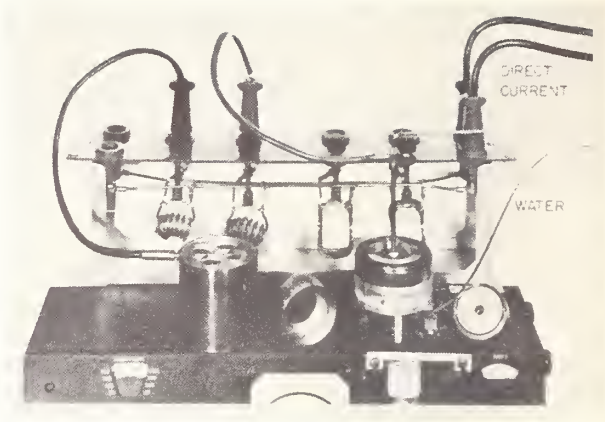


FIGURE 2. *Permeameter with d-c outer conductor (left) and shorting plate (right) removed.*

The filters are shown behind the Q-meter.

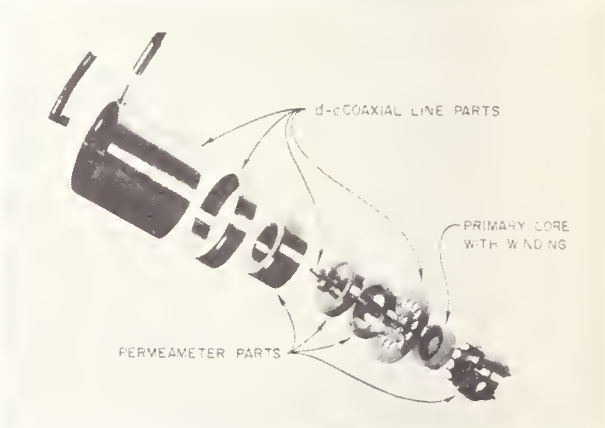


FIGURE 3. *Exploded view of permeameter of the type shown in figure 1b showing how d-c coaxial line is constructed through center of the permeameter.*

input impedance, Z_{in}, of the permeameter for the following conditions:

1. Secondary open; $Z_{in}\equiv Z_1$.
2. Secondary closed; no sample enclosed; $Z_{in}\equiv Z_o$.
3. Secondary closed; standard sample of known impedance Z_s enclosed; $Z_{in}\equiv Z_f'$.
4. Secondary closed; test sample enclosed; $Z_{in}\equiv Z_f$.

The d-c current is then turned on and Z_f measured at each desired value of current. The permeability, μ', and loss, μ'', of the test sample are then calculated at each current setting from

$$\mu''+j(\mu'-1)=\frac{Z_s}{\omega L_a}\left(\frac{Z_{f'}-Z_1}{Z_{f'}-Z_0}\right)\left(\frac{Z_f-Z_0}{Z_f-Z_1}\right) \qquad (1)$$

where L_a is the equivalent air inductance of the test sample and ω is the angular frequency. All input impedances (Z_1, Z_0, $Z_{f'}$, Z_f) are measured with the d-c leads connected to the coaxial line surrounding the permeameter.

3. Variable Length Re-entrant Cavity, 50 to 100 Mc/s

The upper frequency limit of the permeameter is about 50 Mc/s. Above 50 Mc/s slotted lines or cavities are in general used for initial permeability measurements. Slotted lines have been used in the 50 to 100 M/cs range but are cumbersome due to their large length. Somewhat smaller re-entrant cavities have also been used in the same frequency range by noting variations in either capacity or frequency for determining their resonant properties [9–12]. Measurements with a re-entrant cavity may also be made in terms of variations in length of the cavity [9]. The latter method has several advantages over the variable capacitance or the variable frequency method in that no capacity calibration need be made or no expensive frequency measuring equipment is needed. It is also possible to more easily minimize errors due to supports and discontinuities in the line which may become significant at higher frequencies.

On the basis of the above information, an investigation was carried out to determine if a variable length re-entrant cavity can be adapted for obtaining reversible permeability measurements. As in the case of the permeameter, it is necessary to isolate the d-c and rf circuit from each other. However, above 50 Mc/s it is difficult to use lumped circuit chokes for this purpose. On the other hand, the successful use of quarter wavelength transmission lines for this purpose by other investigators using slotted lines for their reversible permeability measurements suggests that the quarter wave choke method may also be applicable for re-entrant cavities.

The arrangement which we have found to give the best results is shown in figure 4. In this case the d-c current passes through the center conductor of the cavity in order to provide the circumferential field in the vicinity of the sample. The d-c current was confined to the center conductor by placing a thin sheet of insulation between the mounting flange of the quarter wavelength short and the cavity. No rf current could be detected in the d-c circuit outside of the cavity when the choke was tuned to approximately a quarter wavelength. It was observed that the setting of the quarter wavelength line is not critical for completely isolating the d-c and rf circuits.

The complex permeability may be evaluated from a determination of the change in the resonance length and Q of the cavity due to insertion of the sample. The equations and procedure for obtaining data are the same as those given in the next section describing variable length half-wave coaxial cavities. Essentially the same equations have also been given by van der Burgt, Gevers, and Wijn [9] who describe a variable length re-entrant cavity for initial complex permeability measurements. However, in their instrument, copper rings having the same dimensions as the sample were measured in order to obtain simplified equations.

The restrictions on the use of such instruments noted by the above authors such as high Q, small

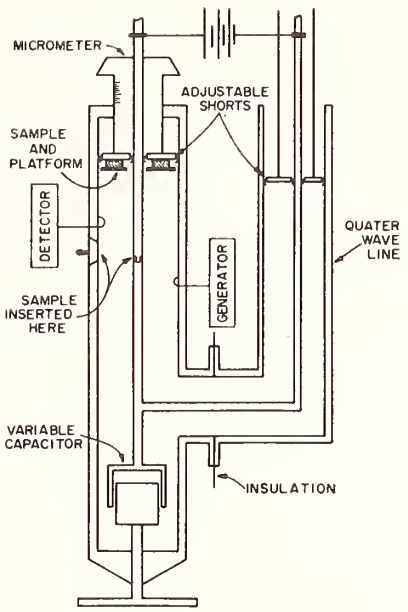


FIGURE 4. *Variable length re-entrant cavity with quarter wavelength line used from 50 to 100 Mc/s.*

The length of the quarter wavelength line relative to the reentrant cavity will in general be somewhat greater than indicated in the above simplified schematic.

length changes, etc., are also applicable to our instrument. A Q of several hundred, obtained with our instrument, appears to be satisfactory for determining the loss of most materials from 50 to 100 Mc/s. In addition, we have found it necessary to give consideration to possible errors due to temperature changes of the center conductor resulting from the large d-c currents required to obtain the bias fields. Such errors can be either eliminated by water cooling the center conductor or corrected for by making two identical runs, one with the sample in the cavity and one with the cavity empty. In our case, the latter method proved to be more feasible since some mechanical difficulty was experienced in obtaining water cooling in a cavity of this type.

4. Variable Length Half-Wave Cavities, 100 to 3000 Mc/s

The combination lumped and transmission line circuitry used in the previously described re-entrant cavity may result in somewhat more difficulty in the mechanical design of the instrument and in the rigorous analysis of the working equations than is the case for completely distributed parameter systems. As such, most magnetic spectra studies of the initial complex permeability in the range above about 100 Mc/s have utilized slotted line techniques. Somewhat less information has been reported on the use of variable length coaxial half-wave cavities for obtaining data of this type [13].

The measurement of parallel reversible permeability at these higher frequencies has been confined exclusively to the use of slotted lines in which the d-c circuit is isolated from the rf circuit by means of a quarter wavelength line [1, 3, 4]. On the other hand, an examination of variable length cavities suggests that an insulated d-c conductor could be placed through a hollow rf center conductor of such a cavity and thereby isolate the d-c from the rf signal without the use of a quarter waveline. Because of this advantage and our previous successful use of variable length cavities for initial permeability measurements, we developed two variable length cavities with hollow center conductors containing insulated d-c conductors. No coupling between the rf and d-c circuits was observed in either instrument. Since complete isolation is obtained without the use of a quarter waveline, this is a distinct advantage over slotted line techniques for reversible permeability measurements.

The design of both of these cavities was based on the Chipman method for impedance measurements [14]. Both cavities were mounted in a vertical position in order to avoid any supports on the center conductor. One of the instruments was designed for measurements below 300 Mc/s, while a smaller, more precise instrument was used for obtaining data above 300 Mc/s.

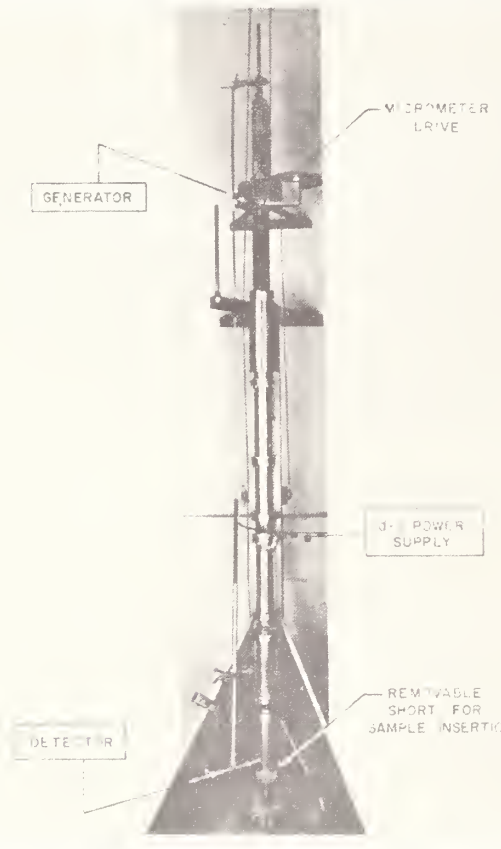


FIGURE 5. *Half-wave variable length cavity used at 100 Mc/s, 200 Mc/s, 300 Mc/s, etc.*

The larger instrument shown in figure 5 is presently designed for obtaining data at frequencies in the vicinity of 100, 200, and 300 Mc/s using 1, 2, or 3 half wavelengths respectively. The Q at these frequencies varies from 400 to 700. Continuous frequency coverage at these low frequencies would require a rather long micrometer drive. In lieu of this approach, the line has been constructed in removable sections to provide for coarse adjustments in length if information at other frequencies should be desired in the future. The general features of the instrument are similar to those shown in figure 6 which is actually a schematic for the higher frequency cavity described below.

Although the above described cavity can be used above 300 Mc/s, it was more convenient to design and construct a smaller, more precise instrument capable of continuous frequency coverage for measurements from 300 to approximately 3000 Mc/s. As in the case of the larger cavity, multiple half wavelengths can be used to obtain resonance at the higher frequencies. In addition, this cavity has the added advantage of a continuously variable coarse adjustment of length for changing the frequency range of the instrument. This adjustment was obtained by

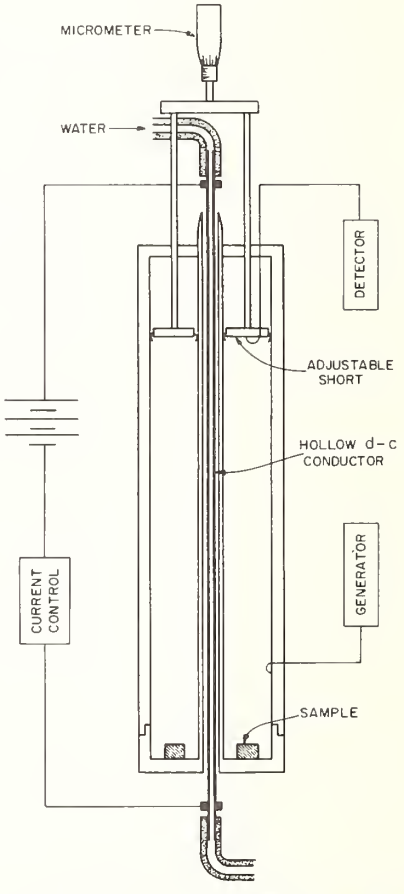


FIGURE 6. *Drawing of the 300–3000 Mc/s variable length cavity.*

A d-c field is applied to the sample by sending direct current through a d-c conductor which is placed inside of and insulated from the hollow rf center conductor.

threading the outside conductor of the cavity for a considerable distance as can be seen in figure 7. The cavity also has a higher Q, ranging from 500 to 3000, depending on frequency. To avoid any discontinuities near the sample at high frequencies, the center conductor and shorting plate of the 300 to 3000 Mc/s cavity were constructed as one piece as can be seen in figure 6. This arrangement gives a uniform distribution of rf current around the center conductor where it joins the shorting plate, and hence a uniform rf field for the sample which rests on the shorting plate. The cavity is opened for inserting the sample by removing the whole center conductor-short assembly. The cavity was prevented from heating up when the direct current was applied by making the d-c conductor out of ⅛ in. tubing through which water was circulated for cooling.

The complex permeability may be readily evaluated from a determination of the changes in resonant length and Q of the cavity when the sample is inserted. The toroidal sample which is placed against the shorted end of the line has a thickness small compared to a quarter wavelength in the material in order that dielectric effects do not influence the magnetic permeability measurements. Our experience has indicated that it is not difficult to obtain sufficiently thin ferrites so that the dielectric effect is negligible at least to 1 Gc/s and in some cases to 3 or 4 Gc/s.

For sufficiently small changes in the parameters of high quality lines, the complex permeability ($\mu^* = \mu' - j\mu''$) may be calculated from the equations

$$\mu' - 1 = \frac{\Delta L}{d}$$

$$\mu'' = \frac{\Delta \ell_f - \Delta \ell_e}{2d}$$

where it is assumed that the sample of thickness d has the same radial dimensions as the coaxial line. The remaining quantities in the equations are obtained from the following length measurements.

ΔL=The difference between the resonance length of the empty cavity and the resonance length of the cavity when the sample is placed against the short.

$\Delta \ell_e$=The difference between the two lengths of the empty cavity for which the power level is ½ the maximum value at resonance.

$\Delta \ell_f$=The difference between the two lengths of the cavity with the sample for which the power level is ½ the maximum value at resonance.

An excellent analysis of variable length cavities for initial permeability measurements in which these equations are developed in essentially the above form has been given by Eichacker [13, 15]. His work also demonstrated that variable length lines and slotted lines both work with the same evaluation formula. In this respect, it should be noted that the above equations have also been obtained in similar form by other authors using slotted line techniques [16].

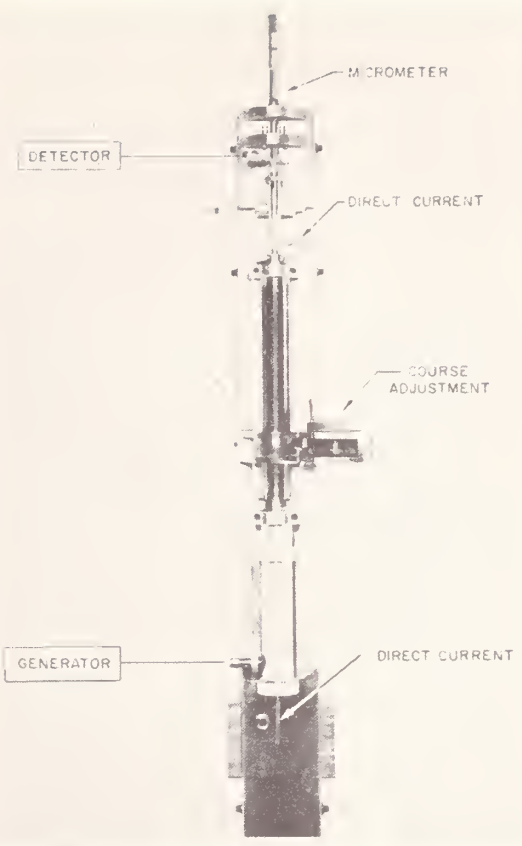


FIGURE 7. *Half-wave variable length cavity used at 300–3000 Mc/s.*

In many cases, it may not be convenient to obtain samples which have the same radial dimensions as the coaxial line. However, following the approach used by Bady and Franklin [10], the above equations may be written in the following form for loose fitting samples.

$$\mu' - 1 = \frac{\log \frac{B}{A}}{\log \frac{b}{a}} \frac{\Delta L}{d}$$

$$\mu'' = \frac{\log \frac{B}{A}}{\log \frac{b}{a}} \frac{(\Delta \ell_f - \Delta \ell_e)}{2d}$$

where the toroidal sample of inner radius a and outer radius b is located in the coaxial line of radii A and B.

5. Associated Equipment and Results

The rf equipment used with the half-wave cavities and re-entrant cavity consisted of commercially available signal sources with pads and matching stubs for the input circuit as well as facilities for monitoring the frequency of the system. The out-

put circuit consisted of matching stubs and calibrated IF detector systems or standing wavemeters depending upon the frequency. For the permeameter, the impedances are measured with a Q meter or Maxwell-type bridge.

Since in the cavities as well as the permeameter, large currents are needed to obtain d-c fields large enough to saturate many samples, a 6 v nickle-cadmium battery capable of high discharge current was used. It was desirable to be able to vary the current from 0.5 to 500 amp and also to have the current remain constant at any desired value during a measurement. To do this the current was passed through a number of power transistors connected in parallel. The large current I_c through the transistors can be varied or held constant by controlling the small bias current I_B. The circuit is shown in figure 8. The voltage drop V_s across the high current shunt is compared to a control voltage, V_c, set by the operator. The difference between these two voltages is amplified and used to drive a servomotor which changes the bias current in such a way as to increase or decrease I_c until V_s is equal to V_c. The current I_c is held constant at a value $I_c = V_c/R_s$ where R_s is the shunt resistance. A millivoltmeter across the shunt is used to measure I_c above 10 amp. Below 10 amp a clip on d-c ammeter is used.

A typical spectrum of the parallel reversible permeability of a magnesium ferrite obtained with the above described equipment is shown in figure 9. Data is obtained by measuring the reversible permeability and loss as a function of the d-c field at each frequency. An example of curves of this type for μ' only is given in figure 10. A complete set of these curves at all desired frequencies enables the spectrum to be obtained for any desired current. This data indicates that the above described equipment provides a highly satisfactory method for obtaining reversible permeability spectra information.

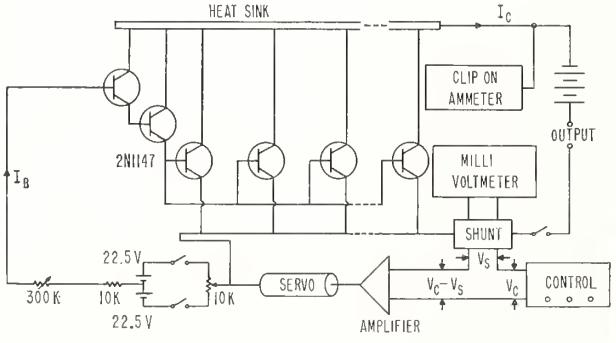


FIGURE 8. *Direct current control for varying or holding constant large d-c currents.*

The output current I_c passed by the power transistors is controlled by the bias current I_B. A control voltage V_c is compared to V_s and the difference amplified to drive a servo which adjusts I_B in such a way so as to increase or decrease I_c until $V_s = V_c$. I_c is changed by changing V_c.

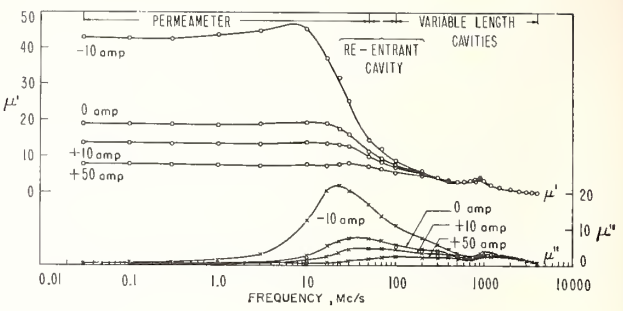


FIGURE 9. *Typical spectrum of the parallel reversible permeability of a magnesium ferrite.*

Data obtained from curves such as shown in figure 10. Sample dimensions outside diameter=0.961 in., inside diameter=0.678 in., thickness=0.040 in.

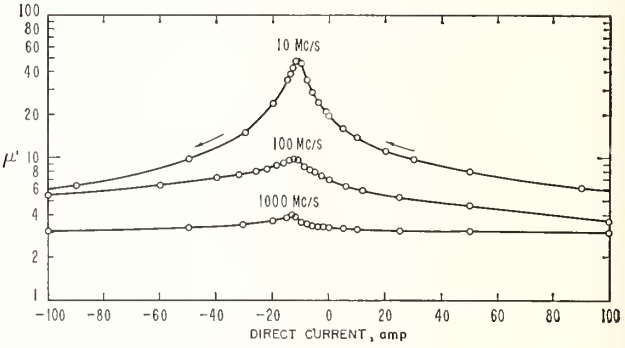


FIGURE 10. *Typical results of parallel reversible permeability versus applied d-c current.*

The sample was cycled at ±100 amps several times before measurements were made in the direction indicated by the arrows.

The authors thank W. A. Pittman for making the many measurements and calculations, and M. B. Lindell and G. A. Boschen for construction of the instruments.

6. References

[1] Brastins, A., and Williams, E. M., Separation of magnetic losses in UHF ferrites, IRE Trans. Instr. **I-10,** No. 2-63–67 (Sept. 1961).

[2] Fomenko, L. A., An investigation of magnetic spectra of of solid solutions of some NiZn-ferrites in the radio frequency range, Soviet Physics JETP **3,** No. 1, 19–28 (Aug. 1956).

[3] Harvey, R. L., Gordon, I., and Braden, R. A., The effect of a d-c magnetic field on the UHF permeability and losses of some hexagonal magnetic compounds, RCA Review **22,** 648–657 (Dec. 1961).

[4] Miles, P. A., Westphal, W. B., and von Hippel, A., Dielectric spectroscopy of ferromagnetic semiconductors, Reviews of Modern Physics **29,** No. 3, 279–307 (July 1957).

[5] Lombardini, P. P., Schwartz, R. F., Ferrites for high-power R-F tuning, Elec. Mfg. **62,** 60–71 (Aug. 1958).

[6] Haas, P. H., A radio-frequency permeameter, J. Research NBS **51,** No. 5, 221–228 (Nov. 1953) RP2454.

[7] Rasmussen, A. L., Hess, A. E., R-F permeameter techniques for testing ferrite cores, Elec. Mfg. **61,** No. 5, 86 (May 1958).

[8] Hoer, C. A., Rasmussen, A. L., Equations for the radio-frequency magnetic permeameter, J. Research NBS **67C,** (Eng. and Instr.) No. 1, 69–76 (Jan.–Mar. 1963).

[9] van der Burgt, C. M., Gevers, M., Wijn, H. P. J., Measuring methods for some properties of ferroxcube materials, Philips Technical Review **14,** No. 9, 245–256 (Mar. 1953).

[10] Bady, I., Franklin, R. J., Measurement of permeability and Q of magnetic materials over the frequency range of 50 to 500 Megacycles, Engineering Report 1159, Signal Corps Engineering Laboratories, Fort Monmouth, N.J. (Sept. 30, 1955) Published in 1957 IRE National Convention Record **5,** Pt 5, 137–145 (1957).

[11] Lindenhovius, H. J., van der Breggen, J. C., The measurement of permeability and magnetic losses of non-conducting ferromagnetic materials at high frequencies, Philips Research Reports **3,** 37–45 (Feb. 1948).

[12] Harrington, R. D., Powell, R. C., Haas, P. H., A re-entrant cavity for measurement of complex permeability in the very-high-frequency region, J. Research NBS **56,** No. 3, 129–134, (Mar. 1956) RP2657.

[13] Eichacker, R., A material-characteristics test assembly for determining the electromagnetic material constants of solid and liquid media at frequencies between 30 and 7000 Mc and temperatures between −60 and +240 °C, Rohde and Schwarz-Mittcilungen, No. 11, 185–205 (Dec. 1958), English Edition (Aug. 1959).

[14] Ginzton, E. L., Microwave measurements, p. 287 (McGraw-Hill New York, N.Y., 1957).

[15] Eichacker, R., Stoffmessung mit Messleitungen, Rohde and Schwarz-Mitteilungen, No. 15, 15–27 (Apr. 1961).

[16] Conrad, E. E., Porter, C. S., Doctor, N.J., and Franklin, P. J., Extension of the "thin-sample method" for the measurement of initial complex permeability and permittivity, J. Appl. Phys. **27,** No. 4, 346–350 (Apr. 1956).

(Paper 67C3–137)

Measurement and Standardization of Dielectric Samples*

H. E. BUSSEY†, SENIOR MEMBER, IRE, AND J. E. GRAY†

Summary—The selection of a material suitable for use as a standard of dielectric properties at microwave frequencies is discussed, and tests are described which indicate that a glass and a glass ceramic are satisfactory for such standards. The probable accuracy of measurement of the real part of the dielectric constant is estimated at ±0.3 per cent. Loss measurements are discussed. A correction is developed for the error resulting from the small airgap often present around the sample in transmission-line measurements. The effects of humidity and temperature variations are examined, and preliminary results of measurements to 800°C are given.

Introduction

STANDARD SAMPLES, *i.e.*, samples with the limits of error well known for the property in question, are useful in a laboratory in order to improve or confirm the accuracy of the measuring procedures and equipment being used.

There are two main requirements for standard samples of dielectric materials. The material properties should be understood sufficiently to avoid such problems as those associated with inhomogeneity, sensitivity to the environment, and chemical changes or aging. After satisfying this first requirement, the measurements then must be investigated sufficiently so that their accuracy can be specified with confidence. The emphasis in this paper is on accurate dielectric measurements. Material problems have been reduced by choosing materials with known isotropy, homogeneity, and long-term stability.

* Received September 4, 1962. Presented at the 1962 International Conference on Precision Electromagnetic Measurements as Paper No. 4.4.

† National Bureau of Standards, Boulder, Colo.

The Material for Dielectric Samples

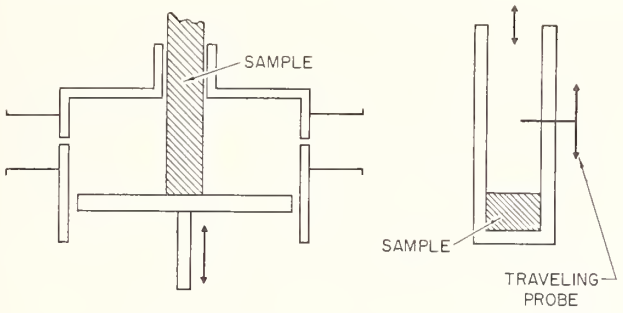
The main materials of the present investigation are a glass and a glass ceramic, both of which are available in optical quality. This ensures that the homogeneity, isotropy, and aging requirements are satisfied. There remains mostly the dependence of the complex dielectric constant on temperature, pressure, relative humidity, and surface conditions.

The temperature and relative humidity dependence of the glass standard samples were investigated over a reasonable range of values near usual laboratory conditions. In addition, a high temperature measuring system has been developed.

The Measurement of Dielectric Constant

The accuracy of the measuring procedures is determined mainly by comparing several independent methods. Corrections are made for some small departures of the measuring system from the ideal mathematical model.

Dielectric measurements depend upon knowing electromagnetic wave solutions in some space containing the material in question. The different methods investi-



(a) (b)

Fig. 1—TE_{01} resonator and TE_{11} line or resonator.

TABLE I

COMPARISON OF TE_{011} RESONANT CAVITY RESULTS FOR k' (PERMITTIVITY) WITH THOSE FROM THE IMPEDANCE METHOD USING TE_{11} MODE CIRCULAR WAVEGUIDE

Sample	TE_{011} resonant cavity, 9200 Mc 0.25-inch diameter rod	TE_{11} line, 8600 Mc, 1.0030-inch diameter		
		Uncorrected	Corrected for airgap	Sample diameter
Glass: from one corner of sheet	6.195	6.128	6.204	1.0001
From another location	6.196	6.127	6.203	1.0000
From another location	6.202	6.139	6.207	1.0004
Glass ceramic, from one corner of sheet	5.730	5.647	5.711	1.0001
From another location	5.708	5.648	5.708	1.0002
From another location	5.731	5.651	5.714	1.0001
Al_2O_3 porcelain (All samples from same batch but random locations in the ceramic body.)	8.784	8.642	8.794	1.0002
	8.775	8.674	8.822	1.0003
	8.775	8.594	8.743	1.0002
Average porcelain	8.778	8.636	8.786	—

gated are represented by the diagrams in Fig. 1. In Fig. 1(a) there is a circular cylindrical dielectric rod centered in a circular cylindrical cavity.[1] For the TE_{011} mode of resonance used, the corrections for imperfections due to the insertion port, the irises, and the gap around the plunger are small and known. These are obtained by perturbation theory using as a basis the exact solution of the ideal resonator with the sample in place.

In Fig. 1(b) there is a circular transmission line operated in the TE_{11} mode with a disk sample at a shorted end. Either by impedance measurement with a traveling probe or by resonating the line with another short, or both, the dielectric properties may be determined. The main error here is due to airgaps between the sample and the walls of the line. Such gaps not only permit the insertion and removal of the sample, but also leave it free to rest squarely on the short circuit, a condition which must be met if an even larger error is to be avoided.

A comparison of the two methods showed that the airgap around the disk sample contributed a significant error to the real part of the permittivity. The results in Table I show that when the TE_{11} data for the disk are corrected for the fit of the sample in the line there is good agreement with the TE_{011} rod data. The glass showed the best agreement. The glass and glass ceramic used were not of optical quality, but both are available in such quality. The results of such measurements on optical quality material are not expected to differ from those given here, except that the scatter of measured values may be less as a result of greater homogeneity of the optical quality material. The Al_2O_3 ceramic body, as might be expected, shows some variation from sample to sample, but the averages show good agreement between the two methods.

A general treatment of the error due to an airgap around a sample is not available, though some specific cases are covered.[2,3] Therefore, a general treatment based on perturbation theory, which seems to be adequate for practical cases, will be given.

[1] H. E. Bussey, "Cavity resonator dielectric measurements on rod samples," *1959 Annual Rept. of the Conf. on Electrical Insulation*, Pocono Manor, Pennsylvania, October 26–28, 1959, publication 756 of the Nat'l Research Council of the Nat'l Acad. of Sci.

[2] W. B. Westphal, "Techniques of Measuring the Permittivity and Permeability of Liquids and Solids in the Frequency Range 3 c/s to 50 kMc/s," M.I.T. Laboratory for Insulation, Cambridge, Mass., Res. Tech. Rept. No. 36; 1950.

[3] E. S. Hotston, "Correction term for dielectric measurements with cavity resonators," *J. of Sci. Instr.*, vol. 38, pp. 130–131; April, 1961.

Correction for Gap Around a Disk Sample

The complex propagation constant γ in the section of transmission line containing the sample is obtained from appropriate measurements and calculations and then the relative permittivity $k=\epsilon/\epsilon_0$, which may be complex to indicate losses, is obtained from

$$(\omega/c)^2 k = \gamma^2 + k_c^2, \tag{1}$$

where k_c is the characteristic or cutoff wave number for the mode and cross section of the line.

Based on perturbation theory[4,5] the change in the propagation constant of a TE or TEM line, due to an airgap between the dielectric cylinder and the cylindrical wall, is

$$\gamma_{\text{gap}}^2 - \gamma_{\text{no gap}}^2 = \Delta(\gamma^2) = (k-1)\left(\frac{\omega}{c}\right)^2 \frac{\int_{\text{gap}} \boldsymbol{E}_1 \cdot \boldsymbol{E}_2 dS}{\int_S |\boldsymbol{E}_1|^2 dS}, \tag{2}$$

where $\boldsymbol{E}_1$ is the electric field of the dielectric filled line with no gap, $\boldsymbol{E}_2$ is the electric field with an airgap present, S is the cross section of the line and dS is an element of S. When k is complex the change in γ is complex, and a loss correction is obtained concomitantly.

For a first-order theory the normal (n) and tangential (t) components of E_2 are obtained from E_1 by satisfying boundary conditions at the dielectric interface. These require that $E_{2,n}=k'E_{1,n}$ and $E_{2,t}=E_{1,t}$, where k' is the real part of k. Using these components in (2)

$$\Delta(\gamma^2) = (k-1)(\omega/c)^2 \frac{\int_{\text{gap}} \{k'|E_{1,n}|^2 + |E_{1,t}|^2\} dS}{\int_S |E_1|^2 dS}. \tag{3}$$

The second term becomes appreciable only for large gaps because E_t vanishes at the conducting wall. Very large gaps are only practicable when using the TE_{0m} modes, in which case the usually predominant term in E_n vanishes; a case already treated.[3] From (1) and (3), the error in a measurement of permittivity k due to an airgap is

$$\Delta k = -(k-1) \frac{\int_{\text{gap}} \{k'|E_{1,n}|^2 + |E_{1,t}|^2\} dS}{\int_S |E_1|^2 dS}. \tag{4}$$

[4] H. A. Bethe and J. Schwinger, "Perturbation theory for cavities," NDRC, Washington, D. C., Rept. No. DI-117; 1943.

[5] The general method is described by E. L. Ginzton in "Microwave Measurements," McGraw-Hill Book Co., Inc., New York, N. Y., ch. 10; 1957.

TABLE II

Comparison of Loss Tangent Measurements on Rods and Disks

Material	Rod (TE_{011})	Disk
Glass	53×10^{-4}	$56^*\times10^{-4}$
Glass	53	56
Glass ceramic	2.4	3.8
Glass ceramic	2.7	3.5
Glass ceramic	2.4	3.6
Al_2O_3 porcelain	9.7 ± 0.2	9.7 ± 0.2†

* Slotted line, except last sample.
† Disk in TE_{113} resonator at room temperature.

TABLE III

Measurement of a Plastic Sample Before and After a Reduction in Diameter to Illustrate the Effect of an Airgap. The Transmission Line Diameter was 1.0029 Inches

Diameter	k' Uncorrected	k' Corrected	$\tan\delta\times10^4$ Uncorrected	$\tan\delta\times10^4$ Corrected
1.0028	2.5330	2.5332	5.72	5.72
0.9808	2.4694	2.5364	5.17	5.26
0.9601	2.4188	2.5412	5.46	5.65
0.9186	2.3588	2.5525	5.68	6.03
0.8505	2.2085	2.5478	5.27	5.94

TABLE IV

Loss Tangent$\times10^4$ Measured at 9200 Mc vs Temperature and Humidity Variations

Material	4 Per Cent R.H. 24°C	31.5 Per Cent R.H. 35°C	48 Per Cent R.H. 23°C	83 Per Cent R.H. 24°C	Unclean surface 83 Per Cent R.H. 24°C
Glass ceramic	2.2	2.3	2.2	2.5	3.6
Glass ceramic	2.2	2.3	2.3	2.6	3.3
Glass ceramic	2.2	2.3	2.2	2.6	4.3
Glass ceramic	2.2	2.3	2.2	2.5	3.5
Glass	53	52	53	53	55
Glass	53	52	53	53	54
Glass	53	52	53	54	55

TABLE V

Complex Permittivity of Al_2O_3 Ceramic vs Temperature

Temperature °C	Dielectric Constant	Loss Tangent$\times10^4$
23	8.81	9.9
308	9.32	11
505	9.71	14
813	10.58	34

For comparison see M.I.T. Tables.[6]

[6] "Tables of Dielectric Materials," M.I.T. Laboratory for Insulation Research, Cambridge, Mass., Tech. Rept. No. 119, vol. V, p. 40; 1957.

Eq. (4) was applied to a TEM coaxial transmission line and a TE_{01} rectangular line and it agreed with the usual correction[2] for these modes when the gap was small. Eq. (4) may be applied to the important case of a circular TE_{11} transmission line for which no correction for fit was previously available. For a sample with radius $b-\Delta b$ in a waveguide of radius b,

$$\Delta k' = -k'(k'-1)(0.8368)\Delta b/b. \quad (5)$$

This equation was used to correct the results in Table I, as shown.

Experimental

The samples compared in Tables I and II were obtained as follows. Pairs of samples, a rod and a disk were cut side-by-side from the various corners of a glass and a glass-ceramic sheet about 12 inches square. Samples from the same batch of Al_2O_3 porcelain were available in each shape, but were not pairs from side-by-side locations.

In order to estimate the size of the airgap that may be treated by (4) without too much error, measurements were made with larger airgaps. The results appear in Table III. The same sample was used throughout this experiment, its diameter being cut down in steps as indicated. It was carefully centered in the line to minimize mode conversion.

To determine the effects of temperature and humidity variations on the measured properties, the TE_{011} resonant cavity and samples were enclosed in a glove box where the temperature and humidity could be controlled. The results of these measurements appear in Table IV (loss variation); the real part to two decimal places did not vary.

The high temperature equipment uses the method of Fig. 1(b). A section of silver circular TE_{11} transmission line is closed with a second short circuit which contains a coupling iris, forming a TE_{113} cylindrical resonant cavity whose Q and resonant frequency yield the complex permittivity of the sample. The first measurements made with this system are reported in Table V.

Discussion

The 1723 glass appears to be quite homogeneous and isotropic, based on the results in Table I. It seems probable that the various systems are capable of measurements on k' to an accuracy of 0.3 per cent.

The results shown in Table II confirm the accuracy of the present loss tangent measurements to ± 10 per cent ± 0.0001. It is expected that future work will improve this accuracy.

Based on Table III we estimate that with large gaps, corrections of more than 4 or 5 per cent will leave inaccuracies of 1 per cent or more in the final results for k'. The results make it doubtful whether the imaginary part of (4) should be used. When there is a gap there may be extra mode-conversion losses that makes up for the missing sample material. The dielectric constant at 9200 Mc of these materials, for two decimal accuracy, is independent of temperature and relative humidity over the normal range of laboratory conditions. Relative humidity and surface contamination do alter the apparent loss, as shown in Table IV, presumably by lowering the surface resistivity of the sample. The magnitude of this effect should vary for different measuring systems which may concentrate the applied electric field at the surface of the sample to varying degrees. It does appear, however, that when measurements can be made at low relative humidity, say below 30 per cent humidity, corrections are unnecessary.

Acknowledgment

The authors wish to thank Dr. W. H. Barney of the Corning Glass Works, Corning, N. Y., and L. E. Ferriera of the Coors Porcelain Company, Golden, Colo., for furnishing various samples.

Reprinted from IRE TRANSACTIONS
ON *INSTRUMENTATION*
Volume I-11, Numbers 3 and 4, December, 1962

Absolute Determination of Refractive Indices of Gases at 47.7 Gigahertz

A. C. Newell and R. C. Baird
National Bureau of Standards, Boulder, Colorado
(Received 2 April 1965)

The refractive index of a gas is determined to high accuracy by measuring the change in the resonant frequency when the gas is admitted to a previously evacuated microwave cavity. This measurement was made at a frequency of 47.736 GHz, and is the first such measurement in this frequency range.

The resonant frequency of the cavity is detected by means of a hybrid tee bridge circuit. When the cavity is matched to the waveguide the null at resonance is very sharp, and the frequency can be set to ±5 parts in 10^8.

Two types of resonant structures were used in this measurement, a conventional cylindrical cavity and a Fabry–Perot resonator using one plane and one spherical mirror. Measurements were made with both structures on each of the gases, and the agreement between the two was within the limits of experimental error. Because the Fabry–Perot resonator has no side walls, air could be passed through it much more easily, and it is therefore preferred for use with moist air.

The values of $(n-1)\times10^6$ for the gases at 0°C and 760 Torr are as follows:

Dry CO_2 free air	288.13±0.05,
Argon	277.48±0.05,
Carbon dioxide	495.16±0.08,
Helium	34.51±0.05,
Hydrogen	135.77±0.05,
Nitrogen	293.81±0.07,
Oxygen	266.95±0 05.

The primary values of these results lies in their increased accuracy over previous measurements and their confirmation of existing accurate data on these gases. In addition the use of the Fabry–Perot resonator as a refractometer is demonstrated and its advantages under certain conditions discussed.

INTRODUCTION

THIS work originated from the need for an accurate means of measuring the index of refraction of air in conjunction with a precision determination of the speed of light. This need required that the refractometer yield absolute values with an accuracy of one part in 10^7 when operated at the same frequency as used in the speed-of-light measurement.

One of the most accurate means of making this measurement is to measure the change in resonant frequency caused by the introduction of air into an evacuated cavity. This method has been used with some variations by several groups,[1,2] but the most accurate work appears to be that of Essen and Froome.[3] Our method was essentially that of Essen and Froome with some modification toward increasing the accuracy. A Fabry–Perot resonator was also developed and used as an independent check on the conventional cylindrical cavity.

THEORY OF MEASUREMENT

The index of refraction of a medium is defined as the ratio of the speed of an electromagnetic wave in a vacuum to its speed in the medium. Therefore,

$$n=\frac{c}{v}=\frac{\lambda f_e}{\lambda f_g}=\frac{f_e}{f_g}, \quad \text{or} \quad (n-1)=\frac{f_e-f_g}{f_g}, \qquad (1)$$

where n is the refractive index, λ the resonant wavelength of the cavity, f_g and f_e the filled and evacuated resonant frequencies, respectively.

The measurement is seen to be dependent upon maintaining the resonant-wavelength constant, or in other words the dimensions of the cavity must remain the same. As is discussed later, there are small changes in the dimensions of the resonant structures which lead to correction terms.

For air and most gases, the change in frequency, f_e-f_g, is a small fraction of the operating frequency. This fact, together with the required accuracy of the measurement, makes it imperative to have a stable frequency source and a precise means of setting and measuring the resonant frequency.

Resonance of the cavity is detected by means of the hybrid tee bridge circuit shown in Fig. 1. Microwave

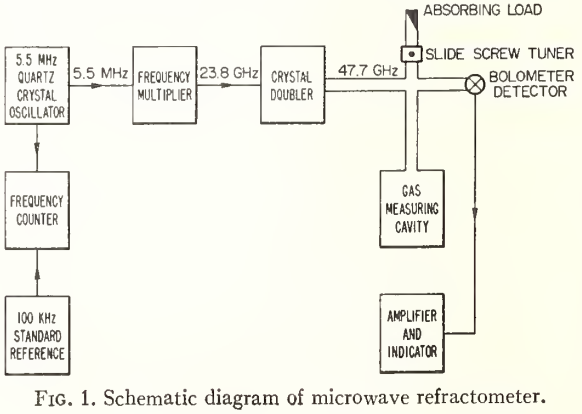


FIG. 1. Schematic diagram of microwave refractometer.

[1] C. M. Crain, Phys. Rev. 74, 691 (1948).
[2] G. Birnbaum, S. J. Kryder, and H. Lyons, J. Appl. Phys. 22, 95 (1951).
[3] L. Essen and K. D. Froome, Proc. Phys. Soc. B64, 862 (1951).

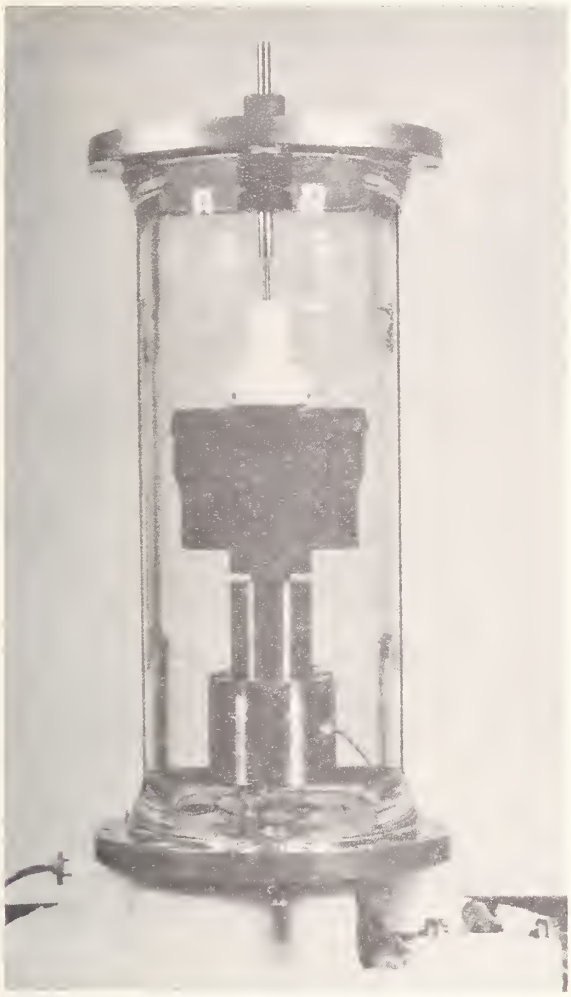


FIG. 2. Cylindrical cavity enclosed in bell jar.

power is derived from a frequency multiplier chain driven by a variable crystal-controlled oscillator. The signal is then fed into the E arm of the hybrid tee. A low-VSWR load and tuner are attached to one colinear arm of the tee and the cavity is attached to the opposite arm. The coupling hole of the cavity has been adjusted in size so that at resonance the cavity is matched to the waveguide. The reflected signal from the cavity, and hence the signal at the detector, is a minimum at resonance. This minimum is a precise indication of resonance. Its sharpness, and therefore the precision of setting to resonance, is determined by the symmetry of the tee and how well it is matched, the matching of the cavity and the load, and the adjustment of the load-arm tuner.

These factors are optimized as follows. As nearly symmetrical a tee as possible is chosen and E–H tuners attached to the input and output arms. With matched loads on the two colinear arms, the tuners are adjusted for minimum VSWR at their inputs. The tee is thus sufficiently well matched with isolation between the E and H arms typically 40 dB. This matching does not change the frequency at which the minimum output signal occurs, but it does make the minimum much sharper. This has been verified experimentally and is discussed in detail in the Appendix. Once the tee is matched the cavity is attached in place of one matched load, and a slide-screw tuner inserted between the tee and the other load. With no tuning in the load arm, the frequency is adjusted for minimum output, and then the tuner is adjusted to cancel out the small reflection from the cavity. This procedure produces a very sharp minimum which makes it possible to set the frequency to a precision of ± 5 parts in 10^8.

Since the output frequency of the multiplier chain is a constant multiple of the oscillator frequency, the frequency measurement is accomplished by counting the frequency of the oscillator with reference to a 100-kHz national standard signal. This reference is maintained accurate to at least one part in 10^{10}, and the short-term stability of the oscillator is 3 parts in 10^9.

CAVITY DETAIL

The first cavity used in the experiment is shown in Fig. 2. This is a conventional cylindrical cavity operating in the TE_{017} mode. It was constructed of invar to reduce its temperature coefficient, and the interior walls were silver plated. After the silver was deposited on the walls, a stainless-steel broach was passed through the cylinder to compress the silver plating. This increases the Q of the cavity, which for this one was about 30 000. The cavity is equipped with a tuning plunger which can be used to trim the resonant frequency; however, in use the plunger is left fixed during a measurement. Considerable effort was spent in eliminating undesired modes in the cavity. Some of these modes were introduced by the tuning plunger mechanism and were eliminated by altering the physical configuration of the plunger housing. A choke was cut in the end plate of the cavity to shift the frequency of the TM_{117} mode out of the working range. The final result is a cavity completely free of all modes save the desired TE_{017} mode throughout the cavity tuning range. The VSWR at resonance, looking into the cavity, remained less than 1.03 over the entire frequency range of the source, which was about 14 MHz, extending from 47 743 to 47 729 MHz. The cavity is mounted in a vacuum-sealed bell jar so that it can be filled with the gas under test and evacuated. The external tuning rod does not make rigid contact with the micrometer head, insuring that movements of the bell-jar end plates are not transmitted to the cavity.

Figure 3 shows a picture of the Fabry–Perot structure which was also used in the experiment. This instrument was designed by Zimmerer[4] who suggested its use as a refractometer. The resonator consists of a spherical brass mirror with a 22-in. radius of curvature

[4] R. W. Zimmerer, M. V. Anderson, G. L. Strine, Y. Beers, IEEE Trans. 2, 142 (1963).

facing a flat brass mirror. Power is coupled into the structure through a single hole in the center of the flat mirror. The mirrors are separated by three invar rods 11 in. long. The whole structure is enclosed in a bell jar in such a way that it is isolated from any movement of the bell-jar end plates. This apparatus was used as an independent check on the cylindrical cavity, but it also has some distinct advantages when measuring the refractive index of room air. When the air is admitted to the bell jar containing the refractometer there will be a period of time in which the water vapor will be adsorbed by the surfaces of the bell jar and refractometer. If the bell jar is sealed off soon after filling with air, this absorption results in a decrease in gas pressure and an increase in resonant frequency. This increase in frequency is due to the loss of water vapor from the gas in the cavity, and there is no measurable change in the resonant properties of the cavity caused by the adsorbed layer of water on the walls. This was determined by measuring the evacuated resonant frequency as a function of time soon after the air had been pumped out. If the layer of water vapor on the cavity walls had changed the resonant frequency there should have been a decrease in frequency as the adsorbed layer was pumped off. However, no such change was observed. Thus, when measuring the refractive index of air it is necessary to circulate the air through the refractometer to allow the system to remain at a condition of water vapor and thermal equilibrium. This circulation is also desirable to sample continuously the air and thereby detect the fluctuations in refractive index.

The cylindrical cavity had only small holes to allow air to enter the cavity, and this made it more difficult to arrive at water-vapor equilibrium. The small inlet holes also caused a good deal of turbulence within the cavity when air was pumped through it, making it necessary to stop the flow of air when the frequency was being set.

The Fabry–Perot refractometer is especially suited for use with air since it has no side walls to impede the flow of air between the mirrors. A large flow of air can be passed through the refractometer without appreciable turbulence, thereby making possible its use as a continuous monitor of refractive index. Also with this cavity, it was only necessary to measure the evacuated resonant frequency and associated rod temperatures two or three times a day. From these measurements the evacuated resonant frequency at any given rod temperature could be calculated. This eliminated the need to evacuate the cavity after each filled reading and greatly improved the room-air measurements. Using the Fabry–Perot refractometer in this way it was possible to measure the refractive index of the air in the laboratory to an accuracy of at least one part in 10^7.

The measurement of cavity temperature changes is very critical if the desired accuracy is to be realized, since any temperature change will cause a change in the cavity dimensions and therefore the resonant frequency. The temperature will rise sharply when gas is admitted to the evacuated bell jar and will decrease sharply when it is pumped out again. While these changes are at most only a few tenths of a degree Celsius, failure to allow for their effect can cause significant errors. For the accuracy required, it is necessary to measure absolute temperature accurately to 0.01°C and temperature changes to an accuracy of 0.002°C. It is also necessary that the temperature sensor be in intimate thermal contact with the cavity itself without transmitting any mechanical stress from the rest of the system. These requirements were fulfilled by using thermistor beads mounted in 8/32 screw heads. For the cylindrical cavity a small hole was tapped in the invar wall of the cavity and the screw and thermistor placed in the hole. A small piece of aluminum foil was placed over the end of the assembly to help shield the thermistor from external thermal radiation. The thermistor leads can be seen extending from the cavity wall in Fig. 2. In the Fabry–Perot structure it

FIG. 3. Fabry–Perot structure with hybrid tee attached.

was found necessary to use three thermistors to measure accurately the average temperature of the rods. One was placed in the middle of one rod, and one on each end of a different rod. The average temperature of the rods is then given by

$$\bar{T}=aT_B+bT_m+CT_t, \tag{2}$$

where T_B, T_m, T_t refer to the temperature at the bottom, middle, and top of the rods, respectively. The constants a, b, and C, as well as the temperature coefficient of the cavity α were determined experimentally by heating the instrument slowly and measuring the rod temperatures and resonant frequency. The constants were then calculated from a least-squares fit of the data to the equation

$$f_{\bar{T}}=f_{\bar{T}_0}+\alpha(\bar{T}-\bar{T}_0), \tag{3}$$

where $f_{\bar{T}}$ and $f_{\bar{T}_0}$ are the resonant frequencies at temperatures $\bar{T}$ and $\bar{T}_0$.

EXPERIMENTAL PROCEDURE

For the actual measurements, gas was admitted to the bell jar and a few minutes allowed for the gas and the metal to come to thermal equilibrium. The frequency of the microwave source was adjusted for minimum detector signal and the frequency was then measured. This setting to minimum and the measuring of frequency was repeated six times and the average used as the filled resonant frequency f_g. The temperature as indicated by the thermistors was recorded and the pressure was measured with a U-tube manometer. For the cylindrical cavity, the cavity temperature was used as the gas temperature; however, it was found that with the larger volume of the Fabry–Perot structure the gas temperature could be significantly different from the rod temperature. It was necessary in this case to use a thermistor bead attached to the glass bell jar and extending a short distance into the air space to measure the gas temperature. After frequency, pressure, and temperature measurements had been made, the bell jar was evacuated and again it was necessary to allow a few minutes for the system to reach thermal equilibrium.

The frequency was again set and measured six times and the cavity temperature recorded. The measured evacuated resonant frequency f_e' was corrected for the change in temperature of the cavity giving, for the cylindrical cavity,

$$f_e=f_e'+\alpha(T_g-T_e), \tag{4}$$

where α is the temperature coefficient of frequency of the cavity, T_g and T_e are the filled and the evacuated temperatures, respectively.

The corresponding equation for the Fabry–Perot resonator can be obtained from Eq. (3) and written as,

$$f_e=f_e'+\alpha(a\Delta T_m+b\Delta T_B+C\Delta T_t). \tag{5}$$

The index of refraction is then

$$(n-1)=(f_e-f_g)/f_g. \tag{6}$$

All of the measurements were made at a mean temperature of 20°C and at pressures ranging from 600 to 760 Torr, and it was therefore necessary to use an extrapolation equation to arrive at the refractive index at standard conditions, namely 0°C and 760 Torr. For the accuracy of the present measurement, it is sufficient to assume that the refractivity of a gas is proportional to its density, and with the use of a gas law the density can be expressed as a function of the temperature and pressure. Various gas laws were considered, and the ones which were used were chosen on the basis of their agreement with gas-density tables in the temperature and pressure range used. These tables, contained in National Bureau of Standards Circular No. 564,[5] were compiled from the work of a number of experiments and represent a good check on the accuracy of a gas law.

The ideal-gas law was used for helium and hydrogen and the extrapolation equation is then

$$N_{T,P}=(n_{T,P}-1)\times 10^6=N_{0,760}273.15P/760T, \tag{7}$$

where $N_{T,P}$ is the refractivity at T°K and P Torr, and $N_{0,760}$ is the refractivity at standard conditions.

Barrell and Sears[6] arrived at an empirical-extrapolation equation from refractive-index measurements on dry CO_2-free air at optical wavelengths. These results give,

$$N_{T,P}=N_{0,760}P[1+(1.049-0.0157t)\times 10^{-6}P]/[760.606(1+0.03661t)], \tag{8}$$

where t is in degrees Celsius and P in Torr. This equation was used for nitrogen, oxygen, and argon as well as for dry CO_2-free air.

Berthelots equation of state for small pressures was chosen for carbon dioxide and is,

$$Pv=RT\{1+9PT_c(128P_cT)^{-1}[1-(6T_c^2/T^2)]\} \tag{9}$$

or

$$Pv=RT(1-B_TP), \tag{10}$$

where

$$B_T=9T_c(128P_cT)^{-1}[(6T_c^2/T^2)-1]. \tag{11}$$

In the above equations P is the gas pressure, v the molar volume, T the absolute temperature, T_c and P_c are the critical temperatures and pressures which for CO_2 have the values

$$T_c=304.20\text{°K} \quad \text{and} \quad P_c=55\ 366 \text{ Torr}.$$

From Eq. (10) the density is given by

$$d=1/v=(P/RT)(1-B_TP)^{-1}. \tag{12}$$

[5] J. Hilsenrath *et al.*, Natl. Bur. Std. (U. S.) Circ. 564 (1955).
[6] H. Barrell and J. E. Sears, Phil. Trans. Roy. Soc. A238, 2 (1940).

Since B_TP is small compared to unity, this may be approximated by

$$d=(P/RT)(1+B_TP), \tag{13}$$

and since the refractivity is proportional to density the extrapolation equation becomes

$$N_{T,P}=N_{0,760}P(1+B_TP)T_0/[P_0(1+B_0P_0)T], \tag{14}$$

where B_T and B_0 are the values of B at temperatures T°K and 0°C, $T_0=273.15$°K, $P_0=760$ Torr. The extrapolated value of refractivity using Eq. (14) are in very good agreement with those using tables of gas density. In the pressure range from 500 to 760 Torr and temperatures from 0° to 25°C they differ by only 2 parts in 10^8.

One additional correction of $+0.30\times10^{-6}$ is applied to allow for the increase in the dimensions of the cavities when the bell jar is evacuated. This is referred to in the literature as the hydrostatic correction and is due to the expansion of the metal when the pressure of the gas is removed. This change in the length for both cavities is

$$\delta l=Pl(1-2\nu)/E, \tag{15}$$

and the change in the radius of the cylindrical cavity is

$$\delta r=-P(1+\nu)(1-2\nu)r/E, \tag{16}$$

where P is the gas pressure, l the length of the cavity, ν is Poisson's ratio, E is Young's modulus, which for invar have the values $\nu=0.290$, $E=21.4\times10^6$ psi, and r is the radius of the cylindrical cavity. The derivation of these equations is discussed in Appendix B.

ACCURACY OF RESULTS

One possible source of error, which has been eliminated in the present measurement, is that due to a phase shift in the tee when the frequency is changed. The signal in the detector arm is the vector sum of reflected waves from the load and cavity arms plus a signal coupled directly from input to output, and the minimum detector signal occurs when this vector sum is a minimum. The phase relationship of these signals is determined by the guide wavelength and the physical distance of the reflection from the junction. When the bell jar is evacuated and the frequency is changed, this phase relationship is altered due to the change in the guide wavelength and the minimum output signal could occur at a different setting of the cavity than for the filled condition. This phase shift causes an error in the frequency-difference measurement and therefore an error in n. To eliminate this source of error, we have sealed off the hybrid tee from the air surrounding it and removed the seal which had been across the connection between the cavity and the tee. This is accomplished as shown in Fig. 4, by using a 0.005-in. Teflon window across the input and output-waveguide flanges and O-rings at the load and cavity

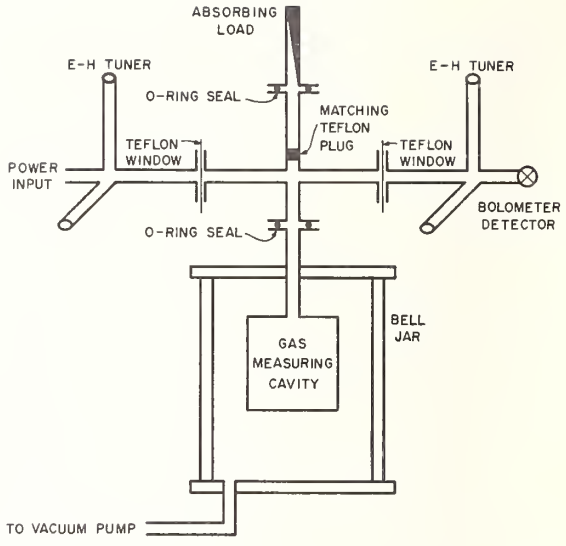


FIG. 4. Hybrid tee arrangement for eliminating phase shift.

flanges. A half-wavelength Teflon plug slightly larger than the waveguide to insure a tight fit is used in place of the tuner in the load arm. It can be moved in the guide relative to the load to accomplish the same tuning effect as the slide-screw tuner. Two of the corners of the plug are cut away slightly to allow for passage of gas past the plug. In this way the tee as well as the cavity is filled with the gas and is evacuated, and the wavelength in the tee remains constant. This is seen from the relationships

$$\frac{1}{\lambda_e^2}=\left(\frac{f_e}{c}\right)^2-\frac{1}{\lambda_c^2}, \tag{17}$$

$$\frac{1}{\lambda_g^2}=\left(\frac{nf_g}{c}\right)^2-\frac{1}{\lambda_c^2}. \tag{18}$$

But from Eq. (1),

$$nf_g=f_e,$$

and therefore,

$$\lambda_e=\lambda_g. \tag{19}$$

In the above equations, λ_e and λ_g are respectively the wavelengths in the tee when evacuated and when filled with the gas of refractive index n, and λ_c is the cutoff wavelength of the guide.

TABLE I. Impurities in each of gas samples in parts per million by volume.

Gas	Impurity in ppm by volume			
	Nitrogen	Oxygen	Carbon dioxide	Water vapor
Argon	10	10		6
Carbon dioxide	40	10		5
Helium	10	5		1
Hydrogen	25		5	1
Nitrogen		50		14
Oxygen	100			6

TABLE II. Error in $(n-1)\times10^6$ due to each contributing factor.

Gas	Both cavities			Cylindrical cavity		Fabry–Perot cavity		rms total 2σ limits	
	Pressure	Impurities	Hydrostatic	Temp.	Random error $2\sigma_N$	Temp.	Random error $2\sigma_N$	Cylindrical	Fabry–Perot
Air	±0.02		±0.03	±0.02	±0.02	±0.10	±0.08	±0.05	±0.13
Argon	±0.02	±0.02	±0.03	±0.02	±0.02	±0.10	±0.04	±0.05	±0.13
Carbon dioxide	±0.03	±0.01	±0.03	±0.04	±0.06	±0.16	±0.16	±0.08	±0.23
Helium	±0.002	±0.01	±0.03	±0.003	±0.04	±0.01	±0.04	±0.05	±0.05
Hydrogen	±0.01	±0.01	±0.03	±0.01	±0.04	±0.04	±0.06	±0.05	±0.07
Nitrogen	±0.02	±0.05	±0.03	±0.02	±0.02	±0.10	±0.02	±0.07	±0.12
Oxygen	±0.02	±0.02	±0.03	±0.02	±0.02	±0.08	±0.06	±0.05	±0.11

The accuracy of the measurement is dependent upon how accurately the frequency difference, the gas temperature and pressure, and the cavity temperature change can be measured, and therefore, every effort was made to reduce the errors in these quantities. By measuring the frequency with reference to a highly stable source and taking the average of six readings, the frequency can be determined to within 2 parts in 10^8.

A precision platinum resistance thermometer was used as the reference for all temperature measurements. It was used to calibrate a precision mercury-in-glass thermometer to within ±0.002°C, and this thermometer was in turn used to calibrate the thermistors at 0.1°C intervals. The thermistor calibrations were periodically checked to allow for any change in their characteristics. The gas-temperature error for the cylindrical cavity is estimated at 0.02°C, while that for the Fabry–Perot is no more than 0.10°C. The larger error for the Fabry–Perot is due to the larger volume of gas in this cavity.

A precision U-tube manometer was used to measure the gas pressure. The scale was calibrated by the Length Standards Section of the National Bureau of Standards to within 0.005 mm and each reading was corrected to give the pressure in terms of the density of mercury at 0°C and standard gravity. The gravity correction was arrived at by using the acceleration of gravity as measured in our laboratory by the U. S. Geological Survey. With these corrections the gas pressure is accurate to within 0.05 Torr.

Another source of error is the uncertainty in the constants used in computing the hydrostatic correction. If a 10% error in these constants is assumed, the corresponding error in refractive index is 3 parts in 10^8.

The gases used were the highest purity available in large enough quantities for this use, and the maximum impurities as stated by the manufacturers are listed in Table I.

Possible sources of systematic error were checked by using different pieces of equipment or making changes in the arrangement of the apparatus. Three different hybrid tees were used and measurements were made with and without the tuning elements on the tee. Two different manometer scales were used, and while most measurements were made at a pressure of 760 Torr, there were also some made in the range from 400–800 Torr. As checks for possible mechanical stress on the cavity, the apparatus was inverted, and readings were also made with the tuning plunger partially inserted and fully withdrawn. In none of these tests were there a measurable systematic change in the measured index of refraction. The use of the two refractometers is also an effective test for systematic error, and the close agreement between the two illustrate that the error should be within the values quoted.

Table II gives the error in $(n-1)\times10^6$ for each of the factors involved. Since these were arrived at by using the estimated maximum error in the measured quantities, the resultant errors in the refractivity is assumed to be equivalent to 2σ limits. To arrive at an estimate of the total error, we have taken the rms sum of these errors with twice the standard deviation of the mean value, $2\sigma_{\bar{N}}$, where $\sigma_{\bar{N}}$ is the standard deviation of the mean value of at least 15 different determinations. This rms sum is then a 2σ limit of the estimated total error, and it is this value which is quoted in the results.

RESULTS

The results for the gases measured are shown in Table III, along with some values as quoted by previous authors. Those values quoted by Maryott and Buckley[7] were arrived at by taking the mean of several measurements made by different authors at frequencies ranging from rf to optical. Their values are given in terms of $(\epsilon-1)$ at 20°C and 760 Torr and have been converted to $(n-1)$ by using the relationship

$$n-1=(\mu\epsilon)^{\frac{1}{2}}-1$$

and extrapolated to 0°C using the appropriate gas law.

The limits of error given by them are the average deviation of the individual values used in computing the mean value. This is about the same as $2\sigma_{\bar{N}}$ when $\sigma_{\bar{N}}$ is the standard deviation of the mean. The errors given by Essen and Froome are the rms sum of the estimated maximum systematic errors together with the standard

[7] A. M. Maryott and F. Buckley, Natl. Bur. Std. (U. S.) Circ. 537 (1953).

deviation of one observation, and as such are similar to 2σ limits. For a more extensive comparison with the results of individual experimenters the reader may wish to consult Essen and Froome[8] or Boudouris.[9]

While the present measurement is the only one at this frequency, there is no reason to expect any dispersive effects for any of the gases except oxygen. The agreement between the two cavities and with previous measurements for air, nitrogen, argon, and hydrogen within the stated experimental error help to verify the stated accuracy of the results. The difference in the values for oxygen was to be expected as a result of the oxygen absorption at 60 GHz, and this is also illustrated by Essen and Froome's work done at 9 [10] and 72 GHz.[11]

There is a rather large difference in the results for helium when compared with the values quoted by Mayrott and Buckley and Essen and Froome, although there have been measurements made with helium that are in close agreement with ours.[12,13] There could be various reasons for this, one of which could be different impurity levels in the gases used. With its very low refractivity, small traces of water vapor or other gases could raise the measured value appreciably. The fact that this is the only gas where there is an appreciable disagreement would tend to support this.

CONCLUSIONS

The close agreement between the two independent refractometers, the cylindrical cavity and the Fabry–Perot, as well as the agreement with other measurements demonstrates the accuracy of this microwave refractometer. The results which we obtained should increase the accuracy of measurements performed in this frequency range when a precise knowledge of the refractive index is needed. The Fabry–Perot refractometer is not difficult to build and offers many advantages when one requires a continuous monitor.

There is a need for further work to determine the refractive index of water vapor and this work is planned as time will permit.

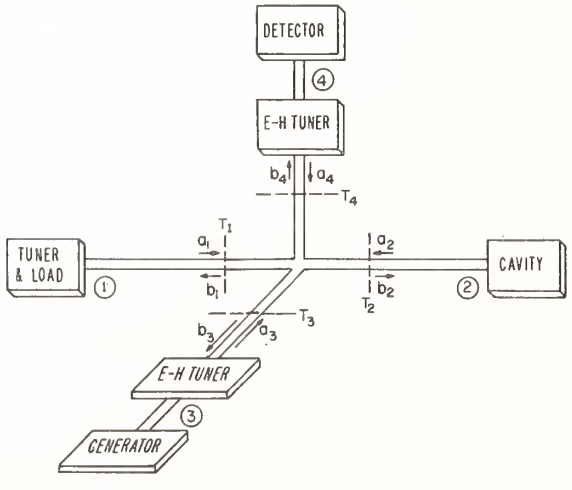


FIG. 5. Schematic diagram of hybrid tee bridge circuit.

TABLE III. Comparison of results.
Index of refraction of gases at 0°C, 760 Torr.

Author	Cavity type	$(n-1)\times10^6$	Frequency	Date
	Dry CO_2-free air			
Maryott–Buckley		288.1 ±0.2		
Essen		288.10±0.10	9 GHz	1953
Essen–Froome		288.15±0.10	24 GHz	1951
Froome		287.66±0.11	72 GHz	1955
Newell–Baird	Cylindrical	288.13±0.05	48 GHz	1964
Newell–Baird	Fabry–Perot	288.15±0.13	48 GHz	1964
	Argon			
Maryott–Buckley		277.6 ±0.2		
Essen–Froome		277.8 ±0.2	24 GHz	1951
Froome		277.7 ±0.2	72 GHz	1955
Newell–Baird	Cylindrical	277.48±0.05	48 GHz	1964
Newell–Baird	Fabry–Perot	277.47±0.13	48 GHz	1964
	Carbon dioxide			
Maryott–Buckley		495.4 ±0.5		
Essen–Froome		494.7 ±1.0[a]	24 GHz	1951
Newell–Baird	Cylondrical	495.16±0.08	48 GHz	1964
Newell–Baird	Fabry–Perot	495.26±0.23	48 GHz	1964
	Helium			
Maryott–Buckley		34.9 ±0.2		
Essen		35.0 ±0.2	9 GHz	1953
Newell–Baird	Cylindrical	34.51±0.05	48 GHz	1964
Newell–Baird	Fabry–Perot	34.43±0.05	48 GHz	1964
	Hydrogen			
Maryott–Buckley		136.2 ±0.2		
Essen		136.0 ±0.2	9 GHz	1953
Newell–Baird	Cylindrical	135.77±0.05	48 GHz	1964
Newell–Baird	Fabry–Perot	135.77±0.07	48 GHz	1964
	Nitrogen			
Maryott–Buckley		294.1 ±0.2		
Essen		294.1 ±0.1	9 GHz	1953
Essen–Froome		294.1 ±0.1	24 GHz	1951
Froome		294.05±0.11	72 GHz	1955
Newell–Baird	Cylindrical	293.81±0.07	48 GHz	1964
Newell–Baird	Fabry–Perot	293.90±0.12	48 GHz	1964
	Oxygen			
Maryott–Buckley		266.3 ±0.1		
Essen		266.2 ±0.2	9 GHz	1953
Essen–Froome		266.4 ±0.2	24 GHz	1951
Froome		263.9 ±0.2	72 GHz	1955
Newell–Baird	Cylindrical	266.95±0.05	48 GHz	1964
Newell–Baird	Fabry–Perot	267.06±0.11	48 GHz	1964

[a] This value has been corrected by using Berthelot's gas law rather than the ideal gas law.

ACKNOWLEDGMENT

The authors wish to express their appreciation to Dr. J. M. Richardson, under whose direction the work

[8] Reference 3, p. 868.
[9] G. Boudouris, J. Res. Nat. Bur. Std. **D6**, 631 (1963).
[10] L. Essen, Proc. Phys. Soc. (London) **B66**, 189 (1953).
[11] K. D. Froome, Proc. Phys. Soc. (London) **B68**, 833 (1955).
[12] J. G. Jelatis, J. Appl. Phys. **19**, 419 (1948).
[13] C. Cuthbertson and M. Cuthbertson, Proc. Roy. Soc. (London) **A135**, 40 (1932).

was initiated, and who offered many helpful suggestions during the preparation of the manuscript.

APPENDIX A. ANALYSIS OF HYBRID TEE CIRCUIT

A schematic diagram of the hybrid tee circuit is shown in Fig. 5. If we adopt reference planes as indicated by the T_j's in the figure, then the scattering matrix of the system is given as

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{21} & s_{22} & s_{23} & s_{24} \\ s_{31} & s_{32} & s_{33} & s_{34} \\ s_{41} & s_{42} & s_{43} & s_{44} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \hat{S} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}, \tag{20}$$

where a_i is the magnitude of the input signal and b_i is the magnitude of the outgoing signal at T_i. The s_{ij}'s are the scattering coefficients relating the input signal at T_j to the output at T_i. For the system as used,

$$a_1 = b_1\Gamma_1,$$
$$a_2 = b_2\Gamma_2,$$
$$a_3 = b_3\Gamma_3 + b_g,$$
$$a_4 = b_4\Gamma_4,$$

where the Γ's are the reflection coefficients of the various arms and b_g is the incident wave which the generator would emit into a nonreflecting load. With these substitutions Eq. (20) becomes

$$\begin{bmatrix} b_1 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \hat{S} \begin{bmatrix} b_1\Gamma_1 \\ b_1\Gamma_1 \\ b_2\Gamma_2 \\ b_3\Gamma_3 + b_g \\ b_4\Gamma_4 \end{bmatrix}. \tag{21}$$

If we write out the equations implied in (21) and solve for the detected signal b_4, we obtain the following:

$$b_4 = b_g \frac{\begin{vmatrix} s_{11}\Gamma_1 - 1 & s_{12}\Gamma_2 & s_{13} & -s_{13} \\ s_{21}\Gamma_1 & s_{22}\Gamma_2 - 1 & s_{23}\Gamma_3 & -s_{23} \\ s_{31}\Gamma_1 & s_{32}\Gamma_2 & s_{33}\Gamma_3 - 1 & -s_{33} \\ s_{41}\Gamma_1 & s_{42}\Gamma_2 & s_{43}\Gamma_3 & -s_{43} \end{vmatrix}}{\begin{vmatrix} s_{11}\Gamma_1 - 1 & s_{12}\Gamma_2 & s_{13}\Gamma_3 & s_{14}\Gamma_4 \\ s_{21}\Gamma_1 & s_{22}\Gamma_2 - 1 & s_{23}\Gamma_3 & s_{24}\Gamma_4 \\ s_{31}\Gamma_1 & s_{32}\Gamma_2 & s_{33}\Gamma_3 - 1 & s_{34}\Gamma_4 \\ s_{41}\Gamma_1 & s_{42}\Gamma_2 & s_{43}\Gamma_3 & s_{44}\Gamma_4 - 1 \end{vmatrix}}. \tag{22}$$

The numerator can be reduced substantially and (22) can be written in a simplified form as

$$\frac{b_4}{b_g} = \frac{\Gamma_1\Gamma_2 D - \Gamma_1 m_{22} - \Gamma_2 m_{11} + s_{43}}{\text{Den.}}, \tag{23}$$

where

$$D = \begin{vmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{41} & s_{42} & s_{43} \end{vmatrix}; \quad m_{11} = \begin{vmatrix} s_{22} & s_{23} \\ s_{42} & s_{43} \end{vmatrix}; \quad m_{22} = \begin{vmatrix} s_{11} & s_{13} \\ s_{41} & s_{43} \end{vmatrix},$$

and Den. means the denominator of Eq. (22). The significant aspect of Eq. (23) is that the numerator does not depend upon Γ_3 and Γ_4. Setting the numerator equal to 0 the condition for the null is obtained as

$$\Gamma_1 = (\Gamma_2 m_{11} - s_{43})/(\Gamma_2 D - m_{22})$$

or

$$\Gamma_2 = (\Gamma_1 m_{22} - s_{43})/(\Gamma_1 D - m_{11}). \tag{24}$$

Each of these equations is a linear fractional transformation and the solutions are unique unless $s_{43}D - m_{11}m_{22} = 0$ in which case Γ_1 and Γ_2 are either constant or the solutions are meaningless. If, as indicated in the figure, T_3 and T_4 are chosen between the junction and the E–H tuners, then the only effect a change in the E–H tuners has on the system is to alter Γ_3 and Γ_4. All of the scattering coefficients remain unchanged. The conclusion is that a null is obtained when Eq. (24) is satisfied, and the frequency at which the null occurs will be independent of the condition of the E–H tuners. Therefore, these tuners may be used to improve the sensitivity without fear of introducing any significant errors into the frequency determinations.

For a truly symmetrical tee, $s_{34} = s_{43} = 0$, $s_{11} = s_{22}$, $s_{14} = s_{24}$, and $s_{13} = -s_{23}$. Under these conditions $\Gamma_1 = \Gamma_2$, and for a high-quality tee one can then expect $\Gamma_1 \approx \Gamma_2$ when a null is obtained.

Since the measurement of refractive indices involves a change of frequency, there is a possibility of small errors arising from changes in the values of the scattering coefficients with frequency. A precise knowledge of how the scattering coefficients vary with frequency would be difficult if not impossible to obtain; however, the following observations can be made. First, as discussed previously, the process of evacuating and filling the tee with the gas being measured along with the cavity keeps the wavelength in the junction constant. This eliminates any phase change in the scattering coefficients from this source. Second, the frequency change is relatively small, ranging from about 0.04% of the resonant frequency for CO_2 to less than 0.004% for He. Experimental tests were performed on the load used on Arm 1 of the bridge and on the tee itself. Variations in the amplitude and phase of the reflection coefficients were checked over a frequency range 20 times as great as the largest frequency shift encountered in practice. The data exhibited some small, slowly varying frequency effects, but the magnitudes and phases of the reflection coefficients appeared to be constant in the range of interest. This implies that to a good approximation the scattering coefficients may be regarded as constants over the limited frequency range used. Finally, since the cavity Q is approximately 30 000, the phase and amplitude of the cavity reflection coefficient change extremely fast at frequencies near resonance. Hence when the frequency is being set, Γ_2 changes rapidly until Eq. (24) is satisfied and the other circuit parameters remain essentially constant. This analysis plus the many experimental tests made on the

system lead us to the conclusion that there is no significant error inherent in the microwave bridge circuit as used in this experiment.

APPENDIX B. DERIVATION OF HYDROSTATIC PRESSURE CORRECTION

The basic equations used in this derivation can be obtained from any good text on elasticity, for example Wang,[14] or Timoshenko and Goodier.[15] In cylindrical coordinates these equations are as follows.

Stress equations:

$$\frac{\partial\sigma_r}{\partial r}+\frac{\sigma_r-\sigma_\theta}{r}=0;\quad \frac{\partial\sigma_z}{\partial z}=0. \tag{25}$$

Stress–strain relations:

$$\begin{aligned}\sigma_r-\nu(\sigma_\theta+\sigma_z)&=E\epsilon_r,\\ \sigma_\theta-\nu(\sigma_r+\sigma_z)&=E\epsilon_\theta,\\ \sigma_z-\nu(\sigma_r+\sigma_\theta)&=E\epsilon_z.\end{aligned} \tag{26}$$

Strain–displacement relations:

$$\epsilon_r=\frac{\partial u}{\partial r};\quad \epsilon_\theta=\frac{u}{r};\quad \epsilon_z=\frac{\partial w}{\partial z}. \tag{27}$$

In the above equations σ_i is the stress component in the ith direction, ϵ_i the strain component in the ith direction, u and w are the displacements of a point in the r and z directions, respectively, E is Young's modulus, and ν is Poisson's ratio.

For the conventional cavity a long hollow cylinder was assumed with the stress on a plane $z=$constant equal to the stress on the ends of the actual cylinder. If it is assumed that $w\equiv 0$, $\epsilon_z=0$ and from (26),

$$\sigma_z=\nu(\sigma_r+\sigma_\theta), \tag{28}$$

and

$$\begin{aligned}(1-\nu^2)\sigma_r-\nu(1+\nu)\sigma_\theta&=E\epsilon_r,\\ (1-\nu^2)\sigma_\theta-\nu(1+\nu)\sigma_r&=E\epsilon_\theta,\end{aligned} \tag{29}$$

or

$$\sigma_r-\sigma_\theta=[E/(1+\nu)](\epsilon_r-\epsilon_\theta). \tag{30}$$

[14] C. Wang, *Applied Elasticity* (McGraw-Hill Book Company, Inc., New York, 1953), pp. 1–54.

[15] S. Timoshenko and J. N. Goodier, *Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York, 1951), pp. 1–60.

The σ_θ can be eliminated from (29) to yield

$$\sigma_r=[E/(1+\nu)(1-2\nu)][(1-\nu)\epsilon_r+\nu\epsilon_\theta]. \tag{31}$$

Combining Eqs. (25), (30), and (31) results in the following differential equation for u:

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}-\frac{1}{r^2}u=0. \tag{32}$$

The solution of this equation is

$$u=Ar+(B/r), \tag{33}$$

and by (27)

$$\epsilon_r=A-(B/r^2);\quad \epsilon_\theta=A+(B/r^2). \tag{34}$$

Eq. (31) can then be written as

$$\sigma_r=\frac{E}{1+\nu}\left[\frac{A}{1-2\nu}-\frac{B}{r^2}\right].$$

But $\sigma_r=$constant$=-P$, where P is the hydrostatic pressure, hence $B=0$ and

$$A=-(P/E)(1+\nu)(1-2\nu). \tag{35}$$

In addition, from (34) $\epsilon_r=\epsilon_\theta$ and by (30) $\sigma_r=\sigma_\theta$. Therefore, according to (26) with $\epsilon_z=0$,

$$\sigma_z=\nu(\sigma_r+\sigma_\theta)=-2\nu P; \tag{36}$$

thus, a pressure of $2\nu P$ must be applied to the ends of the cylinder to keep the displacement zero. Therefore, when P is applied the displacement corresponds to an effective pressure of $P(1-2\nu)$ and the change in the length of the cavity is

$$\delta l=(Pl/E)(1-2\nu), \tag{37}$$

where l is the length of the conventional cavity or the length of the rods separating the two mirrors when the equation is used with the Fabry–Perot resonator.

The change in the radius of the cylindrical cavity is given by Eq. (33) as

$$\delta r=-(P/E)(1+\nu)(1-2\nu)r, \tag{38}$$

where r is the radius of the cavity.

Reprinted from Journal of Applied Physics, Vol. 36, No. 12, 3751–3759, December 1965

A Radio-Frequency Permittimeter*

R. C. POWELL†, MEMBER, IRE, AND A. L. RASMUSSEN†

Introduction

THE measurement of low-impedance materials, such as conductors, semiconductors, electrolytes, and high-permittivity materials has always been difficult since errors due to electrode impedance, contact potentials, interaction of electrodes with the materials, series inductance and connection impedance can cause errors many times the actual quantity being measured. Methods that have been used at low frequencies involve four-terminal bridges [1], and double transformers using an electrodeless ring of the material as a coupling loop [2]. Such measurements at microwave frequencies have been made in a circular electric

* Received by the PGI, June 23, 1960. Presented at the 1960 Conference on Standards and Electronic Measurements as paper 3-5. The work reported here was partially supported by the Dept. of the Navy under a Bureau of Ships contract.

† National Bureau of Standards, Boulder, Colo.

Reprinted from IRE TRANSACTIONS
ON INSTRUMENTATION
Volume I-9, Number 2, September, 1960

field [3] or by using samples large compared to a wavelength. None of these methods are useful at radio frequencies where many materials exhibit interesting and useful properties.

The RF permittimeter (so-called because of its analogy to the RF permeameter [4]) also utilizes a circular electric field in a ring-shaped sample to eliminate electrode and series inductance problems. The circular electrical field is created in a transformer yoke of appropriate symmetry. This instrument is a two-terminal coaxial system designed for making measurements on conventional HF impedance measuring equipment.

Since the yoke of the transformer is made of magnetic materials, the principal limit of the range of the instrument is the maximum initial permeability of yoke materials available. This falls from about 5000 at 10 kc to about 50 at 100 Mc and determines the effective impedance of the single turn sample and hence the sensitivity of the instrument. Using bridges with 0.1 per cent precision and 1 per cent error, the maximum resistivity which can be measured to 1 per cent error ranges from about 10^{-2} ohm meters at 10 kc to about 1 ohm meter at 100 Mc while the minimum relative permittivity which can be measured to the same accuracy ranges from about 10^{8} at 10 kc down to about 10^{2} at 100 Mc. These accuracy figures were obtained empirically by measuring known impedance rings. By changing the number of turns on the primary winding, the permittimeter can be matched to many impedance measuring instruments, but the range of materials that can be measured can be changed only by small amounts.

The instrument is analyzed as a T network and calibrated using the cases of no secondary, a low-resistance copper ring secondary, and a known resistance secondary consisting of a ring of fine resistance wire such that skin effect is eliminated. The specimen to be measured consists of a solid ring of material or a liquid in a ring-shaped container. Since the magnetic path must be broken and reformed for insertion of solid rings, precision of greater than 1 per cent is difficult to obtain.

Measurements made to date have shown values of permittivity of many ferrites to be much higher than those obtained by electrode methods, explaining resonance effects at frequencies lower than those previously calculated. Measurements of strong electrolytic solutions have shown that the conductivity has a strong out-of-phase component which can be either leading or lagging, depending on the nature of the solution, and thus contributes either a positive or negative component to the permittivity. Good agreement with extrapolated values for the in-phase component of the conductivity of these solutions as well as several metals was obtained.

The permittimeter shown in Figs. 1 and 2 was made from a solid piece of high-permeability ferrite using an ultrasonic cutter; the plate and mating surfaces were then ground and lapped. The optimum polish on the surfaces is somewhat rougher than a mirror finish. It was found that too fine a surface caused poor reproducibility, probably because of microscopic magnetic coupling across the boundary.

The following analysis produces the equations necessary to determine the impedance around the ring of material and also the complex permittivity or dielectric constant. The complex conductivity σ^* can be obtained from the complex permittivity K^*K_0 by

$$\sigma^* = j\omega K^* K_0. \tag{1}$$

It should be realized that, while conductivity and permittivity are two different phenomena, it is difficult to separate them experimentally. The measured value generally is the sum of the two effects; hence, such a conversion is acceptable.

Fig. 1—An RF permittimeter with standard ring, shorting ring and various rings of materials to be measured.

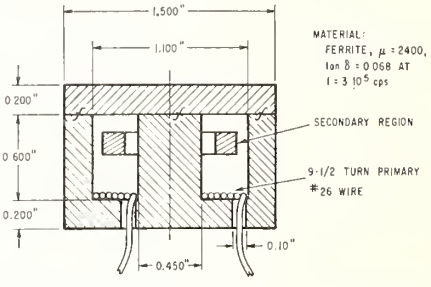


Fig. 2—Section drawing of an RF permittimeter.

Analysis of the Permittimeter

At a single frequency, the permittimeter with a secondary impedance can be represented by an equivalent loaded T network [5] as shown in Fig. 3.

Using the symbol definitions given in Table I, the impedance Z_L can be written in terms of this circuit as

$$Z_L = \frac{Z_2 Z_3 - (Z_{in} - Z_1)(Z_2 + Z_3)}{Z_{in} - Z_1 - Z_2}. \tag{2}$$

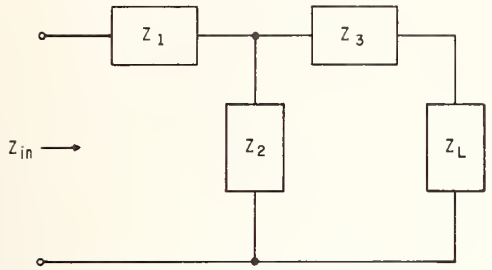


Fig. 3—Equivalent Tee network.

TABLE I

DEFINITION OF SYMBOLS

Capacitance	
C^*	The complex self capacitance of the secondary region in farads.
Dimension	
a	Inside diameter of a ring in meters.
b	Outside diameter of a ring in meters.
c	$(b-a)/2$ in meters
D	Diameter of a ring equal to $(a+b)/2$ in meters.
d	Diameter of the section of a ring of circular section in meters.
h	Height of a ring of rectangular section in meters.
Greek Symbol	
δ	Effective depth of penetration of the field into the material; equal to the skin depth in meters $=(\lambda_0/\mu\pi\sigma Z_0)^{1/2}$
K^*	The complex relative permittivity of the material.
K_0	The permittivity of free space, $8.85\cdot10^{-12}$ farads per meter.
λ_0	Free-space wavelength
μ^*	The complex relative permeability of the material.
μ_0	The permeability of free space, $4\pi\cdot10^{-7}$ henries per meter.
σ^*	The complex conductivity equal to $j\omega K_0K^*$ in mhos per meter.
ω	The angular frequency $2\pi f$ in radians per second.
Impedance	
Z_1, Z_2, Z_3	The impedances of the equivalent T network representing the permittimeter in ohms.
Z_L	The difference between the impedance of the secondary and the impedance of the shorting ring, Z_S, in ohms.
Z_{in}	The impedance looking into the terminals of the permittimeter in ohms.
Z_A	The impedance of a ring of known impedance used as a calibration standard in ohms.
$Z_{in\,A}$	Z_{in} when the ring of known impedance is inserted into the secondary region.
Z_0	The characteristic impedance of free space $=(\mu_0/K_0)^{1/2}$ $=376.7$ ohms.
$Z_{in\,0}$	Z_{in} when no material is inserted into the secondary region in ohms.
Z_S	The impedance of the shorting ring in ohms.
$Z_{in\,S}$	Z_{in} when the shorting ring is inserted into the secondary region in ohms.
Z_X	The impedance of the ring of material to be measured in ohms.
$Z_{in\,X}$	Z_{in} when the ring of material to be measured is inserted into the secondary region.
Inductance	
L_A	Inductive part of Z_A determined from dimensions and expressed in henries.
L^*	The complex self inductance of the secondary region in henries.
Resistance	
R_A	Resistive part of Z_A measured directly in ohms.

If no material is inserted into the secondary region, Z_L can be considered to have infinite impedance and

$$Z_2 = Z_{in\,0} - Z_1. \tag{3}$$

If the shorting ring is inserted into the secondary region, Z_L has exactly zero impedance and

$$Z_3 = \frac{(Z_{in\,S} - Z_1)(Z_{in\,0} - Z_1)}{Z_{in\,0} - Z_{in\,S}}. \tag{4}$$

Therefore, (2) can be rewritten as

$$Z_L = \frac{(Z_{in\,0} - Z_1)^2}{(Z_{in\,0} - Z_{in\,S})} \quad \frac{(Z_{in\,S} - Z_{in})}{(Z_{in} - Z_{in\,0})} \tag{5}$$

and so

$$\frac{Z_X - Z_S}{Z_A - Z_S} = \left(\frac{Z_{in\,0} - Z_{in\,A}}{Z_{in\,A} - Z_{in\,S}}\right)\left(\frac{Z_{in\,X} - Z_{in\,S}}{Z_{in\,0} - Z_{in\,X}}\right) \tag{6}$$

which is the equation used to determine Z_X and hence K^* of the material measured in terms of Z_A a known impedance, Z_S an effective short circuit, and the indicated measurements.

The impedance Z_L is described by

$$Z_L = j\left(\omega L^* - \frac{1}{\omega C^*}\right)_{\text{secondary}} - j\left(\omega L^* - \frac{1}{\omega C^*}\right)_{\text{shorting ring}} \tag{7}$$

so the complex self inductance and capacitance of the shorting ring, the known impedance ring and the ring of material to be measured must be calculated to interpret the results.

For the case where the secondary is a circular ring of circular section (Fig. 4) with diameters D and d, respectively

$$C^* = \frac{K^*K_0}{2}\left[D - (D^2 - d^2)^{1/2}\right] \tag{8}$$

and

$$L^* \sim \frac{\mu_0}{2} D\left(\frac{\mu^*}{4} - 2 + Ln\,\frac{8D}{d}\right) \tag{9}$$

if the current penetrates the entire secondary, as is assumed in the case of the known impedance ring and the ring of material to be measured. If the current penetrates a relatively small depth δ as in the case of the

shorting ring

$$C^* = \frac{K^*K_0}{2}\left\{\left[D^2 - (d - 2\delta)^2\right]^{1/2} - (D^2 - d^2)^{1/2}\right\} \quad (10)$$

and

$$L^* \sim \frac{\mu_0}{2} D\left(-2 + Ln\frac{8D}{d}\right). \quad (11)$$

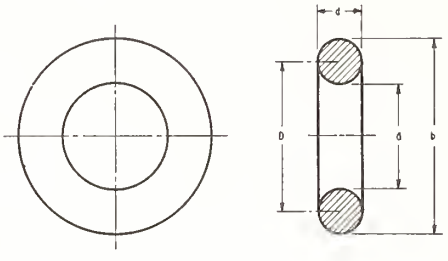


Fig. 4—Dimensions of a ring of circular section.

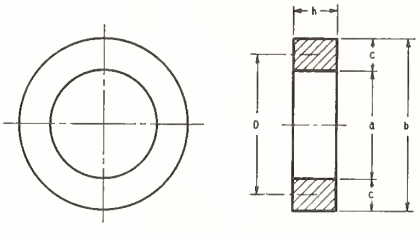


Fig. 5—Dimensions of a ring of rectangular section.

For the case where the secondary is a circular ring of rectangular section (Fig. 5) with an inside diameter a, an outside diameter b, and a height h

$$C^* = \frac{K^*K_0h}{2\pi} Ln\frac{b}{a} \quad (12)$$

and

$$L^* \sim \frac{\mu_0}{2} D\left[\frac{\mu^*\pi(c^2 + h^2)}{24ch} - 2 + Ln\frac{17.9D}{c + h}\right] \quad (13)$$

if the current penetrates the entire secondary as in the case of the known impedance ring and the ring of material to be measured. If the current penetrates only a small depth δ as in the case of the shorting ring

$$C^* = \frac{K^*K_0}{2\pi}\left[hLn\frac{b(a + 2\delta)}{a(b - 2\delta)} + 2\delta Ln\left(\frac{b - 2\delta}{a + 2\delta}\right)\right] \quad (14)$$

and

$$L^* \sim \frac{\mu_0}{2} D\left[-2 + Ln\frac{17.9D}{c + h}\right]. \quad (15)$$

The above expressions for capacity are exact and can be computed directly or obtained by dividing the inductance of a ring of the same geometry in which the magnetic field occupies the position and direction of the electric field by Z_0^2. The expressions for inductance are approximations of sufficient accuracy for this use. More nearly exact expressions and formulas for other configurations are available in the literature [6]–[9]. A sample calculation is given below, and Tables II and III give examples of results. The accuracies for graphite, the films, lead, and molybdenum are lower due either to uncertainties of the dimensions (the metals), or imperfect field penetration (graphite).

Example of a Measurement of the Complex Dielectric Constant of a Ferrite Ring Core

Using the permittimeter shown in Fig. 2, a ring of ferrite of rectangular section as the material to be measured, Z_X, a copper ring of rectangular section for Z_S and a ring of resistance wire of circular section for Z_A, the following data were obtained. All impedance measurements were made on a Maxwell-type bridge circuit.

Ferrite Ring Z_X

$a_X = 1.53 \cdot 10^{-2}$ meters

$b_X = 2.48 \cdot 10^{-2}$ meters

$h_X = 3.82 \cdot 10^{-3}$ meters

$\mu_X^* = 1950 - j146.$

Copper Shorting Ring Z_S

$a_S = 1.43 \cdot 10^{-2}$ meters

$b_S = 2.34 \cdot 10^{-2}$ meters

$h_S = 4.15 \cdot 10^{-3}$ meters

$\sigma_S^* = 5.65 \cdot 10^7 + j0$ mhos per meter (hard drawn copper).

Resistance Wire Ring Z_A

$D_A = 2.22 \cdot 10^{-2}$ meters

$d_A = 6.99 \cdot 10^{-5}$ meters.

Measured Impedances at a Frequency of $3 \cdot 10^5$ cps

$Z_{in\ 0} = (122.7 + j1582)$ ohms

$Z_{in\ S} = (16.49 + j127.6)$ ohms

$Z_{in\ A} = (722.8 + j1043)$ ohms

$Z_{in\ X} = (178.6 + j1636)$ ohms

$R_A = 23.25$ ohms.

TABLE II

RESULTS FOR VARIOUS MATERIALS USING SMALL RF FIELDS AT 24°C

Material		Frequency in Mc	σ' in mhos per meter	σ'' in mhos per meter
Ferrite ring I	$\mu^* = 615 - j63$	0.3	43.7	1.9
	$\mu^* = 582 - j61$	0.6	44.7	4.0
	$\mu^* = 576 - j138$	1.0	45.6	6.0
Ferrite ring II	$\mu^* = 1000 - j28$	0.3	1.36	3.10
	$\mu^* = 1050 - j73$	0.6	2.39	4.42
	$\mu^* = 1050 - j193$	1.0	3.91	6.63
Ferrite ring III	$\mu^* = 3800 - j36$	0.1	17.9	9.6
	$\mu^* = 3510 - j1530$	0.3	22.6	24.3
	$\mu^* = 1850 - j2310$	0.6	31.2	50.0
CsCl solution	5.00 molar	0.3	41.5	−0.4
	2.50 molar	0.3	25.6	0.0
	0.25 molar	0.3	13.4	0.0
	0.625 molar	0.3	7.2	0.0
$BaBr_2$ solution	2.50 molar	0.3	21.7	−0.2
	1.25	0.3	17.2	0.2
	0.625 molar	0.3	9.8	0.1
CdI_2 solution	1.25 molar	0.3	2.92	−0.15
	0.625 molar	0.3	1.92	−0.09
$Pb(NO_3)_2$ solution	1.25 molar	0.3	7.72	−0.63
	0.625 molar	0.3	5.05	−0.19
Evaporated gold	$7.1 \cdot 10^{-5}$ inch thick	0.3	$4.08 \cdot 10^{7}$	$0.50 \cdot 10^{7}$
Evaporated copper	$1.1 \cdot 10^{-5}$ inch thick	0.3	$9.1 \cdot 10^{6}$	$0.1 \cdot 10^{6}$
Evaporated antimony	$3.3 \cdot 10^{-5}$ inch thick	0.3	$2.37 \cdot 10^{5}$	$0.41 \cdot 10^{5}$
Evaporated bismuth	$5.05 \cdot 10^{-4}$ inch thick	0.3	$4.33 \cdot 10^{4}$	$-0.14 \cdot 10^{4}$
Solid graphite	(99.9 per cent purity)	0.3	$8.6 \cdot 10^{4}$	$0.1 \cdot 10^{4}$
Solid lead	(99.9 per cent purity)	0.3	$4.92 \cdot 10^{6}$	$0.02 \cdot 10^{6}$
Solid molybdenum	(99.9 per cent purity)	0.3	$1.61 \cdot 10^{7}$	$0.04 \cdot 10^{7}$

TABLE III

COMPARISON OF RESULTS FOR FERRITE CORES USING SMALL RF FIELDS AT 300 K AND 24°C

μ'	μ''	σ^*† in mhos per meter	σ^*‡ in mhos per meter
970	13	$3.44 \underline{\lvert 50.8°}$	$3.19 \underline{\lvert 47.0°}$
3140	122	$5.74 \underline{\lvert 43.3°}$	$4.74 \underline{\lvert 40.4°}$
3510	1530	$33.3 \underline{\lvert 46.9°}$	$34.5 \underline{\lvert 44.8°}$
3560	810	$9.6 \underline{\lvert 54.9°}$	$9.6 \underline{\lvert 52.4°}$
3880	2100	$63.2 \underline{\lvert 5.0°}$	$39.2 \underline{\lvert 0.8°}$

† From measurements using the permittimeter.

‡ From measurements of capacitance and conductance of a disk with evaporated gold electrodes.

Calculated Quantities

$C_S^* = -j1.08 \cdot 10^{-3}$ farad calculated from (14) where

$\delta_S = (\lambda_0/\pi\sigma_S Z_0)^{1/2} = 1.22 \cdot 10^{-4}$ meters

$L_S^* = 1.96 \cdot 10^{-8}$ henry calculated from (15) where

$\delta_S = 1.22 \cdot 10^{-4}$ meters

$Z_S = j(\omega L - 1/\omega c) = (4.91 \cdot 10^{-4} + j3.70 \cdot 10^{-2})$ ohms

$L_A^* = 8.49 \cdot 10^{-8}$ henry calculated from (11)

$Z_A = (23.25 + j0.16)$ ohms.

$L_X^* = (6.40 \cdot 10^{-6} - j4.77 \cdot 10^{-7})$ henry calculated from (13)

$C_X^* = 2.597 \cdot 10^{-15} K^*$ farad calculated from (12)

$Z_X = (187 - j257)$ ohms calculated from (6)

$C_X^* = 1.618 \cdot 10^{-9} \underline{\lvert -34.6°}$ farads

$K_X^* = 5.14 \cdot 10^{5} - j3.54 \cdot 10^{5}$

or

$\sigma_X^* = 5.90 + j8.57$ mhos per meter.

T Network Parameters

$Z_1 = (1430 \underline{\lvert 85.5°})$ ohms from (5)

$Z_2 = (154 \underline{\lvert 86.0°})$ ohms from (3)

$Z_3 = (137 \underline{\lvert -94.0°})$ ohms from (4).

Acknowledgment

L. A. Steinert calculated the inductance of a ring of rectangular section of which (13) is an approximation. W. A. Pittman made most of the measurements. C. A. Hoer programmed and processed the data on an electronic computer. Dr. P. M. Gruzensky, H. C. Leistner, P. L. London, A. J. Pannone, W. K. Stephenson and H. C. Stump prepared the specimens and ferrite parts.

References

[1] P. A. Miles, W. B. Westphal, and A. von Hippel, "Dielectric spectroscopy of ferromagnetic semiconductors," *Rev. Mod. Phys.*, vol. 29, pp. 279–307; July, 1957.

[2] S. R. Gupto and G. J. Hills, "A precision electrode-less conductance cell for use at audio-frequencies," *J. Sci. Instr.*, vol. 33, pp. 313–314; August, 1956.

[3] Collie, Hasted, and Ritson, "The cavity method of measuring the dielectric constants of polar liquids in the centimeter band," *Proc. Phys. Soc. (London)*, vol. 60, pp. 71–82; January, 1948.

[4] A. L. Rasmussen and A. E. Hess, "R-F permeameter techniques for testing ferrite cores," *Elec. Manufacturing*, vol. 61, pp. 86–91, 308; May, 1958.

[5] W. L. Everitt, "Communication Engineering," McGraw-Hill Book Co., Inc., New York, N. Y.; 1937.

[6] C. Snow, "Formulas for Computing Capacitance and Inductance," Natl. Bur. of Standards Circular 544, U. S. Govt. Printing Office, Washington, D. C.; 1954.

[7] E. B. Rosa and F. W. Grover, "Formulas and Tables for the Calculation of Mutual and Self-Inductance," (Revised), Sci. Papers of the Bur. of Standards No. 169, U. S. Govt. Printing Office, Washington, D. C.; 1948.

[8] F. W. Grover, "Inductance Calculations," D. Van Nostrand Co., Inc., New York, N. Y.; 1946.

[9] F. E. Terman, "Radio Engineering Handbook," McGraw-Hill Book Co., Inc., New York, N. Y.; 1943.

[10] D. E. Bromley and P. T. Good, "Capacitance-coupled cell for electrolytic conductivity measurements," *J. Sci. Instr.*, vol. 36, pp. 326–327; July, 1959.

8.a. **International comparison of dielectric measurements,** H. E. Bussey, J. E. Gray, E. C. Bamberger, E. Rushton, G. Russell, B. W. Petley, and D. Morris, IEEE Trans. Instr. & Meas. IM-13, No. 4, 305-311, (December 1964).

Three materials, i.e., fused silica, glass, and alumina, were selected for comparison based on known or expected homogeneity, isotropy, and stability. Measurements were made by the three Government laboratories (U.K., U.S.A., and Canada), both in the radio-frequency range, using capacitor-type holders either with or without an air gap, and at microwave frequencies, using either cavity-resonance methods or transmission-line impedance methods. The range among the laboratories on the real part of the permittivity is 0.4, 0.8, and 0.2 percent for the three materials, respectively. The agreement on loss tangent is of the order of 0.0001 in many cases, but larger discrepancies exist. An introductory statistical analysis for systematic differences between laboratories is given.

8.b. **Ferrimagnetic resonance measurements using IF substitution techniques,** W.E. Case, R. D. Harrington, and L. B. Schmidt, J. Res. NBS 68C (Eng. & Instr.), No. 4, 255-259 (Oct.-Dec. 1964).

Advantages of using an IF substitution method with a waveguide below cutoff attenuator for obtaining ferrimagnetic resonance line width of polycrystalline materials are discussed. An improvement in the IF system involving phase locking the local oscillator to the signal generator is described. Measurements were obtained which indicate that the above IF attenuator method compares favorably with accurate frequency measuring techniques for plotting entire resonance curves.

8.c. **Calibration of vibrating-sample magnetometers,** W. E. Case and R. D. Harrington, J. Res. NBS 70C (Eng. & Instr.), No. 4, 255-262 (Oct.-Dec. 1966).

An evaluation of two of the most widely accepted methods for calibrating vibrating-sample magnetometers is given. The comparison method uses a material of known magnetization such as pure nickel. In the slope method, the magnetometer is calibrated from the low field linear slope of the magnetization curve of a sample of high permeability.
The primary source of error in the comparison method arises from an uncertainty in the absolute magnetization of nickel and its dependence on evironmental conditions. The study indicated that better accuracy can be expected from the slope method. The use of pure iron in this method was found preferable to high permeability ferrites.

8.d. **Ferromagnetic resonance relaxation, wide spin-wave coverage by ellipsoids,** Allan S. Risley and Howard E. Bussey, J. Appl. Phys. 35, No. 3 (Part 2), 896-897 (March 1964).

The role of degenerate spin-wave coupling in ferromagnetic resonance (FMR) losses has been studied as a function of sample shape. Measurements of linewidth $\triangle H$ and peak absorption intensity s'' both show that FMR loss above the 90°, $\kappa \rightarrow 0$ limit of the spin-wave spectrum is much less than within the spectrum. An investigation of unsymmetrically magnetized spheroids, i.e., with the external field H_0 not along the symmetry axis revealed that there are regions of the spin-wave spectrum which can be reached by more than one sample and that in these regions of overlap the different shapes yield different values of $\triangle H$ and s''. A similar result can be calculated from the usual equation of motion using either Landau-Lifshitz or Bloch-Bloembergen damping. (The asymmetric case has not yet been treated by the spin-wave model.) Despite the strong variations of $\triangle H$ and s'', the gyromagnetic ratio changed less than 1% under all variations of shape and orientation. The variable shape technique gives greater coverage of the spectrum than previous variable frequency work. Another advantage is that the unknown effects of changing frequency on other possible loss mechanisms are avoided.

8.e. **Polycrystalline spin wave theory of ferromagnetic resonance compared with the tilting experiment,** A. S. Risley, E. G. Johnson, Jr., and H. E. Bussey, J. Appl. Phys. 37, No. 2, 656-668 (February 1966).

Schlomann's spin wave theory of polycrystalline ferromagnetic resonance has been tested in detail using the tilting method reported earlier. Among the quantities measured were the linewidth $\triangle H$, and the maximum value of the imaginary part of the effective susceptibility $(\chi_m'')^{\text{eff}}$. The measurement method is equivalent to measuring asymetric $(N_x = N_y)$ ellipsoids and thus introduces another shape dependence in addition to that due to spin waves. Direct comparison with the symmetric $(N_x = N_y)$ form of the theory is inaccurate. A proper comparison can be made in two ways: (1) generalizing the theory to include the $N_x = N_y$ case and then using the data directly, (2) transforming the data and then comparing with $N_x = N_y$ theory. Both comparisons are reported and their equivalence is demonstrated. It is concluded that Schlomann's theory does not fit the data within the spin wave (SW) manifold. It is suggested that the theory of Sparks may provide a better fit. The difference between coupling models seems to be the basic difference between the two theories. Experimentally and theoretically, $\triangle H$ and $(\chi_m'')^{\text{eff}}$ are different quantities. The dispersive effects of spin wave coupling are shown to be the cause of the difference between the theoretical $\triangle H$ and $(\chi_m'')^{\text{eff}}$.

8.f. **Interpretation of ferromagnetic resonance measurement made in a nonresonant system,** A. S. Risley and H. E. Bussey, IEEE Trans. Instr. & Meas. IM-15, No. 4, 393-396 (December 1966).

The nonresonant shorted waveguide, NRS, method for measuring the ferromagnetic resonance line width, $\triangle H$, is discussed. The major aim of the paper is to test the usual resonator formula for getting unloaded Q from

loaded Q_L, written here as $\triangle H=H_L/(1+\beta)$. Using this formula, the value of $\triangle H$ is very nearly constant over a wide range of values of $\triangle H_L$ and β. These quantities were varied on a given sphere by changing the angle ϕ between the RF and the dc fields. We believe that this is the first time, with one sample and with no change of either surface finish or axis of magnetization, that $\triangle H$ has been shown to be independent of coupling—a necessary requirement for the method to be accurate. We conclude that the NRS method should be both practical and accurate over a wide range of $\triangle H$.

8.g. **Tensor permeability measurements at *L*-band frequencies using a degenerate mode cavity,** L. B. Schmidt, R. D. Harrington, and W. E. Case, J. Res. NBS 71C (Eng. & Instr.), No. 1, 69-75 (Jan.-Mar. 1967).

The exact solution for the field equations of a cylindrical TM_{110} mode cavity has previously allowed accurate measurements of tensor permeability to be obtained at *X*-band frequencies. It is demonstrated that this method is also applicable at frequencies down to 1 GHz. A brief description of the cavity and measurement system for obtaining data at these lower frequencies is given. Both intrinsic and external permeability results on three commercially available polycrystalline garnets are shown. The larger size rods required for measurements at these frequencies result in some sample size effects in the data. In addition, a previously unreported absorption in the external tensor permeability of the materials was observed.

See also 10.c.

9. Quasi-Optics and Millimeter Waves

Papers

Abstracts

* Correspondence

Measurement of Laser Energy and Power

GEORGE BIRNBAUM, SENIOR MEMBER, IEEE, AND MILTON BIRNBAUM

Abstract—**Measurements of laser energy and power are discussed with emphasis on those methods which attempt to establish the accuracy or standards for such measurements. Devices to attenuate laser radiation are summarized. The principles and limitations of the methods are stressed rather than design details of the apparatus. It is found that standards for the measurement of laser power and energy are yet to be definitely established. Thus caution should be exercised in accepting stated accuracies for laser output measurements.**

Manuscript received February 6, 1967; revised March 6, 1967. This paper is an extension of a report by G. Birnbaum, presented at URSI XVth General Assembly, Munich, Germany, 1966.

G. Birnbaum is with the North American Aviation Science Center, Thousand Oaks, Calif.

M. Birnbaum is with Aerospace Corporation, El Segundo, Calif.

I. Introduction

THE field of lasers and its applications in science, technology, and medicine have matured to the point where the need for accurate measurements of the characteristics of laser radiation, particularly energy and power, is clear.[1,2] To quantitatively determine the accuracy of a measurement, or to establish a standard for a measure-

[1] The measurement of the stability of laser oscillators and the accurate determination of laser wavelength is discussed in an accompanying paper by G. Birnbaum, elsewhere in this issue.

[2] A precise phase adjustment of laser radiation is necessary in some laser applications. A phase shifter has been developed which does not need calibration and has an estimated error of 1 degree [1].

ment, the experimenter must carefully search for all significant systematic and random errors in the apparatus until it is clear that the apparatus is performing as intended, and that its limits of performance can be stated quantitatively. Although little has been accomplished which meets such exacting requirements, considerable effort has been expended in developing methods for measuring laser output.

Although all methods for measuring laser output are considered in this review, the emphasis is placed on accurate methods of measurement of laser energy and power, and, in particular, those methods attempting to establish standards for such measurements. The principles and limitations of techniques are stressed rather than design details of the apparatus. Most of the measurements of laser energy and power used the 6943 Å pulsed ruby laser as the source, which reflects the fact that the ruby laser continues to be among the most useful. Methods to attenuate the intensity of laser radiation are summarized. The work reviewed covers the period of 1960, the advent of lasers, to December, 1966.[3]

II. General Considerations

Lasers have been developed which span the range of wavelengths from submillimeter to the vacuum ultraviolet wavelengths, and with power output capabilities ranging from milliwatts to tens of gigawatts. The scope of the problem of determining laser output is apparent from the enormous wavelength and power range to be covered.

There are several basic types of laser output to be measured, namely, CW, pulsed, Q-switched [2] and mode-locked [3], [4]. CW lasers (gas or solid state) have outputs typically in the range of 10^{-3} to 10^{3} watt. In the case of pulsed lasers, pulse durations of 10^{-3} to 10^{-6} second with peak powers of several megawatts and total output energy of several hundred joules may be encountered. In Q-switched laser operation, the output pulses are usually less than 100 nanoseconds in duration with peak powers as high as a gigawatt. Mode-locked operation of CW, pulsed, and Q-switched lasers has resulted in pulse durations as short as 10^{-12} to 10^{-13} second with peak powers as high as 10^{10} watts. The enormous range of laser pulse widths, output powers, and energies greatly complicates the measurement problem. It is very unlikely that any one method or instrument can be used to measure the output power and energy of all lasers.

Most measurements of laser output, including the most precise, have been made with CW and pulsed lasers operating in the visible region of the spectrum. By far the most difficult of the outputs to be measured is that from mode-locked lasers because the direct determination of the pulse shape of mode-locked lasers is usually not possible due to the extremely short duration of the pulses.[4] Measurement of the power of Q-switched lasers can usually be effected by observation of the pulse shape with fast photoelectric or photoconductive detectors, and measurement of the energy can be obtained by integration of the laser output. In the case of pulsed lasers, photodetectors are employed to measure power and energy. A combination of radiometric and calorimetric devices is frequently used in order to improve the accuracy of the output measurements. The output of CW lasers can be obtained by calorimetric or radiometric methods. Because of the high peak powers of the Q-switched lasers and particularly the mode-locked Q-switched lasers, numerous precautions are required to avoid damage to the optical components and to stay within the linear range of the detectors. Attenuators are necessary in most cases to protect sensitive detectors and to stay within their linear range.

The high output power available from lasers has resulted in several novel techniques for power measurement which utilize the nonlinear susceptibility of crystals. Measurement of the dc polarization [5] and second harmonic intensity have been utilized [6]. Other methods include vaporization of known amounts of material and thermionic emission from a metal surface in a vacuum [7].

Laser output detectors ideally should satisfy the following conditions:

1) The response should be uniform over the area of the receiving aperture.

2) The response should be independent of the direction of the incident radiation so that it has the same response to the highly collimated laser beam as it has to the radiation from the calibrating source.

3) The response should be linear so that energy from the relatively weak calibrating source incident, perhaps for 1 second, and the energy from the pulsed laser, incident for 10^{-6} second or much less, can be compared.

Unless the detector is an absolute device, i.e., the constant of the device is known theoretically, the detector must be calibrated. This is usually the case, and because the detector may have a more or less complicated wavelength response, it must be calibrated at the laser wavelength. This may be accomplished by the use of a thermopile calibrated at the laser wavelength, or a standard lamp whose spectral radiance is known.

III. Power Measuring Device[5]

A practical and widely used system for measuring the output of pulsed lasers uses a photoelectric detector with its output displayed on an oscilloscope [10]–[12]. When the laser is a single-shot device, it is necessary to record the oscilloscope trace photographically. Energy is measured by integrating the power pulses. In the case of Q-switched lasers, the pulse is so short, usually of the order of 10 nanoseconds, that attention must be paid to the circuits used in conjunction with the photodetectors if the pulse shape is to be faithfully recorded.

Two types of photodetectors are in common use, the vacuum phototube and the photodiode (e.g., the silicon photodiode [13]) which is capable of handling much higher power levels. The latter may have a detecting area smaller

[3] A review of laser output measurements through 1964 was given by G. Birnbaum [2].

[4] *Note added in proof:* A novel method of making such measurements has been described by J. A. Armstrong, "Measurement of picosecond laser pulse widths," *Appl. Phys. Lett.*, vol. 10, pp. 16–18, January 1967.

[5] Other reviews of methods of measuring laser output are given in [8] and [9].

than the cross-sectional area of the laser beam. If this is the case, a diffusing screen should be used to reduce errors due to nonuniform intensity distribution in the laser beam.

In using photodetectors, it should be noted that the sensitivity may vary appreciably not only between units, but also over the area of the photosurface. Each phototube must be calibrated separately at the laser wavelength and frequently because of variations of phototube properties. Temperature control may be necessary to obtain reproducible results with photomultiplier tubes. In addition, errors may be introduced by the anode shadow on the cathode of a phototube in measurements where the use of a diffusing screen is not desirable. Since the spread of the laser beam may increase with laser energy, the effect of the anode shadow will vary with laser output. This introduces nonlinearity even in the linear region of the phototube [14].

A solar cell (silicon photodiode) may be used to measure laser output [15], [16]. Since its response peaks approximately at 870 millimicron, it may be used in the near infrared region.

A novel method of measuring intense laser radiation, in which the beam passes through the detector with negligible loss of intensity, uses the optical rectification effect in crystals [17]. In crystals which have no center of inversion, an intense laser beam produces a dc polarization which can be measured as a voltage appearing across the crystal [5], [18]. This polarization is equal to the light intensity multiplied by a second-order nonlinear susceptibility tensor, which is the same as that for the electro-optic effect. The principle of this method has been demonstrated using a quartz crystal as the optical rectifier [5].

IV. Attenuators

To protect photodetectors from overload or damage, as well as to extend the range of all types of devices for measuring laser output, it is frequently necessary to reduce the intensity of radiation incident on the detector. For this purpose, a variety of attenuators have been developed: neutral density filters [19]; silicon wafers [20]; an integrating sphere [11]; coarse diffraction gratings [21]; and a diffuse reflector such as a block of magnesium oxide whose distance from laser to reflector and reflector to photocell may be controlled [22]. Of these attenuators, the neutral density filter is most susceptible to damage by very high power laser beams. However, a cell filled with a solution which is absorbing at the laser wavelength may be used in place of the neutral density filter. In the silicon wafer attenuator, the laser radiation is completely absorbed producing a large number of free electrons and holes, which recombine by radiative and nonradiative processes [20]. The low efficiency of the radiative process produces sufficient radiation for photodetector observation of the beam pulseform, but not enough to damage the photodetector. Unlike conventional glass attenuation filters, the attenuation factor of silicon disks is stable with usage.

The integrating sphere attenuator is essentially a sphere with two small openings. The laser beam enters one of these, is scattered and rescattered from the inner surface, then emerges from the second opening. The attenuation factor is given approximately by the ratio of the area of the exit hole to the total surface of the sphere [11]. In the diffraction grating attenuator, the monochromatic laser radiation is deflected into a direction determined by the order of the diffraction. The ratio of the intensity of the diffracted beam to that of the beam incident on the grating is given by diffraction theory [21]. The diffuse reflector attenuator will be discussed in Section VI.

Since the calibration factor of attenuators is usually not known, it must be calibrated when used with devices for measuring laser output. However, the calibration of attenuators involves only the measurement of power ratio rather than absolute power, and thus depends for its accuracy on the linearity of the output measuring device.

V. Energy Measuring Devices

A variety of calorimeters for measuring the energy of pulsed and CW laser radiation have appeared [23]–[31]. The energy from the laser is used to raise the temperature of the device, and this rise in temperature is recorded thermoelectrically. Until the whole device has reached a uniform temperature the temperature rise should not be recorded, as nonuniform distribution of the incident energy can cause high local heating. For this reason, a long cooling time constant is generally required to minimize loss of energy by radiation, conduction, or convection from the active element before reaching equilibrium. Also when calibrating the device, the energy can be received over a long interval of time to compensate for the low power of the source. A steady drift, which is slow in comparison with the time required to make a measurement, is tolerable.

Perhaps the simplest calorimeter developed thus far consists of a disk of thin foil with two thermistors attached at the center of the rear surface [13]. To increase the absorption of a gold disk, the surface was painted black, but the paint could be entirely removed by one or two pulses from a Q-switched Nd^{3+} glass laser. Black surfaces of electroplated platinum and nickel proved to be more resistant to damage, but required a number of exposures to laser radiation before the surface stabilized.

A method of increasing the blackness of the detector element is to form the material in the shape of a cone so that the incident radiation is multiply reflected until it is totally absorbed. Broadband calorimeters of this type have been constructed of a carbon cone [23], [24], and blackened and unblackened metal cones [13], [25] with thermoelectric elements such as a thermistor in intimate contact with the cone to measure the temperature rise. To ensure a long cooling time constant, these absorbing elements may be mounted on fine glass supports in thermally insulated enclosures. A difficulty in fabricating cone shaped devices is the requirement that the apex be sharp and the angle of the cone should be narrow to make the light trapping characteristics good.[6]

In a related method, the laser beam is focused onto a

[6] It is claimed that these problems may be circumvented by a device using a stack of one hundred stainless steel razor blades so that the sharp edges represent a cone or wedge which has been folded many times [26].

small hole in a spherical metal shell, located in an evacuated chamber to reduce heat loss [27], [28]. The defocused laser beam hits the inside of the shell and is absorbed uniformly at the inner surface after many reflections.

In calorimeters in which the energy is absorbed in a thin surface layer, care must be exercised to avoid reradiation losses. In this case, the absorbing surface may experience a large increase in temperature, even though the calorimeter indicates a small temperature rise. This effect is usually most serious in measurement of the output of Q-switched lasers.

In the liquid calorimeter [29] (see Section VI), the absorbed energy is distributed throughout a volume and thereby avoids errors which may be caused by surface damage and localized radiative cooling of solid receptors. However, power must be maintained at a sufficiently low level to prevent light scattering from the calorimeter by boiling and convective liquid currents.

With the calorimeters mentioned thus far, it is necessary to wait for temperature equalization by thermal conduction before the temperature rise is measured. Since heat may be lost before equilibrium is reached, an error may be introduced in the determination of the total energy absorbed. This type of error is minimized in the rat's nest wire calorimeter. The laser beam is absorbed in 1000 feet of fine enameled copper wire, the bolometer unit, tangled into a ball which is placed in an isothermal enclosure [30]. The change in resistance of the wire is a measure of the total energy absorbed. Since the change in resistance is independent of the volume distribution of the wire, there is no need to wait for temperature equalization as in the usual calorimeter. By tangling the wire, the incident energy suffers multiple reflections and the unit has the characteristics of a gray body. However, the actual absorption coefficient of a tangle of wires is indeterminate. In addition, the insulating varnish on the wire substantially increases the heat capacity of this type of calorimeter, and makes it difficult to calculate its thermal response. By winding the wire as a grid so that the laser beam suffers only one reflection, a device was obtained which was easily and reproducibly wound and whose performance can be estimated theoretically [13].

The measurement of the linear momentum of radiation by the deflection of a ballistic torsion pendulum has been used to measure the output of a pulsed ruby laser beam [32], [33]. Unlike the devices mentioned previously, the ballistic pendulum transmits the laser beam through the apparatus with negligible absorption. This measurement is absolute in that it does not require calibration. However, accuracy limits are not given for the measurement. In fact, except for an estimate of 10 percent for the carbon cone calorimeter [23] and 20 percent for integrating sphere [28], and for an investigation to be described, no accuracy is given for any of the previously mentioned power and energy measurements.

Killick et al. [13] investigated the performance of a wire calorimeter and a silicon disk photodiode, both of which were calibrated against a standard thermopile. Measurements of the output of a Q-switched Nd^{3+} glass laser were performed using the two systems simultaneously. It was concluded that the laser output could be measured to ± 25 percent.

TABLE I

SUMMARY OF DEVICES FOR ENERGY AND POWER MEASUREMENTS*

Device	Range of operation	Typical response time (s)	Surface damage†
Energy (Joules)			
Cone calorimeters	10^{-2}–2×10^{3}	1–20	10–20 mW/cm^2 in 50 ns
Metal disk calorimeter	10^{-2}–10	10	50 mW/cm^2 in 50 ns
Rat's nest calorimeter	10^{-3}–10	10^{-4}	—
Wire calorimeter	10^{-3}–0.5	10	10 mW/cm^2 in 50 ns
Liquid calorimeter	1–500	10–60	‡
Torsion pendulum	0.5–500	60	—
Integrating photocurrent	10^{-8}–10^{-3}	1	1 mW/cm^2
Thermopile	10^{-6}–1	10^{-1}	300 mW/cm^2
Copper sphere	5×10^{-4}–10	180	—
Power (W/cm^2)			
Phototube	10^{-8}–10^{-3}	3–10 ns	1 mW/cm^2
Photodiode	10^{-4}–6	0.3–4 ns	10 W/cm^2‖
Nonlinear crystal	10^{3}–10^{12}	10^{-5}	10^{12} W/cm^2§

* Adapted from [8] and [13].
† Surface damage may occur at the indicated power density.
‡ Local boiling of the liquid should be avoided.
‖ Above this power, the response of this device is nonlinear.
§ Breakdown of quartz at roughly this power density.

Table I summarizes some characteristics of the energy and power measurement devices described previously.

VI. Standards of Energy and Power Measurements

Attempts at accurate measurement of laser energy and power have been initiated in a few laboratories. We consider first the liquid calorimeter represented schematically in Fig. 1(a) which has been developed by Jennings [34] at the National Bureau of Standards, U.S.A., and used to measure the output energy of a pulsed ruby laser. The heart of the calorimeter is a 3-mm thick absorption cell [Fig. 1(b)] whose back and sides are made of gold-plated silver. The cell is filled with an aqueous solution of $CuSO_4$ or $CuCl_2$, and the concentration is adjusted such that 99.9 percent of the laser beam is absorbed in a 6.0-mm path. The cell is mounted on glass fibers inside a massive brass block for temperature stability, and the entire system is placed in an aluminum box and insulated from it with styrofoam. A calibrated thermocouple measures the temperature rise between the absorption cell and the brass heat sink. The calorimeter was calibrated in two different ways: 1) by a substitution method in which a known amount of electrical energy was put into the cell by means of a heater immersed in the liquid, and 2) by a computation based on a knowledge of the specific heat and mass of all the components, and thermocouple sensitivity. Reflection losses due to the entrance windows were calibrated with a 6328 Å He-Ne laser. The calorimeter was designed to accept up to 30 joules input and peak powers up to 200 megawatts per square centimeter. Two calorimeters of this design were compared and found to agree within two percent.

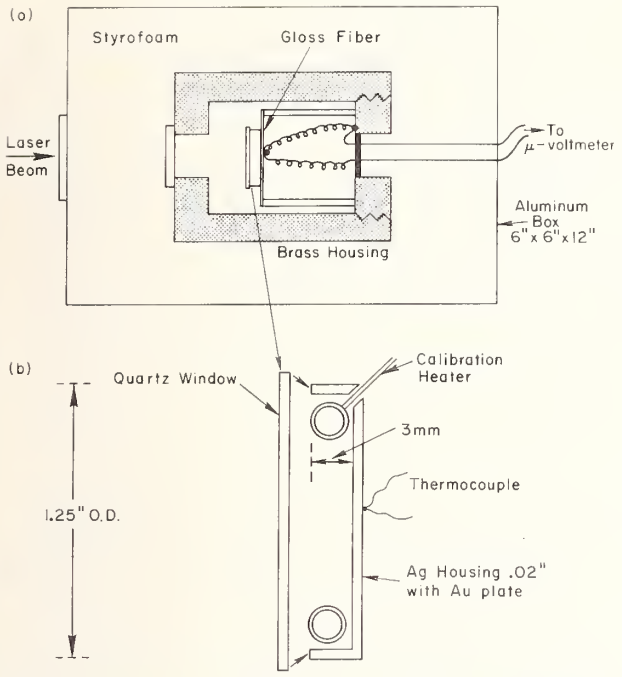


Fig. 1. Liquid calorimeter (Jennings [34]). (a) Schematic drawing. (b) Exploded view of absorption cell which contains $CuSO_4$ or $CuCl_2$ solutions.

McSparron et al. [35], also at the National Bureau of Standards, have developed a radiometric method for measuring laser energy and power as shown in Fig. 2(a). Power from the laser is incident normally on an $MgCO_3$ block and a small part is reflected into a receiver (phototube or thermopile). The ratio of the power entering the receiver to the initial power is given by

$$\alpha = \frac{\beta A \cos\theta}{\pi d^2}, \tag{1}$$

where A is the area of the receiver, β is the reflectance factor of the $MgCO_3$ at θ, and d is the distance from the $MgCO_3$ to the receiver. The quantity β was measured at 20° on the NBS goniophotometer with monochromatic light of the same wavelength as the laser. Two receiver positions are used in order to test for systematic errors. The calibration arrangement for the receiver is shown in Fig. 2(b). Energy from a lamp calibrated for spectral irradiance is passed through a filter to spectrally limit the output of the lamp, and through a shutter to form light pulses of known duration. Simultaneous measurements at the two receiver positions of the energy of a pulsed ruby laser agreed to about 4 percent.

As the first step toward establishing measurement standards in the optical frequency region in Japan, Sakurai et al. [36] have developed a microcalorimeter, shown schematically in Fig. 3(a), to measure CW power output of lasers. The laser thermopile unit consists of an aluminum cone-shaped mount with blackened inside walls and a star thermopile as shown in Fig. 3(b). An absorber, a blackened

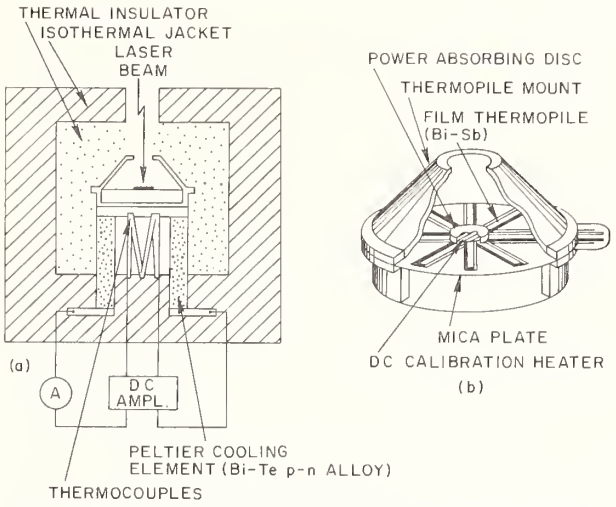


Fig. 2. Radiometric method for measuring laser energy and power (McSparron [35]). (a) Experimental arrangement for measuring laser energy. (b) Method of calibrating the receiver.

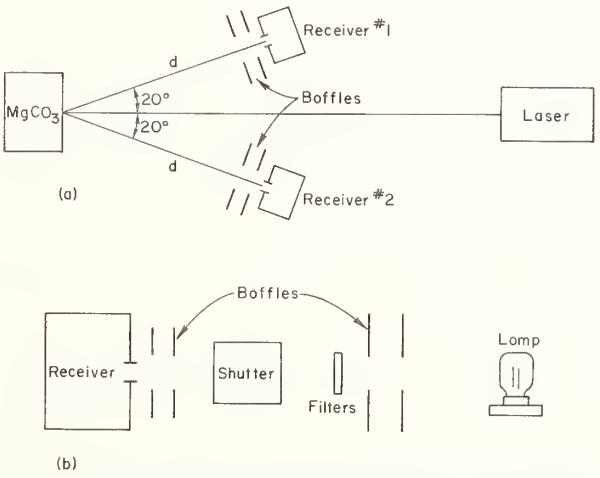


Fig. 3. A microcalorimeter using the dc substitution method (Sakurai et al. [36]). (a) Schematic drawing. (b) The power absorbing disk and thermopile mount.

aluminum disk, within which a dc heater for calibration purposes is embedded, is fixed at the center of the star thermopile. By the use of this unit, the power of a CW He-Ne laser operating at 6328 Å was measured by the current substitution method. However, because there was a difference in sensitivity of laser power measurement and direct current power measurement, it was necessary to determine the effective substitution efficiency of the unit.

The temperature rise of the unit, due to either absorption of laser or dc power, is detected by the thermopile. The thermopile output is amplified by a high-gain dc amplifier which controls the cooling current to the Peltier elements. By this method, the temperature difference between the jacket and the energy absorbing mount is reduced to less than 10^{-5}°C at equilibrium. At this point, the cooling current is proportional to the energy absorbed by the unit. A typical unit has an efficiency of 92.2 percent at 6328 Å.

TABLE II

CHARACTERISTICS OF POTENTIAL STANDARDS FOR LASER POWER ENERGY

	Wavelength (Å)	Power and energy range	Accuracy (%)	Reference
Liquid calorimeter	6943	0.01 J–30 J*	2‡	18
Photodetector	6943	†	4‡	19
Microcalorimeter	6328	1–50 mW	0.5	20

* Maximum power, 200 mW/cm^2.

† The receiver, a phototube or thermopile, can tolerate maximum power levels of the order of milliwatts per square centimeter. However, the incident power is attenuated by reflection from a $MgCO_3$ block.

‡ The radiometric and calorimetric methods were compared in measurements of the energy of a pulsed ruby laser, and the results differed by 9 percent. The reason for this discrepancy is unknown at present.

From an analysis of this technique, it was concluded that the absolute accuracy of power measurements of a CW He-Ne laser (6328 Å) is better than 0.5 percent in the range of 1 to 50 milliwatt. Work is under way to measure the energy of pulsed ruby lasers with this calorimeter.

The characteristics of the energy and power measurements discussed in this section are summarized in Table II. Accurate determinations of laser energy and power output are still in an early stage of development. For example, when the NBS radiometric and calorimetric methods were compared in measurements of the energy of a pulsed ruby laser, the results differed by 9 percent, a discrepancy greater than the sum of the experimental errors given for these methods. Work is in progress to determine the origin of this discrepancy. In any case, it is clear that caution should be exercised in accepting accuracies for laser energy and power measuring devices until standards for such measurements are definitely established.

ACKNOWLEDGMENT

The authors thank Drs. D. A. Jennings and K. Sakurai, and D. A. McSparron, C. A. Douglas, and H. L. Badger for making available their manuscripts prior to publication.

REFERENCES

[1] J. Hamasaki and H. Noguchi, "Variable phase shifter for laser light using birefringent crystals," *Proc. IEEE* (*Correspondence*), vol. 53, pp. 80–81, January 1965.

[2] G. Birnbaum, *Optical Masers*, Advances in Electronics and Electron Physics, suppl. 2. New York: Academic Press, 1964.

[3] M. H. Crowell, "Characteristics of mode-coupled lasers," *IEEE J. of Quantum Electronics*, vol. QE-1, pp. 12–20, April 1965.

[4] A. J. DeMaria, D. A. Stetser, and H. Heynau, "Self mode-locking of lasers with saturable absorbers," *Appl. Phys. Letters*, vol. 8, pp. 174–176, April 1966.

[5] A. K. Kamal and M. Subramanian, *Symp. on Optical Masers*. Brooklyn, N. Y.: Polytechnic Press, 1963, pp. 601–613.

[6] N. Bloembergen, *Nonlinear Optics*. New York: W. A. Benjamin, Inc., 1965, ch. 5, pp. 131–134.

[7] S. H. Khan, F. A. Richards, and D. Walsh, "Time resolution of laser induced electron and ion emission," *IEEE J. of Quantum Electronics* (*Correspondence*), vol. QE-1, pp. 359–360, November 1965.

[8] P. J. Bateman, *IEE Conf. on Lasers and Their Application* (London), pp. 40–41 to 40–47, 1964.

[9] R. C. Honey, in *Laser Parameter Measurement Handbook*, H. G. Heard, Ed. Palo Alto, Calif.: h nu systems, inc., 1966, ch. 4, pp. 4-1 to 4-116.

[10] T. H. Maiman, R. H. Hoskins, I. J. D'Haenens, C. K. Asawa, and V. Evtuhov, "Stimulated optical emission in fluorescent solids, II: Spectroscopy and stimulated emission in ruby," *Phys. Rev.*, vol. 123, pp. 1151–1157, August 1961.

[11] E. Schiel, "Photoelectric energy meter for measuring laser output," *Proc. IEEE* (*Correspondence*), vol. 51, pp. 365–366, February 1963.

[12] K. Kinoshita and T. Suzuki, "Measurement of mean laser output power," *Japan J. Appl. Phys.*, vol. 2, pp. 811–812, 1963.

[13] D. E. Killick, D. A. Bateman, D. R. Brown, T. S. Moss, and E. T. de la Perrelle, "Power and energy measuring techniques for solid state lasers," *Infrared Phys.*, vol. 6, pp. 85–109, June 1966.

[14] M. Stimler and G. P. Worrel, "Nonlinear photocell response due to laser beam spread," *Appl. Optics*, vol. 3, pp. 538–539, April 1964.

[15] E. J. Schiel and J. J. Bolmarcich, "Absolute measurement of GaAs diodes radiation using solar cells," *Proc. IEEE* (*Correspondence*), vol. 51, pp. 1780–1781, December 1963.

[16] S. Deb and M. K. Mukherjee, "Measurement of radiant power output with a solar cell," *IEEE J. of Quantum Electronics* (*Correspondence*), vol. QE-1, pp. 219–220, August 1965.

[17] M. Bass, P. A. Franken, and J. F. Ward, "Optical rectification," *Phys. Rev.*, vol. 138, pp. A534–A542, April 1965.

[18] J. F. Ward, "Absolute measurement of an optical-rectification coefficient in ammonium dihydrogen phosphate," *Phys. Rev.*, vol. 143 pp. 569–574, March 1966.

[19] A. L. Glick, "A method for calibration of laser energy output," *Proc. IRE* (*Correspondence*), vol. 50, p. 1835, August 1962.
——, "Comment on a method for calibration of laser energy output," *Proc. IEEE* (*Correspondence*), vol. 51, p. 1360, October 1963.

[20] N. N. Winogradoff and H. K. Kessler, "Radiative recombination lifetimes in laser-excited silicon," *Appl. Phys. Lett.*, vol. 8, pp. 99–101 February 1966.
——, "Silicon attenuators for laser measurements," *NBS Tech News Bull.*, pp. 182–183, October 1966.

[21] R. Gerharz, "Attenuation of laser light by a diffraction grating, *Proc. IEEE* (*Correspondence*), vol. 52, p. 438, April 1964.
——, "Absence of polarization effects in diffraction-attenuated laser light," *Proc. IEEE* (*Correspondence*), vol. 53, pp. 105–106, Januar 1965.

[22] R. C. C. Leite and S. P. S. Porto, "A simple method for calibration of ruby laser output," *Proc. IEEE* (*Correspondence*), vol. 51, pp 606–607, April 1963.

[23] T. Li and S. D. Sims, "A calorimeter for energy measurements o optical masers," *Appl. Optics*, vol. 1, pp. 325–328, May 1962.

[24] J. A. Calviello, "An optical calorimeter for laser energy measurements," *Proc. IEEE* (*Correspondence*), vol. 51, pp. 611–612, April 1963.

[25] S. Koozekanani, P. P. Debye, A. Krutchkoff, and M. Ciftan "Measurements of the laser output," *Proc. IRE* (*Correspondence*) vol. 50, p. 207, February 1962.

[25a] B. F. Scott, "Fabrication and performance of calorimeters for laser energy measurements," *J. Sci. Instrum.*, vol. 43, pp. 685–687 October 1966.

[26] J. A. Ackerman, "Laser energy measuring device," *Appl. Optics*, vol. 3, pp. 644–645, May 1964.

[27] C. W. Bruce, "Pulse laser instrumentation," U. S. Air Force Weapons Lab. Tech. Rep. 64-127, 1965 (unpublished).

[28] V. S. Zuev and P. G. Kryukov, "Calorimeter for measuring the radiation energy of optical generators," *Instruments and Exper. Tech.*, vol. 3, pp. 563–564, 1963.

[29] E. K. Damon and J. T. Flynn, "A liquid calorimeter for high-energy lasers," *Appl. Optics*, vol. 2, pp. 163–164, February 1963.

[30] R. M. Baker, "Measuring laser output with rat's nest calorimeter," *Electronics*, pp. 36–38, February 1963.

[31] F. Davoine, J. L. MacQueron, and A. Nouailhat, "Etude de l'énergie d'un faisceau L.A.S.E.R. par mesures calorimétriques," *Journal de Physique*, vol. 24, pp. 1103–1106, 1963.

[31a] B. F. Scott, "Laser energy measurements with a liquid absorption cell," *J. Sci. Instrum.*, vol. 43, pp. 940–942, December 1966.

[32] J. J. Cook, W. L. Flowers, and C. B. Arnold, "Measurement of laser output by light pressure," *Proc. IRE* (*Correspondence*), vol. 50, p. 1693, July 1962.

[33] M. Stimler, Z. I. Slawsky, and R. E. Grantham, "Torsion pendulum photometer," *Rev. Sci. Instrum.*, vol. 35, pp. 311–313, March 1964.

[34] D. A. Jennings, "Calorimetric measurement of pulsed laser output energy," *Program, Conf. on Precision Electromagnetic Measurements* (*Abstracts*, pp. 13–14), NBS, Boulder, Colo., 1966.

[35] D. A. McSparron, C. A. Douglas, and H. L. Badger, "Radiometric measurement of laser output," *Program, Conf. on Precision Electromagnetic Measurements* (*Abstracts*, p. 12), NBS, Boulder, Colo., 1966.

[36] K. Sakurai, Y. Mitsuhashi, and T. Honda, "A laser microcalorimeter," *Program. Conf. on Precision Electromagnetic Measurements* (*Abstracts*, pp. 14–16), NBS, Boulder, Colo., 1966.

Calorimetric Measurement of Pulsed Laser Output Energy

D. A. JENNINGS

Abstract—There are several methods by which one may measure the energy output of the pulsed laser. However, the technique which seems to be most promising as far as accuracy and precision are concerned is the calorimetric method. We have designed, built, and calibrated calorimeters for measuring the output energy of the pulsed ruby laser (6943Å). The heart of the calorimeter is a small absorption cell containing an aqueous solution of $CuSO_4$. The temperature of the absorption cell, as measured by a thermocouple, indicates the energy absorbed by the calorimeter. The calorimeter was calibrated in two different ways: 1) the known heat capacity of the absorption cell and the thermocouple sensitivity calibration gives a calorimeter calibration, which agrees within 0.3 percent of 2) an electrical energy substitution calibration which is obtained via a heater wire contained in the absorption cell solution. A method has been devised by which two calorimeters may be intercompared. Calorimeters which we have built and calibrated agree with each other to about 0.7 percent. This specific calorimeter has been designed to measure energies up to 30 J and will take peak powers o- up to 200 MW/cm².

Manuscript received June 23, 1966. This paper was presented at the 1966 Conference of Precision Electromagnetic Measurements, Boulder, Colo.

The author is with the National Bureau of Standards, Boulder, Colo.

Introduction

ONE OF THE IMPORTANT parameters of pulsed laser systems is their output energy. The application of pulsedl asers requires knowing this energy with varying degrees of accuracy, depending on the application. There are several methods by which one may measure the energy output of the laser [1]–[7]. In this paper we would like to describe an optical calorimeter for measuring the output energy of a pulsed laser. The calorimetric technique seems very promising as far as accuracy and precision are concerned. The calorimeter which is described in this paper was designed to measure 0.1- to 30-J laser pulses in the 5000 to 10 000-Å range.

Reprinted from IEEE TRANSACTIONS
ON *INSTRUMENTATION AND MEASUREMENT*
Volume IM-15, Number 4, December, 1966
pp. 161-164

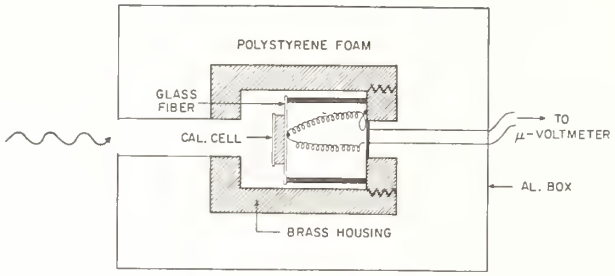


Fig. 1. Cross-sectional diagram of the calorimeter, showing layout of components, method of supporting calorimeter absorption cell, and position of the thermocouple.

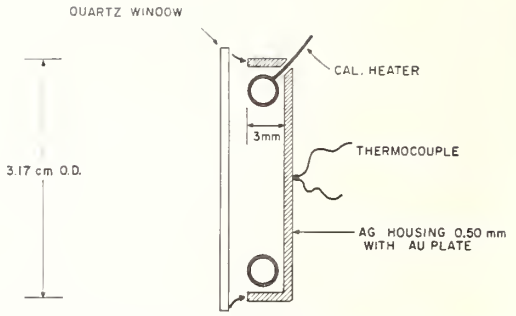


Fig. 2. Cross-sectional diagram of the calorimeter absorption cell, showing the approximate size of components and the position of the heater wire.

General Description

The basis of the calorimetric system is the calorimeter proper, which is shown in Fig. 1. The calorimeter absorption cell, filled with an absorbing solution, was supported in a massive brass housing (heat sink) by means of small glass fibers. The brass housing was placed in polystyrene foam insulation and the entire system enclosed in an aluminum box. One junction of a Cu-constantan thermocouple was attached to the absorption cell, and the other junction was attached to the brass housing. When a pulsed laser was fired into the absorption cell, the cell rose in temperature and the thermocouple generated a voltage proportional to the temperature difference between the absorption cell and the brass housing. The output voltage of the thermocouple was then a measure of the energy in the laser beam. The voltage generated by the thermocouple was measured with a microvoltmeter whose output was fed into a strip-chart recorder.

The Absorption Cell

Since the heart of the system is the absorption cell, we shall describe it in detail. Figure 2 shows a cross-sectional view of the absorption cell. The absorption cell was made of silver, so as to give a fast thermal equalization to the cell. The silver cell was then electroplated with 2.5×10^{-4} cm of gold in order that the tarnish problem be kept to a minimum. The cell diameter was 3.17 cm and the depth was 3.0 mm. The wall thickness was 0.50 mm. The silver cell had three small holes near the edge. One hole was utilized for filling, and the other two were for a heater wire. The heater wire we shall describe later. The entrance window was made of quartz. Eopoxy was used to bond the quartz to the silver cell and also to seal the fill and heater-wire holes. The cell was filled with a one molar solution of $CuSO_4 \cdot 5H_2O$. The addition of 2 drops of Bendix purple ink per 10 cc of solution extends the useful range to 5000 Å, as opposed to only 6500 Å with $CuSO_4$ solution alone. An absorption curve of this solution is shown in Fig. 3. The absorption coefficient α is defined by the equation

$$I = I_0 C^{-\alpha x}$$

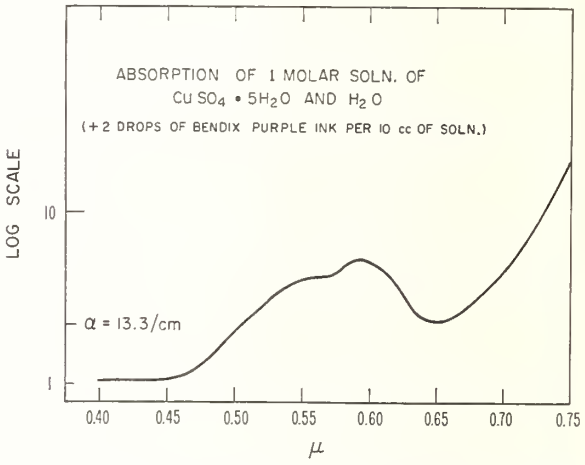


Fig. 3. Absorption spectrum of the solution used in the calorimeter absorption cell. The calibration of the arbitrary log scale is shown and was measured at 6943 Å.

where I is the laser beam intensity at a distance x in the absorbing liquid, I_0 is the initial laser beam intensity, and x is the distance into the liquid as measured from the surface. The absorption coefficient α was measured using a low intensity incoherent light source at 6943 Å, and checked using the high intensity Q-switched ruby laser. The cell filled with this solution will absorb 99.9 percent of the laser beam, not counting the Fresnel losses of the window.

A cell of this description gave about 0.1°C temperature rise for a 1-J input.

Calibration of the Calorimeter System

Great care must be taken in the calibration of the calorimeter since the accuracy is no better than the calibration. Two independent methods of calibration were used. One calibration was based on the heat capacity of the absorption cell and the thermocouple calibration, and the second calibration was based on an electrical energy substitution via the heater wire.

The heat capacity calibration consisted of weighing all the component parts of the absorption cell and then

using the best obtainable values for the specific heats to calculate the heat capacity of the absorption cell. All components of the absorption were accurately weighed. The thermocouple sensitivity was calibrated using standard techniques. The heat capacity calibration factor for this particular calorimeter was calculated to be $3.102 \pm .024\ \mu V/J$. The uncertainty was obtained by taking into account the errors in the specific heats, weighing, and thermocouple sensitivity. This calibration factor has also been adjusted to correct for Fresnel reflection losses of the entrance window. The error here was very small, since the correction was only on the order of 4 percent. The true calibration was obtained by use of the following equation

$$SHC_t = SHC(1 - R)$$

where SHC is the specific heat calibration, SHC_t is the adjusted specific heat calibration, and R is the reflection as calculated from the Fresnel equations.

The electrical energy substitution calibration was accomplished by passing a known dc current through the heater wire for a known length of time. The heater wire was Teflon coated and had a resistance of 16 ohms per foot and a total resistance of 110 ohms.

The dissipation of heat outside the cell from the leads to the heater was small and corrected for in the calibration.

Prior to making the electrical energy substitution calibration, all instruments were calibrated to an accuracy of 0.2 percent.

The electrical energy calibration proceeded in the following manner. The gate output from a preset scaler, counting the line frequency (60 hertz), was used to turn on a transistor switch. The transistor was powered by a constant current source (constant to 0.01 percent), and the absorption cell heater wire was in the collector circuit of the transistor. The voltage drop V across the heater wire was measured via the calibrated strip-chart recorder. Since the voltage V and the heater-wire resistance R and the time t are known, the energy J, put into the absorption cell, can be calculated from

$$J = \frac{V^2}{R} t.$$

A typical calibration trace obtained from the strip-chart recorder is shown in Fig. 4. The overshoot at the beginning of the trace was due to the close proximity of the heater wire to the silver housing of the absorption cell. The error, introduced because the silver housing was warmer during the overshoot than when in an isothermal condition, was small enough to be neglected. An estimation of this error from both convection and radiation cooling shows it to be less than 0.25 percent. Other losses, such as conduction of heat along the wire leads and quartz mounting fiber, were less than 0.1 percent over the entire measurement (20–30 minutes).

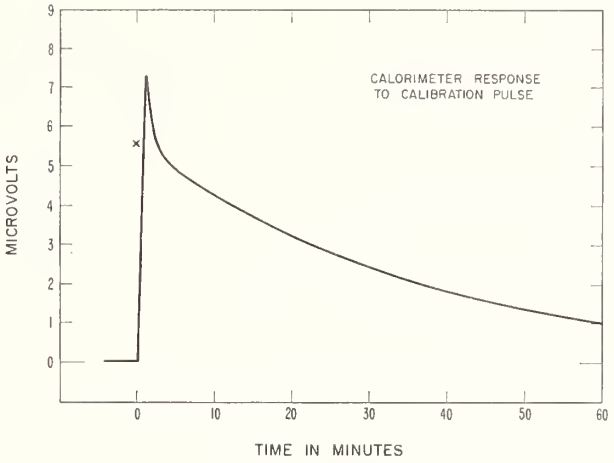


Fig. 4. Typical chart trace of the calorimeter system response to a calibration. Note the overshoot due to the close proximity of heater wire to the silver housing. The total energy input to the calorimeter here was 1.55 J.

The electrical energy substitution calibration was finally obtained by making an exponential extrapolation of the thermal decay back to time = 0, as shown by the x on the chart (Fig. 4).

The precision of the calorimeter system was determined by the reproducibility of the electrical energy substitution calibration. The electrical energy substitution calibration factor for this calorimeter system was $3.092\ \mu V/J$. This factor was the average of 13 calibration runs. The data had a standard deviation of 0.009. The two calibrations, specific heat and electrical energy substitution, compare very favorably with each other. From this we assume that the accuracy and precision is good to at least ± 1 percent.

Results and Intercomparisons of Calorimeters

A typical output of the calorimeter system for a normal ruby laser is shown in Fig. 5. Notice here that there was no overshoot as compared with the electrical calibration. The output energy of the laser was found by making the exponential extrapolation to time = 0, indicated by the x on the chart, and then dividing the deflection at time = 0 by the calibration factor.

Two calorimeter systems of the same basic design were intercompared with a setup as shown in Fig. 6. The technique was to measure the reflectivity R of the beam splitter. This position of calorimeter, as shown in Fig. 6, was the A position and the reflectivity of the beam splitter was R_A. Then the calorimeters were interchanged to position B, and R_B measured. Now if calorimeter No. 1 was assumed to be correct, and calorimeter No. 2 was in error by a fractional amount Δ

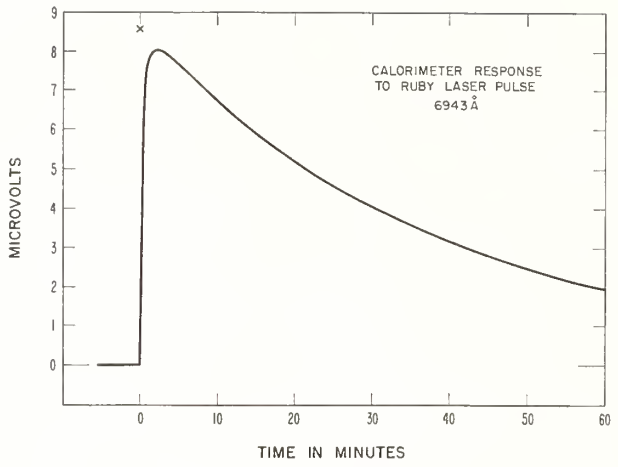


Fig. 5. Typical chart trace of the calorimeter system response to normal ruby laser pulse. Note here that there is no overshoot at the beginning of the trace. The energy in the laser pulse for this shot was 2.77 J.

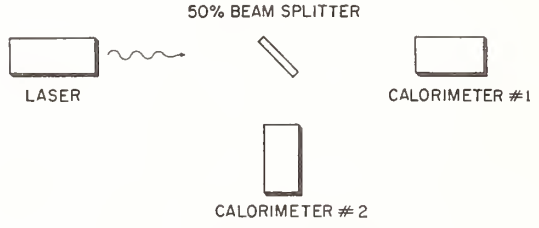


Fig. 6. Block diagram of apparatus arrangement used in the intercomparison of two calorimeters.

where

$$\Delta = \frac{E_2^1 - E_2}{E_2^1}$$

where E_2^1 is the true incident on calorimeter No. 2, and E_2 is the apparent energy, then Δ can be shown to be equal to

$$\Delta = \frac{R_A}{R_B} - 1.$$

For lasers which have a polarized output, i.e. 90° or 60°-oriented ruby laser rods, the laser rod must be oriented properly, relative to the beam splitter. This was easily done to $\pm 1°$, and represented an error of no more than ± 0.1 percent.

Using this technique, we find a Δ for our calorimeter systems of 0.007. This result was from an average of 10 comparison runs. The data had a standard deviation of 0.002.

Discussion

We have made several calorimeters modeled after the same basic design with modifications to include cone shaped absorption cells, a different absorbing liquid ($CuCl_2+H_2O$), and Brewster's angle input. We have tested the calorimeters with laser energies up to 15 J (normal laser) and powers up to 150–200 megawatts (Q-switched), and found no anomalies in the calorimeters. It would seem possible to scale the calorimeter components to make calorimeter systems that would measure very high energies, say 10^4 J with one percent accuracy and precision, by using a large aperature, large volume calorimeter so as not to exceed a certain energy density, in order to avoid a change in state of the absorbing fluid.

Acknowledgment

The author would like to thank Dr. J. L. Hall, of the Joint Institute of Laboratory Astrophysics, for the many stimulating discussions concerning the design of the calorimeters, and A. L. Rasmussen, of the National Bureau of Standards, for many of the measurements.

References

[1] T. H. Maiman, R. H. Hoskins, I. J. D'Haenens, C. K. Asawa, and V. Evtuhov, "Stimulated optical emission in fluorescent solids. II. Spectroscopy and stimulated emission in ruby," *Phys. Rev.*, vol. 123, p. 1151,–1157, August 15, 1961.
[2] S. Koozekanani, P. P. Debye, A. Krutchkoff and M. Ciftan, "Measurements of the laser output," *Proc. IRE* (*Correspondence*), vol. 50, p. 207, February 1962.
[3] T. Li and S. D. Sems, "A calorimeter for energy measurement of optical masers," *Appl. Optics*, vol. 1, pp. 325–328; May 1962.
[4] A. L. Glick, "A method for calibration of laser energy output," *Proc. IRE* (*Correspondence*), vol. 50, p. 1835, August 1962.
[5] R. C. C. Leite and S. P. S. Porto, "A simple method for calibration of ruby laser output," *Proc. IRE* (*Correspondence*), vol. 51, pp. 606–607, April 1963.
[6] J. A. Calviello, "An optical calorimeter for laser energy measurements," *Proc. IRE* (*Correspondence*), vol. 51, pp. 611–612, April 1963.
[7] M. Stimler, "Ballistic torsional pendulum," U. S. Gov. Res. Rept. AD-420, p. 469, October 1963.

Millimeter Wavelength Resonant Structures*

R. W. ZIMMERER†, MEMBER, IRE, M. V. ANDERSON†,
G. L. STRINE†, AND Y. BEERS†, MEMBER, IRE

Summary—**This paper discusses the construction of millimeter wave Fabry-Perot resonators, using both planar and spherical reflectors. It also discusses the equivalent circuits of planar reflectors and the method of obtaining efficient power transfer into the resonators.**

THIS PAPER is a further report on the work on millimeter wave Fabry-Perot interferometers that was started in this laboratory by Culshaw.[1–5] These interferometers have become of wide interest because of their use as resonators in optical and millimeter wave masers. These resonators have many other potential uses as spectrometers, refractometers, and wave meters.

* Received October 23, 1962; revised manuscript received November 30, 1962. A condensed version of this paper was presented at the PGMTT National Symposium in Boulder, Colo., on May 22, 1962.

† Radio Standards Laboratory, National Bureau of Standards, Boulder, Colo.

[1] W. Culshaw, "Reflectors for a microwave Fabry-Perot interferometer," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-7, pp. 221–228; April, 1959.

[2] W. Culshaw, "High resolution millimeter wave Fabry-Perot interferometer," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-8, pp. 182–189; March, 1960.

[3] W. Culshaw, "Resonators for millimeter and submillimeter wavelengths," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-9, pp. 135–144; March, 1961.

[4] W. Culshaw, "Millimeter wave techniques," *Advances in Electronics and Electron Phys.*, vol. 15, pp. 197–263; 1961.

[5] W. Culshaw, "Measurement of permittivity and dielectric loss with a millimetre-wave Fabry-Perot Interferometer," *Proc. IEE*, vol. 109, pt. B, Suppl. No. 23, pp. 820–826; 1961.

In the millimeter region the ratio of wavelength to the mirror dimensions, although small compared to unity, is much larger than in the optical region. Therefore, diffraction losses in the millimeter region tend to be much larger. At the same time modes are separated more widely, and it is usually possible to work with a single mode. In the optical region mirrors are made of semisilvered surfaces or by multilayered dielectric surfaces. As is well known,[6] with such mirrors large reflectivity is incompatible with low resonance transmission loss. Culshaw realized that in the millimeter region other techniques allowing the achievement of both objectives were practical. He evolved a scheme of drilling an array of holes in metallic sheets and started work with metal films with photoetched holes deposited on dielectric slabs. We have further developed this technique and have used thin perforated metal foils stretched on frames. This technique appears to be the best available for use with plane reflectors.

For many applications Fox and Li[7] and Boyd and Gordon[8] have demonstrated the superiority of con-

[6] M. Born and E. Wolf, "Principles of Optics," Pergamon Press, Inc., New York, N. Y., Sec. 7.6; 1959.

[7] A. G. Fox and T. Li, "Resonant modes in a maser interferometer," *Bell Sys. Tech. J.*, vol. 40, pp. 453–488; March, 1961.

[8] G. D. Boyd and J. P. Gordon, "Confocal multimode resonator for millimeter through optical wavelength masers," *Bell Sys. Tech. J.*, vol. 40, pp. 489–508; March, 1961.

cave mirrors over planar ones because of the greatly reduced diffraction losses and the greater ease of alignment. In some of the instruments to be described later we have applied their ideas to the millimeter region. Marcuse,[9] in independently applying them to the millimeter region, coupled the resonator to a waveguide by a single hole. In principle, diffraction losses should be greater with this type of coupling, and probing the periphery of the field with absorbing objects tends to confirm this hypothesis. With both types of coupling we have obtained Q values within a factor of two or three of that computed for solid metal reflectors with no diffraction losses. It is probable that the slightly greater diffraction losses of the single hole feed are compensated by increased reflector losses with the multihole coupling. Unless a high uniformity of field is required for some special application, the single hole feed is to be preferred because of its simplicity.

The Properties of Thin Perforated Reflectors

With particular reference to the Stark spectrometer to be described below, we were concerned with the problem of obtaining good power transfer into the resonator while preserving high field uniformity and high reflectivity. In attacking this problem we adopted the policy of using simple approximate theory as a guide to the extent that such theory is valid and then of relying upon experiment to compensate for higher-order effects. One such theory is the impedance theory.

To check the validity of the impedance theory, we set up the transmission experiment shown in the upper part of Fig. 1. The perforated plate is placed between

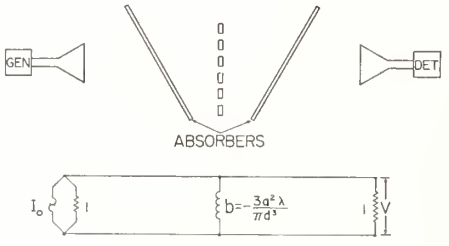


Fig. 1—Transmission through a lossless perforated thin plate.

coaxial transmitting and receiving horns. The signal received with the plate in the beam is compared to that with it absent. Standing waves between the plate and either horn are greatly reduced by placing absorbers at oblique angles on either side of the plate. A magazine is a convenient absorber, and the attenuation can be varied by choosing the number of pages. The significance of the experimental transmission coefficients was enhanced by separately varying the horn spacing and the plate position along the axis. These variations produced no more than 0.5 db change in the transmission coefficients. Transmission factor was taken to be the change in power at the detector as measured by a bolometer.

In the analysis of the experiment, it is assumed that the situation is described by a wave incident upon the plate, a wave reflected into the transmitter space, and a wave transmitted into the detector space. Furthermore, the simplifying assumption is made that these are all plane (TEM) waves of a single plane polarization. Since the beam is bounded in cross section, other modes must actually be present, but the experimental data show that they are of small amplitude. Such assumptions also imply that the plate does not depolarize the beam. In principle, if the rows of holes make oblique angles with the field vectors, depolarization can take place. We have observed no evidence of such depolarization nor any dependence of transmission factor upon orientation. Nevertheless, all measurements reported here were made with the rows of holes aligned with the field vectors.

If the space on the side of the plate away from the source is unbounded, if losses are negligible, and if only one mode is propagated at distances sufficiently removed from the plate, the plate can be represented by a susceptance in a transmission line analogy. Under these conditions, the power carried by the reflected wave on the side facing the source and the power carried by the transmitted wave on the other side must equal the power carried by the incident wave. In a transmission line such a condition is produced by connecting a susceptance across a line of infinite length. A mathematical proof seems hardly necessary to establish this equivalence, but such proofs can be found in textbooks.[10] However, a mathematical treatment is required to evaluate the susceptance. Such a one[11–13] shows that, if the holes are of diameter d and in a rectangular array with spacings a and c, the normalized susceptance is approximately

$$b = -\frac{3ac\lambda}{\pi d^3} . \qquad (1)$$

Eq. (1) is an approximation valid when a, c, and d are small compared to λ. Higher-order terms have been derived by Munushian,[13] who shows that (1) is the appropriate expression for an array of holes. It is interesting to note that when either a or c exceeds λ, the

[9] D. Marcuse, "Maser oscillation observed from HCN maser at 88.6 kMc," Proc. IRE (*Correspondence*), vol. 49, pp. 1706–1707; November, 1961.

[10] J. C. Slater, "Microwave Electronics," D. Van Nostrand Co., Inc., New York, N. Y., Sec. 6.5; 1950.
E. Ginzton, "Microwave Measurements," McGraw-Hill Book Co., Inc., New York, N. Y., Sec. 6.4; 1957.

[11] C. G. Montgomery, R. H. Dicke, and E. M. Purcell, "Principles of Microwave Circuits," McGraw-Hill Book Co., Inc., New York, N. Y., Sec. 6.11; 1948.

[12] N. Marcuwitz, "Waveguide Handbook," McGraw-Hill Book Co., Inc., New York, N. Y., ch. 5; 1951.

[13] J. Munushian, "Electromagnetic Propagation Characteristics of Space Arrays of Apertures-in-Metal Discontinuities and Complementary Structures," Electronics Research Lab., University of California at Berkeley, Lab. Rept., Ser. No. 60, Issue 126; September, 1954.

theory breaks down qualitatively as well, because the amplitudes produced by various holes interfere constructively to give maxima of radiation off the axis. Under these conditions, the plate acts as a grating producing more than one order of constructive interference.

In cases which are of interest in this work b is large in magnitude compared to unity. Then it can be easily shown that the transmission coefficient is

$$T = -20 \log_{10} \frac{2\pi d^3}{3ac\lambda} \qquad \text{expressed in decibels.} \quad (2)$$

The argument of the logarithm is twice the reciprocal of the susceptance. If the theory is valid, all experimental points should lie on a single straight line when the transmission coefficient is plotted against the logarithm of twice the susceptance. With a single plate, the susceptance is varied by varying λ. Such data are plotted along with the theoretical line in Fig. 2. With two plates, the points lie on lines of the same slope as the theoretical line but slightly below it. The latter fact can probably be attributed to losses in the plates, which are neglected in the theory. The curve representing the third plate crosses the theoretical line at high values of the argument (small values of λ). However, this plate had comparatively large spacings between the holes, and at short wavelengths the assumptions of the theory are not fulfilled. With all three plates, the holes were in a square array, and thus $c=a$. The data can be considered to be in reasonable agreement with the impedance theory. The use of this approximate theory as a guide in the design of interferometers is therefore justified.

The experimental data displayed in Fig. 2 were all obtained on thin metal foils stretched on frames: that is, foils whose thickness was small compared to a free-space wavelength. Brief qualitative consideration of transmission line theory indicates that the presence of lossless dielectric backing causes a reduced admittance mismatch, resulting in more energy being transmitted through the perforated plate than would be predicted from (2). When the frequency is such as to cause the thickness of the dielectric to be an integral number of dielectric half wavelengths, the transmission coefficient should agree with (2).

If a plane wave is normally incident upon a plane metal surface and if Γ is the reflection coefficient at the generator, the fraction of the incident power which is dissipated in the metal is

$$t_0 = 1 - |\Gamma|^2.$$

If a slab of lossless dielectric of thickness l and dielectric constant K is placed in contact with the metal on the generator side, the fraction of the power dissipated in the metal is increased unless, of course, the thickness of the dielectric happens to be exactly an integral number of half wavelengths.

In general, the fraction of power dissipated is

$$t = \frac{K}{1 + (K-1)\cos^2\theta}\, t_0, \qquad (2a)$$

where

$$\theta = \frac{2\pi l}{\lambda_1}.$$

t_0 is the value of t in the absence of dielectric, and λ_1 is the wavelength in the dielectric. Fig. 3 shows the relative power dissipated in the metal for three frequently used dielectrics as a function of θ. The following values of dielectric constants have been used: 1) plate glass 9.5, 2) quartz 3.78, and 3) rexolite 2.2.

The power transmitted through a perforated sheet is therefore increased by the factor t/t_0 given by (2a). The preceding discussion applies strictly to solid metal plates. However, our experience indicates that it holds as an excellent approximation for perforated plates.

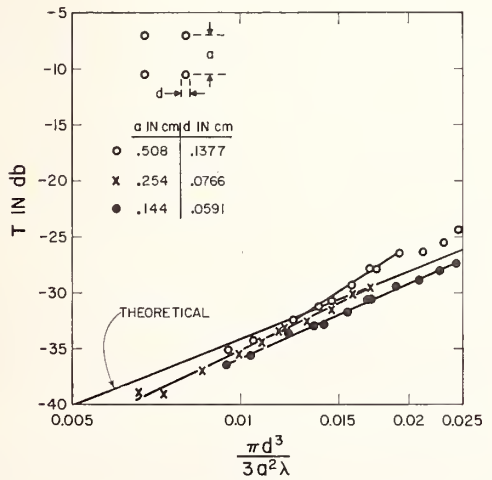


Fig. 2—Transmission measurements of three perforated thin plates.

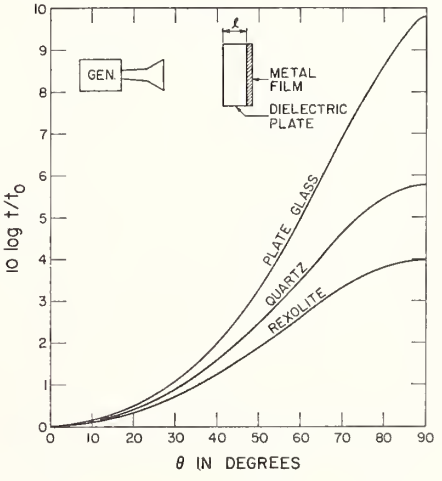


Fig. 3—Increase in absorption in metal film when deposited on a dielectric.

When the thickness of the metal plate is appreciable, it is represented by a three-terminal network, as is well known.[12] If the rows of holes are oblique to the field vectors such that depolarization effects must be considered, it can be supposed that a plate can be represented by a network having two pairs of input terminals and two pairs of output terminals.

The Resonant Frequencies of Parallel Plate Interferometers

In all practical applications of the Fabry-Perot interferometer, observation is made along the central axis peripendicular to the plates. What is considered as a resonance from the microwave point of view corresponds to a maximum in intensity at the center of the field of view in an optical interference pattern. In elementary texts, the behavior is explained in terms of the multiple reflection of plane waves between the reflectors. It is assumed that these plane waves have the same wavelength λ_0 and phase velocity as in an unbounded medium. If the plates are perfectly reflecting, the condition for resonance is then

$$\lambda_0 = (2D/m) \tag{3}$$

where D is the separation and m is any positive integer. The frequency is given by dividing λ_0 into the phase velocity.

Such a description is inadequate to explain all of the recent observations. Schawlow and Townes,[14] in proposing the use of the Fabry-Perot interferometer as a resonator for optical masers, suggested that it should be considered as a rectangular box with four open sides. A transmission line with both ends short circuited resonates at the same frequencies as when both ends are open circuited except that the positions of the nodes and antinodes are interchanged. Analogously, if the interferometer has rectangular plates, each of dimensions A and B separated by a distance D, it can be expected to resonate at the same frequencies as a closed rectangular box of the same dimensions where the free-space wavelengths are given very accurately by

$$\lambda_{m,n,p} = 2\left[\frac{m^2}{D^2} + \frac{n^2}{A^2} + \frac{p^2}{B^2}\right]^{-1/2}, \tag{4}$$

where m, n, and p are nonzero integers.

The theoretical work of Fox and Li[7] investigated the field patterns and showed that they do indeed differ from plane waves. Symmetry considerations require that n and p be odd integers for modes which are observed with coupling which is symmetrical about the central axis of the instrument. In the following we assume this symmetry.

In most practical situations, the second and third terms in the brackets of (4) are small compared to the first. In cases where they may be completely neglected, it can be seen that $\lambda_{m,n,p}$ becomes equal to λ_0 obtained in the elementary theory. In conventional optical situations these terms are generally so small that modes of the same m and differing n and p lie so close together as not to be resolved, and the elementary theory is adequate. However, with masers, the resolution is such that the frequencies emitted as the result of simultaneous oscillation in several of these modes can be resolved. The modes with the lowest values of n and p, namely unity, have the highest Q and give rise to the strongest maser lines. Also they generally produce the strongest and sharpest resonances in the interferometers described in the paper. The original report on the helium-neon gas maser[15] contains excellent experimental verification of the validity of (4).

In that paper the strong signals at 150-Mc intervals are due to beating of modes with different m's but all with $n=1$ and $p=1$. The weaker peaks displaced by 1.5 Mc are due to beats between a mode described by $n=1$ and $p=1$ and a mode with a different m and either n or p equal to 3 while the other of these two quantities remains equal to 1. Quantitatively these values are consistent with (4) and the geometry of the apparatus, in which $B=A$.

For many purposes it is convenient to employ an approximation for (4) by expressing D in terms of λ_0 by (3) and retaining only first-order terms in a binominal expression. B is set equal to A, since this condition usually prevails. Then

$$\lambda_{m,n,p} = \lambda_0\left[1 - \frac{(n^2 + p^2)\lambda_0^2}{8A^2}\right]. \tag{5}$$

Since n and p are never zero, the second term on the right is not identically zero and it must be considered if the Fabry-Perot resonator is used as a wavemeter of the highest available accuracy. λ_0 may be determined by measuring the displacement of one plate between major resonances. Then a correction can be determined by substituting this value into the second term of (5). A can be determined from geometry or by measuring the frequency shift between a main resonance for which $n=1$ and $p=1$ to a subsidiary one where one or both has a higher value.

The Q of Parallel Plate Resonators

The unloaded Q of a parallel plate resonator is given approximately by the following well-known expression[1]

$$Q = \frac{m\pi}{1 - |\Gamma|^2} \tag{6}$$

wherein Γ is the amplitude reflection coefficient of the surfaces. In (6) diffraction losses are neglected, and in the numerator an approximation has been made by

[14] A. L. Schawlow and C. H. Townes, "Infrared and optical masers," *Phys. Rev.*, vol. 112, pp. 1940–1949; December, 1958.

[15] A. Javan, W. R. Bennett, Jr., and R. Herriott, "Population inversion and continuous optical maser oscillation in a gas discharge containing a He-Ne mixture," *Phys. Rev. Lett.*, vol. 6, pp. 106–110; February, 1961.

setting $|\Gamma|^2 \sim 1$ in a more exact expression. This latter approximation is valid under all circumstances which are of interest. This equation has been derived on the basis of the elementary plane wave theory.

According to (6), Q should increase linearly with m (or D). However, at larger spacings diffraction becomes important and limits the Q obtainable. This situation is illustrated by the experimental data presented in Fig. 4. The data lie on a straight line until D becomes comparable to A, which is equal to 14 cm in this instrument. Then Q falls below the line. In this experiment the holes were such as to give weak coupling. The measured Q is therefore essentially the unloaded Q.

The straight line of greater slope in Fig. 4 represents the calculated value of Q if the plates were made of solid aluminum. The fact that the slope of the experimental line is a factor of 8 or so less indicates that the surface losses are correspondingly greater than for solid plates. This is not unreasonable because the plates used in this experiment had thin films barely one skin depth thick deposited on glass. The dielectric slab also increases the natural loss of the metal film as given by (2a). Data obtained in this laboratory on thicker plates indicate that surface losses are no more than 2 or 3 times greater than calculated for solid plates.

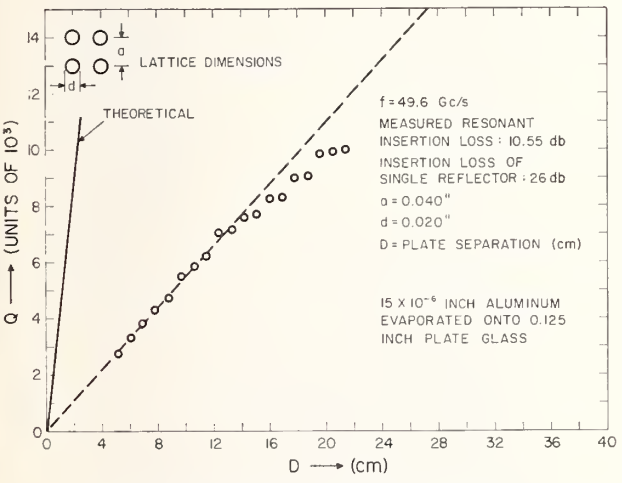


Fig. 4—The measured Q vs mirror separation of a parallel plate Fabry-Perot resonator.

Efficient Power Transfer

For many applications, it is necessary to obtain efficient power transfer into the resonator. With reaction resonators and a single horn serving as both transmitter and receiver, it is first necessary to get good power transfer in order to distinguish the resonance from a large background of reflected power. A transmission resonator is easier to adjust since input and output are naturally separated. If a transmission resonator is used in a maser, weak coupling to the transmitter and close coupling to the detector are desired for optimum sensitivity. Otherwise more of the signal developed by the sample between the plates is wasted in the generator.

In the former case, exactly, and in the latter case, approximately, we may consider one of the two plates to be opaque. To determine the conditions for optimum power transfer to the horn adjacent to the perforated plate, we apply the impedance theory mentioned earlier. This theory assumes the plane wave approximation. At first we shall neglect losses in the perforated plate. This assumption appears to be illogical since a perforated plate might be expected to be intrinsically more lossy than a solid one. However, later considerations will show that, contrary to our intuition, the losses due to this plate do not play an important role. When these losses are neglected, it is more convenient to work with admittance than impedance.

According to basic electromagnetic theory, a solid metallic plane surface at normal incidence can be represented as a normalized admittance of

$$y = g(1 - j) \tag{7}$$

where

$$g = \left(\frac{\sigma}{4\pi\epsilon_0 f}\right)^{1/2} \quad \text{and} \quad \mu = \mu_0$$

and where σ = conductivity of the metal, f is the frequency, and ϵ_0 is the permittivity of empty space. With frequencies in the millimeter wave region and with metals of good conductivity g is a large number, something between 10^3 and 10^4.

As the plane of reference moves away from the metal, the admittance changes in accordance with transmission line theory. It is then possible, in principle, to choose D in such a way that the normalized conductance at the reference plane is unity while the susceptance is positive. Then if the perforated plate is placed here and if the hole pattern is selected in such a way as to make the negative susceptance equal in magnitude to the positive transformed susceptance of the solid plate, a perfect admittance match is obtained.

Because g is so large the Smith Chart cannot be used. For the same reason the analytical expression can be considerably simplified. Applying standard transmission line theory and making various approximations, the following simple results can be obtained. These approximations include 1) neglecting terms of the order of unity and of the order of $\sqrt{g}$ in comparison with terms of the order of g, 2) approximating the tangent of the phase angle by the angle in radians, 3) retaining only lead terms in binominal expansions.

The required value of normalized susceptance for the hole pattern is

$$b = -(2g)^{1/2}. \tag{8}$$

The reflection coefficient of the solid metal surface is

$$\Gamma = \frac{1+j}{g} - 1 \tag{9}$$

and

$$|\Gamma|^2 = 1 - \frac{2}{g} \cdot \tag{10}$$

The required distance D is given by the roots of the equation

$$\tan \beta D = -(1/2g)^{1/2} \tag{11}$$

where

$$\beta = \frac{2\pi}{\lambda_0} \cdot$$

By use of (6) and (7) it can be shown that

$$Q = \frac{\pi m g}{2} \cdot \tag{12}$$

Up to this point losses in the perforated plate have been neglected. There is a theoretical argument which indicates that if the surface admittance of this plate is of the same order of magnitude as that of the solid plate, the effect of these losses is to make only minor changes in the hole diameter and plate separation required for optimum power transfer. Therefore, compensation for these effects can be accomplished by changing slightly these quantities by experiment. Since this theoretical argument is lengthy and probably of little interest, it will not be given here. However, our actual experience indicates that it is justified in practice.

A Parallel Plate Stark Cell

In the design of a Fabry-Perot interferometer for use in observing the Stark effect in millimeter wave molecular spectra we have applied the principles of the previous sections. This device is shown in Figs. 5 and 6. One plate is solid, and the other is composed of a copper foil 0.0015 inch thick with a square array of holes. Both are gold plated. The solid plate is mounted on insulators so that a dc or low-frequency ac voltage can be applied between reflectors.

The ring which supports the insulators is supported by three magnetostriction transducers. These employ nickel armatures 2 inches long and 0.25 inch in diameter. The magnetic circuits are completed with soft iron except for small air gaps. By dissipating a few watts of dc power in the transducer coils it is possible to nearly magnetically saturate the armatures producing a contraction of about 50 microinches. If all three coils are excited, the plate is translated an amount corresponding to detuning the cavity by about the 3-db bandwidth, thus achieving fine tuning. If the trans-

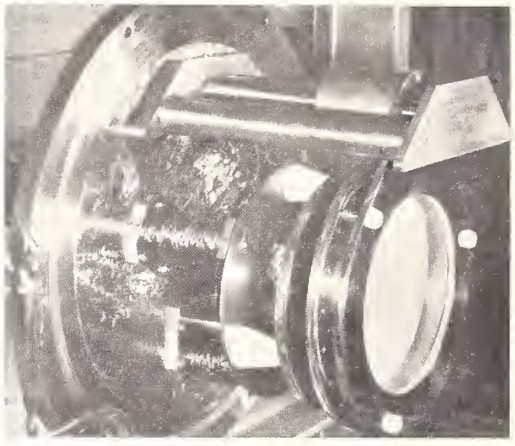


Fig. 5—Parallel plate Stark cell with vacuum cover removed.

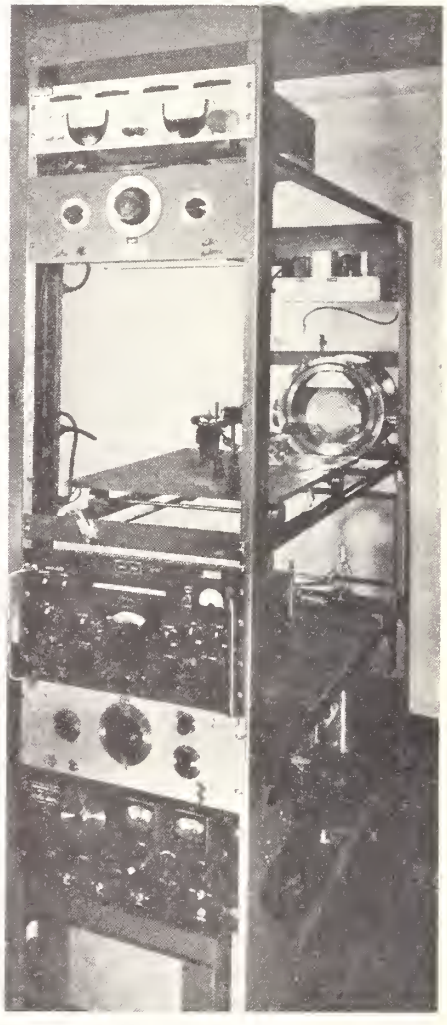


Fig. 6—Assembled view of parallel plate Stark cell showing associated equipment.

ducers are excited differently, this effect may be used to change the alignment. Coarse alignment adjustment is made by three screws holding the perforated reflector to the main frame. Coarse tuning is accomplished by a lead screw in the shaft from which the reflector is supported. The apparatus is provided with a gasketed cover having a plastic front. The lead screw extends through the vacuum envelope allowing coarse tuning adjustment even when the chamber is evacuated.

The foil is stretched on the frame before the insertion of the holes. For the insertion, it is then laid on backing material. The holes used in this interferometer were inserted by a specially made punch, but in other plates with larger holes, drills have been used.

By the use of measurements with calipers, the plates can be made sufficiently parallel to make resonances detectable. The final adjustment makes the resonances as sharp as possible. Putting the vacuum cover in place has only little effect on the response.

Fig. 6 shows the assembly of the interferometer and some of the associated equipment. The assembly employs a frame composed of two standard relay racks tied together by horizontal bars which holds the interferometer, vacuum system, and many of the associated electronic instruments. The horn which feeds the interferometer, as well as the klystron and the associated microwave components, are mounted on a shelf with wheels using two of the horizontal tie bars as tracks. By means of a lead screw adjustable at a panel on each relay rack, the horn-resonator spacing can be conveniently varied to obtain the optimum power transfer when the impedance match is not perfect.

This interferometer was designed with the intent of producing a uniform dc field between the plates. The array of holes occupies a 2-inch square. The horn is approximately 1.5 inches square. Practical experience as well as theory indicate that with close plate spacing the RF field is confined to a cross section slightly larger than the area of the array of holes. The dc field between the plates extends far beyond the RF field, assuring Stark field homogeneity within the active volume. The spacing between the plates can be varied between 7 mm and 45 mm. The widest spacing is slightly less than the size of the hole pattern. For a design center wavelength of 4.29 mm, the m values range from 4 to 10. Consideration of a number of conflicting factors leads to the choice of a moderately small spacing as preferable for the present application. Factors favoring small spacing are 1) relative freedom of inhomogeneity in the Stark field caused by edge effects and 2) relative freedom from pulling of the frequency of spectral lines by the response of the resonator. At close spacings the Q is low and the cavity resonance is broad. The factors favoring large spacing are 1) decrease in relative inhomogeneities in the Stark field caused by the holes and 2) high signal-to-noise ratio because of higher Q. It appears that the latter factors are less important than the former. However, the optimum spacing is to be determined by experiment.

The hole pattern was designed for operation at 3-mm wavelength. A spacing of 0.1 inch was selected as being a convenient value slightly less than the wavelength (0.118 inch). As mentioned earlier a spacing greater than a wavelength is undesirable because of the reinforcement of resonances with maxima off the axis. While the apparatus was being fabricated, it was decided to change the operating wavelength to 4.29 mm (0.169 inch). Under these conditions, for gold, (7) gives $g = 2.28 \times 10^3$, and (8) gives the required susceptance for the holes as -67.4. By (1) the required hole size is 0.029 inch. In order to test the theory the reflection coefficient was determined experimentally and it was found that holes with 0.1 inch spacing but with diameters of 0.037 inch yielded nearly 100 per cent absorption for values of $m = 5$ to $m = 15$ at a frequency of 70 Gc. Subsequent measurements at 55.2 Gc using the same hole pattern indicate an absorption of approximately 10 per cent. This corresponds to an SWR of 9:1 and gives some indication of the bandwidth characteristics of the array of holes as a coupling device.

Experimentally the degree of impedance mismatch is nearly independent of the horn-to-resonator spacing and of the spacing between reflectors. This latter fact is expected if the preceding theory is valid. If the shunt impedance is defined as the ratio of rms electric field strength to magnetic field strength at a reference plane where the latter is a minimum, it can be inferred that this quantity depends only on $|\Gamma|^2$ and not on m. Therefore it is not an explicit function of Q and increasing m does not change the field strength although it increases Q. In these respects this type of resonator differs from the conventional one.

If the above value of g is substituted into (12), Q/m is calculated to be 3.6×10^3. No attempt was made to make an accurate measurement, but a rough measurement indicated good agreement.

A discussion of the Stark effect for the measurement of voltage and the application of this instrument will be discussed by two of the present authors in another paper.[16]

Spherical Resonators

If a uniform Stark field is not needed and relatively narrow spectral lines are to be investigated, the high Q and small size of the spherical plate resonator recommend its use. The sensitivity of a resonant cavity to small changes in cavity loss is well known[17] to be $\Delta V/V = Q\lambda\alpha/2\pi$ where V is the voltage incident on the detector at cavity resonance and α is the free-space attenuation of the gas sample within the resonator. With a loaded Q of 10^5, it is practical to measure very small

[16] Y. Beers and G. Strine, "The Measurement of Voltage by Use of the Stark Effect, presented at Internat'l Conf. on Precision Measurements, Boulder, Colo.; August 14–17, 1962.

[17] C. H. Townes and A. L. Schawlow, "Microwave Soectroscopy," McGraw-Hill Book Co., Inc., New York, N. Y., Sec. 15-11; 1955.

values of α. The limitation imposed by restricting the resonance bandwidth of the cavity to be several times the natural linewidth of the gas sample is ameliorated by an absence of spectral line broadening due to collisions between the molecules and the cavity walls. The gas pressure can be reduced until pressure broadening is equal to Doppler broadening. For oxygen at room temperature this would occur at about 30 microns pressure and give a linewidth of some 0.25 Mc compared with a cavity bandwidth of 0.6 Mc if the Q were 10^5.

Possible saturation effects in a gas contained in a resonator with such a large Q are not as great as Q alone might imply. As pointed out earlier, in this type of cavity the electric field is not enhanced in the same way as in a simple R-L-C equivalent circuit. From the definition of Q and (6) it can be shown that for a closely coupled resonator with a loaded Q of Q_L the ratio of the electric field within the cavity to that in the transmission line driving it is approximately $(Q_L A_0/A_r m\pi)^{1/2}$, where A_r and A_0 are the effective areas of the resonator and input transmission line, respectively. This ratio is about unity in the spectrometer to be described.

The spherical plate spectrometer shown in Fig. 7 was designed to isolate the resonator as well as possible from any mechanical forces which might tend to deform it when evacuated. This allows it to be used as a refractometer by measuring the detuning of the cavity as gas is admitted to different pressures. The resonator itself consists of a spherical brass surface of 20-inch radius of curvature facing a flat brass surface nominally 10 inches away. The waveguide feed terminates in the center of the flat plate with an 0.063-inch hole coupling the waveguide to the resonator. The spherical surface is supported on three legs above the flat surface and can be screwed along its axis for tuning. This adjustment can be made through the vacuum container by a retractable finger which can be disengaged to avoid communicating forces to the structure when the pressure is changed. The open aperture of the mirrors is 4.5 inches and the whole resonator fits within a 14-inch length of 6-inch O.D. glass or plastic pipe.

At the top end of the transparent vacuum container a scale is affixed in order to record the axial position of the spherical mirror. The screw thread is metric so that the spectrometer serves also as a precision wavemeter when required. Moving the mirror between two resonances is a translation quite close to a half wavelength which can be read directly in millimeters on the scale.

The brass surfaces were finished to a high polish and precision of a few ten thousandths of an inch. The loaded Q is close to 10^5 at wavelengths of 5 mm. The parameter $a^2/b\lambda$ used by Fox and Li to compute diffraction losses is 1.3 for 5-mm wavelength, suggesting that higher modes than the fundamental TEM_{q00} may be supported. This notation was introduced by Fox and Li.[7] Indeed the axially symmetric TEM_{q01} mode has been identified and behaves qualitatively in all respects as predicted by Fox and Li.

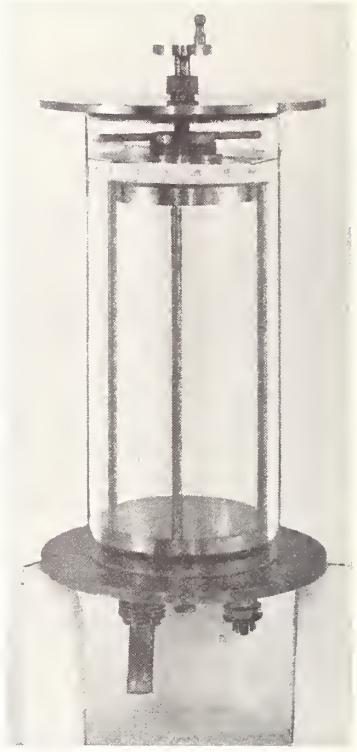


Fig. 7—Spherical plate spectrometer inside transparent vacuum cover.

A spherical plate resonator of small dimensions has been constructed specifically for use as a wavemeter over the waveguide band 50 to 75 Gc. The radius of curvature of the spherical mirrors is only 2 inches, about 10 wavelengths, but it performs quite well. The details of this instrument are discussed elsewhere.[18]

Acknowledgment

The authors want to express thanks to their many colleagues both within and without this laboratory for the many stimulating discussions they have had. In particular we want to express our appreciation to M. Humpal and G. Wichman for the painstaking efforts they took in the fabrication of the resonators discussed here.

[18] R. W. Zimmerer, "New wavemeter for millimeter wavelengths," *Rev. Sci. Instr.*, vol. 33, pp. 858–859; August, 1962.

Reprinted from IEEE TRANSACTIONS
ON MICROWAVE THEORY AND TECHNIQUES
Volume MTT-11, Number 2, March, 1963

Spherical Mirror Fabry-Perot Resonators*

ROBERT W. ZIMMERER†, SENIOR MEMBER, IEEE

Summary—An experimental investigation of the Fabry-Perot Interfometer (FPS) using spherical mirrors is reported. The FPS was operated as a microwave resonant cavity at 60 to 70 Gc. Measurements were made of the loss and coupling as a function of mirror spacing. The electric field variation within the resonator was also measured. Other characteristics of the spherical Fabry-Perot resonator were observed and are discussed.

A qualitative discussion of the behavior of a spheroidal cavity resonator is presented and its relation to the FPS and beam waveguide is demonstrated.

Introduction

THE SPHERICAL mirror Fabry-Perot Interferomter (FPS) was first introduced as a new optical instrument by Connes [1]. In a series of papers [1], [2], he developed a geometrical optics theory and application of the instrument. With the advent of the laser the FPS was employed as a resonator and an electromagnetic theory of its operation was developed by several investigators at Bell Telephone Laboratories [3]–[7]. In a parallel development of the beam waveguide for the transmission of quasi-optical microwave power, Goubau and his associates have developed an electromagnetic theory [8] which has many applications to the FPS. The application and experimental verification of these various theories has been most rapid. The direct observation of laser output [9], the successful operation of the microwave FPS [10], [11], the transmission line studies of the beam waveguide [12]–[15], all verified the theoretical soundness of the work.

We have constructed and operated a variety of microwave Fabry-Perot resonators of both planar and spheri-

* Received April 29, 1963; revised manuscript received June 20, 1963. A preliminary note on this work has been published by the author, "Experimental investigation of Fabry-Perot interferometers," PROC. IRE (*Correspondence*), vol. 51, pp. 475–476; March, 1963.

† National Bureau of Standards, Boulder, Colo.

cal type [16]. Experimental investigation of the FPS at microwave frequencies is particularly attractive because of the comparative ease with which direct measurements of phase and amplitude can be made. We have made our measurement at wavelengths between 8 and 4 mm where klystron generators of good stability and power output are conveniently available. At these wavelengths not all the restrictions of the optical theory are satisfied while the theory of the beam waveguide is not yet extensive enough to include all aspects of the microwave FPS.

Evolution of the FPS from a Spheroidal Resonator

In our work with the FPS we have found it quite instructive to consider it as being evolved from a microwave cavity formed by rotating an ellipse about its minor axis. Such a closed surface with rotational symmetry is a special form of ellipsoid called an oblate spheroid [17]. This surface can be described by the set of equations

$$z = \frac{b}{2}\xi\cos\theta$$

$$r = \sqrt{x^2+y^2} = \frac{b}{2}\sqrt{1+\xi^2}\sin\theta \qquad (1)$$

where z is measured along the axis of rotation and r perpendicular to it.

For a given b there is an oblate spheroid for every value of ξ between 0 and ∞ and a hyperboloid for each value of θ between 0 and $\pi/2$. These confocal oblate spheroids and hyerboloids constitute an orthogonal coordinate system in which Maxwell's equations can be expressed. If the wave equation is expressed in this coordinate system and the eigenfunctions are examined, a most remarkable behavior is found [18], [19]. The eigenfunctions are significantly different from zero only in the neighborhood of $\theta\sim 0$. The θ for which each eigenfunction decreases to a certain fraction of its maximum value, increases with the order of the eigenfunction and with the ratio λ/b. Thus for an oblate spheroidal resonator $b\gg\lambda$, the portion of the cavity walls far off the axis of symmetry play no appreciable role in the boundary condition that the field should vanish at the oblate spheroidal surface. Only an area on the z axis ($\theta=0$) is pertinent to this boundary condition. If the remainder of the spheroidal surface is discarded, what is left is a pair of approximately spherical mirrors. By virtue of the orthogonal coordinate system of ξ and θ, we know that the spheroidal surfaces within the resonator are surfaces of constant phase and the hyperboloids are surfaces of constant amplitude. If $b\gg\lambda$, the solution is nonzero only for θ near zero and the surfaces of constant phase are paraboloids. An examination of (1) shows that if $b\to 0$ in such a way that $b\xi/2\to\rho$ a spherical coordinate system results.

In the general case of a spheroidal cavity it is necessary to know the curvature of the approximately spherical portion on the z axis. This can be calculated from the generating ellipse. The radius of curvature of an ellipse on the minor axis is given by the square of the major axis divided by the minor axis. From (1) it follows that the radius of curvature b' of the spheroidal surface at $\theta=0$ is

$$b' = \left|\frac{1+\xi^2}{\xi}\right|\frac{b}{2}, \qquad (2)$$

which can be recognized as (23) of Boyd and Gordon [5]. The relationship of the oblate spheroidal resonator to the FPS now becomes clear. The family of oblate spheroids belonging to a pair of foci of separation b generates all the possible combinations of spherical mirrors of radius of curvature b_1 and b_2 spaced a distance d. This is illustrated in Fig. 1.

$$d_1 = b_1 \pm \sqrt{b_1^2 - b^2}$$

$$d_2 = b_2 \pm \sqrt{b_2^2 - b^2}. \qquad (3)$$

In the actual operation of the FPS the parameter b is not as useful as the spacing d of the mirrors.

Eliminating b from (3) we have

$$\frac{d_1}{2} = \frac{d(b_2-d)}{b_1+b_2-2d}$$

$$\frac{d_2}{2} = \frac{d(b_1-d)}{b_1+b_2-2d} \qquad (4)$$

where $d_1+d_2=2d$. In this discussion it is also useful to express b in terms of the physical parameters of the FPS,

$$\frac{b}{2} = \frac{\sqrt{d(b_1-d)(b_2-d)(b_1+b_2-d)}}{b_1+b_2-2d}. \qquad (5)$$

In this discussion the convention that $b_1<b_2$ is employed. As developed in part VI of Boyd and Kogelnik [6], all combinations of radii of curvature and mirror spacing are not resonant structures. From (5) it can be seen that the interfocal distance of the prototype spheroidal resonator becomes zero at four different values of the mirror spacing d. The limit $b=0$ results in a spherical coordinate system which has the familiar Bessel functions for eigenfunctions. For $b_1<d<b_2$ and $b_1+b_2<d$, b is imaginary. These are the high loss regions.

The upper diagram in Fig. 1 shows the general case for nonidentical mirrors. Note that convex mirrors, as well as concave, are allowed. The emphasized ellipse has the minimum radius of curvature on the axis and determines the special case of confocal identical mirrors symmetrically spaced about the origin. The unique degeneracy of the confocal placement of identical mirrors can be seen by noting that, from (3), any oblate spheroid with $b<b'$ will generate the required confocal mirror surfaces. In the limit $b\rightarrow 0$ the spheroid becomes a sphere with its center at one mirror or the other. This limiting case for nonidentical mirrors is illustrated in Fig. 1(b).

The general eigenfunctions for the oblate spheroidal resonator and in particular their asymptotic form for $b\gg\lambda$ are discussed in much detail by Flammer [18]. If $b\gg\lambda$ the angular functions can be expressed in terms of the Laguerre polynomials. In the limit $b/\lambda\rightarrow\infty$ the expansion contains only the first term. This is the solution obtaining for paraboloidal coordinates as developed by Pinney [20]. In the optical limit with $b/\lambda\sim 10^7$ the spherical mirror is indistinguishable from a paraboloidal mirror; in the microwave region with $b/\lambda\sim 10^2$ there is a measurable difference between these two mirrors.

In the experiment described by Christian and Goubau [12] at 24 Gc, lenses corresponding to parabolic mirrors were used and the electric field variation across the aperture was observed to be a smoothly changing function with a single maximum on the axis of the beam waveguide. In measurements of the field variation within a confocal parabolical resonator at 9.3 Gc, Beyer [14] reports very good agreement within the eigenfunction of the dominant lowest-order mode of the beam waveguide. The electric field variation measured perpendicular to the axis of a FPS at several axial positions is shown in Fig. 2. In this measurement the frequency was 61 Gc and b/λ was 115. (A discussion of the experimental method used to record these measurements is deferred to later.) It is evident that at least a fourth-order symmetric field distribution is present in what is the lowest mode of this FPS. The flat mirror was a highly polished brass surface and the spherical mirror a polished quartz mirror of optical quality with an evaporated copper surface. It is interesting to note that an equivalent spherical mirror of brass did not produce nearly as sharp a field structure. In all probability this was due to the less precise surface figure of the brass mirror which tended to "blur" the distinction between sphere and paraboloid.

The complex field structure of the FPS at microwave frequencies may be a disadvantage in some applications but this is strongly compensated by the ease with which it is aligned. The spherical surface of the mirror can be defined by a point, the center of curvature. The axis of the FPS is the line containing the two centers of curvature. If this axis intersects each mirror near enough to its center so that the electromagnetic fields are confined

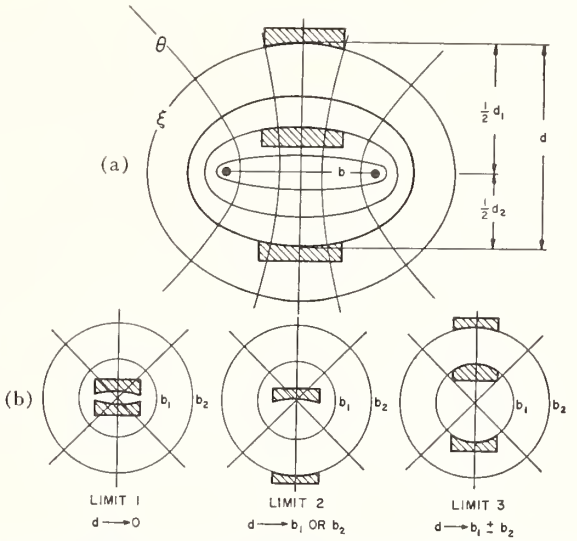


Fig. 1—Spherical mirror Fabry-Perot resonator derived from confocal oblate spheroids. (a) General case. (b) Limiting cases.

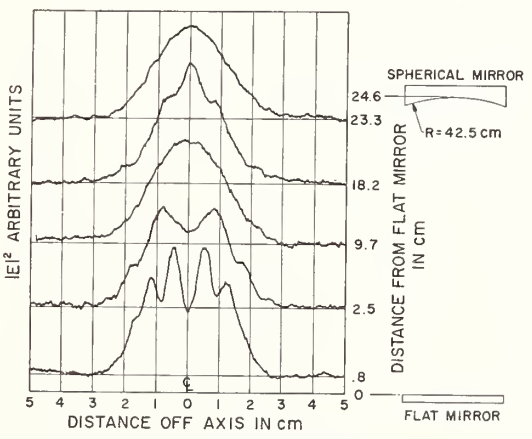


Fig. 2—Electric field variation perpendicular to resonator axis and electric field.

entirely within the mirror aperture, the FPS is aligned. A small rotation of one mirror can be corrected by an axial translation. This is not the case with either the planar Fabry-Perot or the parabolic mirror resonator. These mirror surfaces are characterized by an axis or normal. Alignment is only achieved when the axis of the paraboloids coincide or the normals to the plane mirrors are parallel. A rotation of one mirror cannot be compensated by an axial translation. Resonators using at least one plane mirror can be displaced laterally keeping their axes parallel but two paraboloids have only one relative placement for correct alignment.

In our laboratory we have experienced the difficulty of aligning flat mirrors to make a microwave planar Fabry-Perot resonator. Optical instruments were finally employed to simplify the tedious process. Beyer [14] re-

ports similar difficulty with paraboloids. In sharp contrast, the FPS alignment procedure is trivial. This problem of precise alignment is carried over to the beam waveguide where it affects the stability of a long transmission path containing many lenses.

The Analogy to the Beam Waveguide

Another aspect of the stability problem arises from the practical limitations of making mirrors or lenses identical. The FPS with different focal length mirrors is equivalent to a beam waveguide composed of a periodic sequence of lenses of alternating focal lengths. The transmission line analog was employed by Fox and Li [4] in their numerical calculation of the losses and mode structure of a variety of resonators. The converse analog was employed by Christian and Goubau [12] when they used a resonator to measure the losses of a beam waveguide. Such a transmission line analog is depicted in Fig. 3. One condition for stable operation would require that any image plane should recur periodically at the same relative position. This is illustrated by two planes, each at a distance x from two successive lenses of the same focal length. Using the thin lens equation of geometrical optics, $1/f=1/P+1/S$, and the relation between the focal length and radius of curvature of a mirror, $1/f=2/R$, the position on the optical axis of the image point is

$$x(1,2) = \frac{d(R_2-d) \pm \sqrt{-d(R_1-d)(R_2-d)(R_1+R_2-d)}}{R_1+R_2-2d} \quad (6)$$

where 1 and 2 refer to the two different lenses. It is apparent that both real and complex x are obtained depending upon the spacing of d. For the complex roots of x, where the complex conjugate of x is denoted by $\bar{x}$, $x(1,2)+\bar{x}(2,1)=d$. Referring to (4) and (5) the roots of (6) can be written

$$\begin{aligned} 2x(1,2) &= d_1 + ib \\ 2\bar{x}(2,1) &= d_2 - ib. \end{aligned} \quad (7)$$

This identification is shown symbolically in Fig. 4 where the elliptical section of a spheroidal resonator is superposed on the periodic sequence of lenses. This method could be extended to cover any reiterated sequence of lenses or mirrors of different focal lengths but it would not yield much more understanding of the phenomena. What is of great interest is the stability of such a transmission line when the lens spacing is nearly periodic and the focal lengths are randomly distributed about some design center as would occur in any practical application. To study this requires more sophisticated methods than geometrical optics. Similar stability problems have been considered by Pierce for electron beams [21] and beam waveguides [22] and occur in particle accelerators using periodically spaced magnets.

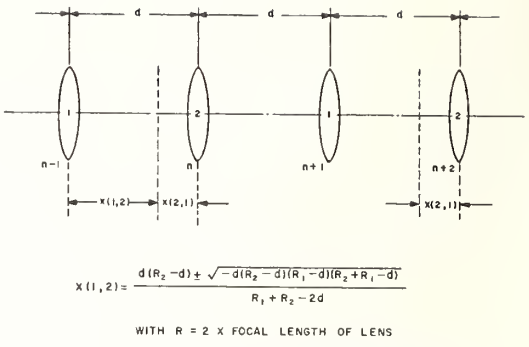


Fig. 3—Periodic sequence of simple lenses with alternating focal lengths.

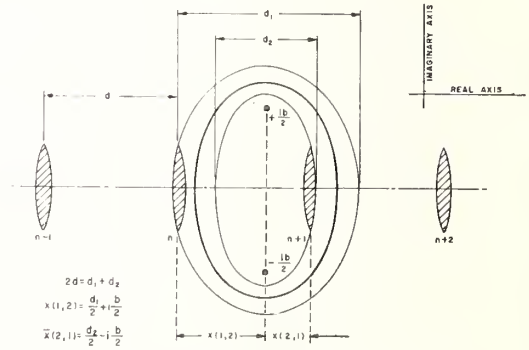


Fig. 4—Relationship of oblate spheroidal resonator to periodic sequence of lenses.

Measurement of Resonator Losses

Experimental measurements of the losses of the beam waveguide lens due to diffraction were made by Christian and Goubau [12], [13] and Beyer [14], and Beyer and Scheibe [15]. These experiments consisted of operating confocal paraboloidal resonators and measuring the Q of the resonator. We have measured the loss of the FPS as a function of mirror spacing for resonators using both identical and nonidentical mirrors. Several resonators were constructed using different combinations of spherical mirrors.

The FPS was operated as a microwave reaction cavity in order to have one mirror free to move. The fixed mirror was 12 cm in diamter with an RG-98u or RG-99u waveguide feed at the center. A coupling hole of diameter equal to the short waveguide dimension was drilled through the mirror. The iris was as thin as possible, about 0.010 inch. The brass mirrors were turned on a lathe and polished to a specular finish. They gave rather good optical images even though surface imperfections could be seen. The mirror figures were probably of the order of 0.001 inch out of a wavelength of 0.180 inch, which is comparable to the finest optical quality obtained at optical wavelengths. Several mirrors were quartz, ground and polished to optical standards and coated with 10^{-4} cm of copper by evaporation. The figure of these mirrors was good to $10^{-4}\lambda$ at microwave

frequencies. The alignment of the FPS was accomplished by autocollimation using a small flashlight. This was a simple procedure because the FPS mirrors produced a good image of the point source.

The Q of the FPS is so high, and stable klystron sources so difficult to operate, a dynamic method of measuring Q was employed. This method was only possible because of the inherent high stability of the particular klystrons used. The klystron was connected to the FPS through a 10-db 4-port directional coupler. The bolometer detector was connected to the same side of the directional coupler as the FPS and measured the power reflected back from the FPS-waveguide junction. The klystron repeller was modulated with a low-frequency sawtooth waveform which deviated the microwave frequency some 10 Mc. The bolometer output was monotored on an osciloscope. Frequency markers were generated by a second simultaneous modulation of the klystron repeller by an RF voltage of several Mc. Such RF modulation produces sidebands of known separation which are resolved by the resonator giving absorption dips on the oscilloscope. The RF modulation is used only to calibrate the sweep length and is not present during measurements of the FPS resonance width. A typical calibration was 2 cm/Mc deviation of the klystron center frequency. The FPS-resonance width varied from 0.5 Mc to several Mc. The ratio of power absorbed by the FPS to that incident on the coupling hole, $1-|\Gamma_c|^2$, was measured simultaneously with the resonance width. The use of a bolometer had the advantage that oscilloscope deflections were proportional to power. The signal-to-noise ratio was better by an order of magnitude than with a video crystal, although there was considerable variation among bolometers and crystals. A disadvantage of the bolometer is its long time constant which requires slow sweep speeds for faithful reproduction of the sharp resonance of the FPS. Some difficulty was experienced due to a background of microwave power which "leaked" into the bolometer. This leakage varied markedly with frequency and was always minimized before making measurements.

The Q of the FPS resonator, neglecting diffraction losses, is given by the well-known relation

$$Q = q\pi/(1 - |\Gamma_m|^2) \tag{8}$$

where $|\Gamma_m|^2 \sim 1$ is the power reflection coefficient of a single mirror and $q = 2d/\lambda$, the longitudinal mode number.

The mirror reflection loss is given by

$$1 - |\Gamma_m|^2 = 4r = 4\sqrt{\frac{\omega\mu}{2g}} \tag{9}$$

where r [23] is the real part of the normalized surface impedance of a metal, and ω, μ and g have their customary meaning. The calculated value of $1-|\Gamma_m|^2$ for a copper surface at 70 Gc is about 10^{-3}.

A set of measurements of $q\pi/Q$ vs d is shown in Fig. 5 for an FPS operated at 69.5 Gc. The upper curve shows the high loss region for $b_1 < d < b_2$. The lower curve was measured using one flat mirror because of the difficulty of making two brass mirrors of identical curvature. The aperture of the movable flat mirror was large enough so that diffraction losses were due entirely to the spherical mirror. The measurements were plotted against $2d$ for ease of comparison. An attempt was made to fit a theoretical curve to the results. It was impossible to get a good fit with the upper curve but the lower curve was closely described by assuming an effective value of $a^2/b\lambda = 1.38$ instead of the actual value of 1.64.

The corresponding measurements of $1-|\Gamma_c|^2$ are shown in Fig. 6. A surprising feature is the apparent low coupling to the FPS in the far region, $d > b_2$.

To better interpret Figs. 5 and 6 it is useful to propose an equivalent microwave circuit for the general FPS. An equivalent series resonant circuit is given in Fig. 7. The FPS may be regarded as a section of beam waveguide transmission line short-circuited at each end. The characteristic impedance of this transmission line is Z_2, the voltage attenuation constant is α, the impedance of each mirror is $(1+j)R$ ohms and the reactance of the coupling hole is jX. Elementary transmission line theory [24] shows that a low but finite impedance transforms into itself in moving q half wavelengths along a low loss transmission line. The resonant nature of the length $q\lambda/2$ of the transmission line is represented by the equivalent lumped reactance and capacitance of equal magnitude $q\pi Z_2/2$. The transmission line equivalent resistance is $\alpha d Z_2$ ohms. The coupling hole transforms the input waveguide impedance into a very low impedance X^2/Z_1. The Q of a series resonant circuit is given by the quotient of the magnitude of the inductive reactance and resistance at resonance. Assuming that $Z_1 \sim Z_2 \sim Z_0$ of free space and neglecting the small quantities jX and jR compared with $q\pi Z_2/2$, we get the expected result [23].

$$Q = \frac{q\pi}{4r + 2\alpha d + 2x^2} \tag{10}$$

where

$$r = \frac{R}{Z_0} \quad \text{and} \quad x = \frac{X}{Z_0},$$

which reduces to (8) for $\alpha = x = 0$. In these experiments the FPS was undercoupled and $2r \sim 5x^2$. The quantity $1-|\Gamma_c|^2$ can be expressed in terms of the equivalent microwave circuit as

$$1 - |\Gamma_c|^2 = \frac{4z}{(1+z)^2} \tag{11}$$

with

$$z = \frac{2r + \alpha d}{x^2}. \tag{12}$$

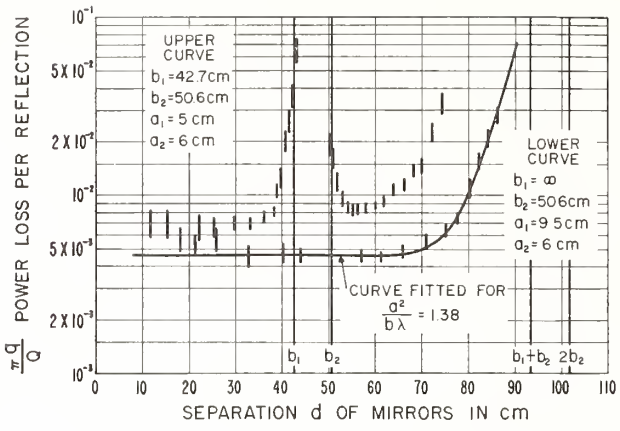


Fig. 5—Power losses in two Fabry-Perot resonators

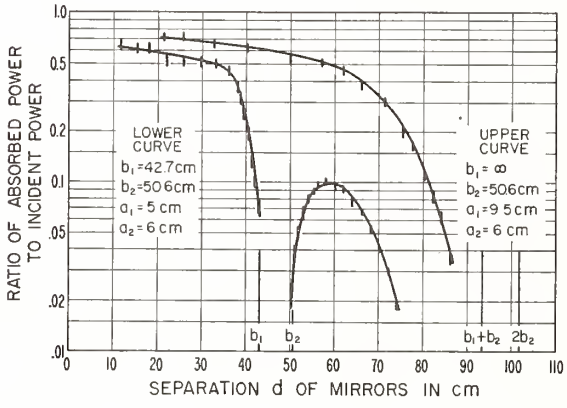


Fig. 6—Power coupling into spherical Fabry-Perot resonators.

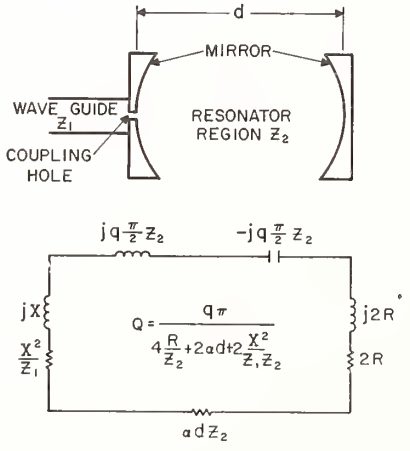


Fig. 7—Equivalent series resonant circuit of Fabry-Perot resonator.

The term αd is entirely due to absorption in the gas contained between the mirrors. At 70 Gc the normal atmospheric loss is about 0.5 db/km ($\alpha \sim 10^{-7}$ cm^{-1}) and rises as high as 15 db/km in the middle of the O_2 absorption band at 60 Gc [25]. In these measurements αd was too small to be resolved as an independent contribution to the resonator loss. The diffraction loss, α_D in the notation of Boyd and Gordon [5], was measured. Rewriting (10) in this notation,

$$Q = \frac{\pi q}{\alpha_R + \alpha_D + 2x^2} \tag{13}$$

and α_R is identified as $4r$. Rewriting (11) we get for large values of $\alpha_D + \alpha_R$,

$$1 - |\Gamma_c|^2 = \frac{(8x^2)}{\alpha_R + \alpha_D} \tag{14}$$

which is the same form as the expression for $Q/\pi q$. Since the measured value of both $1-|\Gamma_c|^2$ and $Q/\pi q$ are plotted on semilog paper, the curves should be quite similar in regions of high loss or low coupling. By superposing Figs. 5 and 6 it can be observed that this indeed is the case.

In fitting a theoretical curve to the data the expression for α_D, suggested by Boyd and Gordon [5], was used. This is

$$\alpha_D = A10^{-BN}. \tag{15}$$

They give the values of $A = 10.9$ and $B = 4.94$ for the TEM_{q00} mode with the square mirror. It was found that better argeement was obtained using the values $A = 29$ and $B = 4.83$ which fit the curve of α_D as given by Beyer and Scheibe [15] and Fox and Li [4] for the TEM_{q00} mode with a circular mirror. From (43) of Boyd and Kogelnik we obtain

$$N_1 = \frac{a_1^2}{b_1\lambda}\left[\frac{2}{s_1} - 1\right]^{1/2}$$

$$N_2 = \frac{a_2^2}{b_2\lambda}\left[\frac{2}{s_2} - 1\right]^{1/2}, \tag{16}$$

with $s_1 = d_1/b_1$, $s_2 = d_2/b_2$, and d_1 and d_2 defined by (3) and (4). Because only one mirror contributed to the diffraction loss in the FPS when using one flat and one spherical mirror, (13) was modified to give

$$\frac{\pi q}{Q} = \alpha_R + \tfrac{1}{2}\alpha_D + 2x^2 \tag{17}$$

as the equation of the curve with α_D determined from (15) and (16). The measured value 4.6×10^{-3} was used for the quantity $\alpha_R + 2x^2$. This compares with $\alpha_R = 10^{-3}$ computed for copper by (9).

It is expected that an extremum should be observed in the far region for both $q\pi/Q$ and $1-|\Gamma_c|^2$ for $b_1 \neq b_2$. From (3) and (16) it can be seen that this should occur when $b=b_1$, whereupon $2d=b_1+b_2+\sqrt{(b_2^2-b_1^2)}$. For these measurements the extremum should occur at $d=60.3$ cm. This agrees well with the $1-|\Gamma_c|^2$ data but not with the $q\pi/Q$ measurements.

An interesting feature of these data is that the coupling to the FPS is lower in the far region, $d>b_2$ than in the near region $d<b_1$ by an order of magnitude. If the FPS axis moved away from the coupling hole in going from $d=40$ to $d=60$ cm, the field of the resonator at the coupling hole could be smaller than the maximum and account for the decreased coupling. Attempts to realign the FPS have shown this not be to the explanation. This effect is real and occurs in all the FPS of this type we have operated. In view of the complex field structure with in the FPS, a possible explanation is a relative decrease of field strength on the axis compared to the average field across the aperture as d increases. This would account for the gradual decrease in coupling shown in the upper curve of Fig. 6 as d goes from 0 to 60 cm.

Measurement of Field Distribution

At this point it is appropriate to mention that the field of the FPS was measured by moving a small lossy paper disk across the mirror aperture and recording the change in power absorbed by the cavity $(1-|\Gamma_c|^2)$. If the object is of small volume compared with the active resonator volume, low loss compared with the natural resonator losses, and dielectric and magnetic properties nearly that of free space, it will have very little perturbing effect on the field and can be represented as an added series resistance in the equivalent circuit. The equivalent resistance will be very nearly proportional to the square of the local electric field. The reflected resistance z of the FPS at resonance normalized to the waveguide impedance varied from about 6 to 9 in these field measurements as the probe moved from zero to maximum field locations. From (11) it can be found that $1-|\Gamma_c|^2$ is linear to 4 per cent for this change in z.

The paper disk was supported by a nylon thread passing through its center and wound on the shaft of a 0.1-per cent helipot. As the helipot shaft was rotated the paper disk moved across the mirror aperture at right angles to the plane of the electric field. A voltage developed by the helipot referenced its lateral position. The paper disk was about 2λ in diameter. To use an unstabilized klystron with a FPS of $Q\sim 10^5$, the klystron was swept in frequency as previously described. The peak of the bolometer output corresponded to $1-|\Gamma_c|^2$ at resonance. Detuning effects and sweep to sweep frequency changes were greatly minimized by using a peak detector to rectify the ac signal from the bolometer. Thus two dc voltages were derived to operate an XY recorder which made the traces of Fig. 2. The complex field pattern of Fig. 2 is strongly dependent on the mirror spacing. This phenomenon appears to be real and not caused by the presence of the probe within the resonator. Further study of this is being undertaken.

Resonance Within the High Loss Region

The experiments indicate that the resonances persist into the high loss region where the present FPS theory is not valid [26]. It is of interest to estimate how close to the high loss regions the approximate theory should be valid. To do this we might say that b, the interfocal distance of the spheroidal prototype resonator, should be at least 20λ. If we let d approach to within ϵ, the 4 limiting mirror separations, we obtain for ϵ the following minimum values:

$$\epsilon = 100\lambda^2\left(\frac{1}{b_1}-\frac{1}{b_2}\right) \qquad \begin{matrix} d\to b_1-\epsilon \\ d\to b_2+\epsilon \end{matrix} \tag{18}$$

$$= 100^2\left(\frac{1}{b_1}+\frac{1}{b_2}\right) \qquad \begin{matrix} d\to \epsilon \\ d\to b_1+b_2-\epsilon. \end{matrix} \tag{19}$$

For this FPS $\epsilon=0.16$ cm, $\epsilon=2.0$ cm, a rather close approach to the boundaries of the high loss region.

Our work demonstrates that within the central high loss region there are two TEM_{q00} resonances present which can be identified as belonging to the two low loss regions. They occur at slightly different values of the mirror spacing d. The one is decaying as the other is growing with increasing d. It seems reasonable to define a wavelength within the resonator to be twice the distance the mirrors must be displaced to change q by unity. From (47) of Boyd and Kogelnik we have

$$\frac{q}{2}=\frac{d}{\lambda}-\frac{1+m+n}{2\pi}\cos^{-1}(1-s) \tag{20}$$

where $s=d/b'$. Calculating $\Delta q/2\Delta d$ we get the Taylor Series expansion.

$$\frac{\Delta q}{2\Delta d}=\frac{1}{\lambda_g}=\frac{1}{\lambda}-\frac{1+m+n}{2\pi b'\sqrt{s(2-s)}} \cdot\left\{1-\frac{1-s}{s(2-s)}\,\frac{\lambda_g}{4b'} +\frac{2s^2-4s+3}{s^2(2-s)^2}\,\frac{\lambda_g^2}{24b'^2}+\cdots\right\} \tag{21}$$

where Δd has been put equal to $\lambda_g/2$. Near the confocal spacing, $s\sim 1$ and (21) simplifies to

$$\lambda_g=\lambda\left[1-\frac{1+m+n}{2\pi}\,\frac{\lambda}{b'}\right]^{-1}. \tag{22}$$

If λ_g is computed for two different FPS resonators each with identical mirrors of radius of curvature b_1 and b_2 spaced approximately confocally, we obtain a fractional difference in wavelength given by

$$\frac{\Delta\lambda_g}{\lambda} = \frac{\lambda}{2\pi}\left(\frac{1}{b_1} - \frac{1}{b_2}\right) \tag{23}$$

where $q = (2d/\lambda) \gg 1$ and λ is very nearly equal to λ_g. Assuming that the two nearly coincident TEM_{qoo} resonances within the high loss region belong to the same q, but to slightly different values of λ_g, one can compute the fractional difference in wavelength from (23) to be

$$\Delta d = q\frac{\Delta\lambda_g}{2}$$

or

$$\frac{\Delta\lambda_g}{\lambda} = \frac{\Delta d}{d}. \tag{24}$$

Measurements on several FPS gave excellent agreement with this interpretation. Δd was of the order 0.004 inch, just about the limit for reading the micrometer mirror adjustment. The assignment of wavelengths to the two low loss regions was in agreement with (22), the near region $d < b_1$ having the longer λ_g and the far region $d > b_2$ having the shorter λ_g.

The definition of λ_g as given by (22) has been verified to within the error of experimental measurements of λ_g by simultaneous measurement of the frequency to some six places. For the larger FPS the free-space wavelength λ agreed with the resonator wavelength λ_g to the 3 figure precision of the λ_g measurement. The FPS has been used with confidence in these measurements to determine the operating wavelength of the klystron. It is surprising that (22) appears to hold even for the small wavemeter constructed for RG-98/u waveguide [26]. In that instrument the radius of curvature of the mirrors is only 10λ and since it is operated confocally b of (5) is also 10λ.

Higher-Order Modes

Qualitative verification was made of the existence of higher-order TEM_{dmn} modes. These were quite conspicuous at close mirror spacing and in fact could be easily confused with the dominant TEM_{qoo} mode because of their prominence. At no time could clear evidence be found of the axially symmetric "bulls eye" modes. These are predicted for circular mirrors from beam waveguide theory [8] as well as being one of the results obtained by Fox and Li [4] (Boyd and Kogelnik [6] compare these formulations in the appendix to their paper). The mirrors could always be readjusted to resolve a single resonance peak of a higher-order mode into two or more individual resonances. These multiplets could be shown to have different numbers of maxima along the electric field and at right angles to it. In this respect they conformed to the square mirror theory of Boyd and Gordon, with the degeneracy in m and n removed by a small difference in the radius of curvature in the x-z and y-z planes.

The true surface of a real mirror can be considered to have one or more zones where the curvature is constant and each zone would form its own FPS. If the wavelength is such that the field overlaps many zones, multiplets might well occur of the type observed in these experiments. As an approximation one could replace the oblate spheroid prototype FPS by an ellipsoid with all three axes different. The FPS mirror would have two principal curvatures along orthogonal diameters. This would resemble the prototype strip mirrors of Boyd and Gordon where the square mirrors are resolved into two strip mirrors along orthogonal directions. The strip mirror in turn can be considered to be generated by an elliptical cylinder. The solution of the resonator or waveguide problem for an elliptical cylinder with interfocal distance large compared to the wavelengths gives the eigenfunctions of the FPS [19].

Many variations of the FPS have already been considered by those working with them such as using two cylindrical reflectors at right angles (circular, elliptical, or parabolic), or unfolding the resonator to make a beam waveguide using mirrors instead of lenses. Undoubtedly many more applications of the FPS will be invented as microwave technology makes increasing use of physical optics.

References

[1] P. Connes, "Enhancement of the product of luminosity by resolution of an interferometer using a path difference independent of incidence," *Rev. Opt.*, vol. 35, pp. 37–43; January, 1956.

[2] ——, "The spherical Fabry-Perot etalon," *J. Phys. Rad.*, vol. 19, pp. 261–269; March, 1958.

[3] Since this work was presented at the Orlando Conference, a further paper by A. G. Fox and T. Li, "Modes in a maser interferometer with curved and tilted mirrors," Proc. IEEE, vol. 51, pp. 80–89; January, 1963, has appeared which discusses the nature of the resonances within the high loss region. These are not allowed in the approximate theory of Boyd and Kogelnik.

[4] A. G. Fox and T. Li, "Resonant modes in a maser interferometer," *Bell Sys. Tech. J.*, vol. 40, pp. 453–488; March, 1961.

[5] G. Boyd and J. Gordon, "Confocal multimode resonator for millimeter through optical wavelength masers," *Bell. Sys. Tech. J.*, vol. 40, pp. 489–508; March, 1961.

[6] G. Boyd and H. Kogelnik, "Generalized confocal resonator theory," *Bell Sys. Tech. J.*, vol. 41, pp. 1347–1369; July, 1962.

[7] D. Kleinman and P. Kisliuk, "Discrimination against unwanted orders in the Fabry-Perot resonator," *Bell. Sys. Tech. J.*, vol. 41, pp. 453–462; March, 1962. Comments on this paper have been made by A. G. Fox, T. Li, D. Kleinman, and P. Kisliuk, *Bell. Sys. Tech. J.*, vol. 41, pp. 1475–1476; July, 1962, and successful application of the suggestion is given by H. Kogelnik and C. Patel, "Mode suppression and single frequency operation in gaseous optical masers," Proc. IRE (*Correspondence*), vol. 50, pp. 2365–2366; November, 1962.

[8] G. Goubau and F. Schwering, "On the guided propagation of electromagnetic wave beams," IRE Trans. on Antennas and Propagation, vol. AP-9, pp. 248–256; May, 1961.

[9] D. Herriott, "Optical properties of a continuous He-He maser," *J. Opt. Soc. Am.*, vol. 52, pp. 31–37; January, 1962. See also, W. W. Rigrod, "Isolation of axi-symmetrical modes," *Appl. Phys. Letters*, vol. 2, p. 51; February, 1963.

[10] D. Marcuse, "HCN molecular beam-type maser," IRE Trans. on Instrumentation, vol. I 11, pp. 187–190; December, 1962.

[11] E. H. Scheibe, "Measurements on resonators formed from circular plane and confocal paraboloidal mirrors," Proc. IRE (*Correspondence*), vol. 49, p. 1079; June, 1961.

[12] J. Christian and G. Goubau, "Experimental studies on a beam waveguide for millimeter waves," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-9, pp. 256–263; May, 1961.
[13] J. Christian and G. Goubau, "Some measurements on an iris beam waveguide," PROC. IRE (*Correspondence*), vol. 49, pp. 1679–1680; November, 1961.
[14] J. Beyer, "Study of the Beam Waveguide," Ph.D. thesis, University of Wisconsin, Madison, 1961; University Microfilms, Inc., Ann Arbor, Mich. A portion of this work has been published: J. Beyer and E. Scheibe, "Loss measurements of the beam waveguide," IEEE TRANS. on MICROWAVE THEORY AND TECHNIQUES, vol. MTT-11, pp. 18–22; January, 1963.
[15] J. Beyer and E. Scheibe, "Higher modes in guided electromagnetic wave beams," IRE TRANS. ON ANTENNAS AND PROPAGATION (*Correspondence*), vol. AP-10, pp. 349–350; May, 1962.
[16] Y. Beers, R. Zimmerer, G. Strine, and M. V. Anderson, "Millimeter wavelength resonant structures," IEEE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MIT-11, pp. 142–149; March, 1963. See also, W. Culshaw, "Resonator for millimeter and submillimeter wavelengths," IEEE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-9, pp. 135–144; March, 1961.
[17] The use of ellipsoidal microwave cavities has not been extensive. This is in part due to the difficulty of theoretical anlaysis. Two such theoretical investigations are recorded in the literature and pertain to this discussion. The word confocal is used both in its geometrical sense as confocal ellipses and optical sense of lenses having a common focal point. J. C. Simons, "Electromagnetic Resonant Behavior of a Confocal Spheroidal Cavity System in the Microwave Region," Ph.D. thesis, M.I.T., Cambridge; 1950. H. Geppert and H. Schmid, "Confocal Ellipsoidal Shells as Electromagnetic Cavity Resonators," Wartime Rept. of the Mathematical Institute of the University of Berlin, Germany; available from ASTIA as Document ATI 32573, May, 1944.
[18] C. Flammer, "Spheroidal Wave Functions," Stanford University Press, Palo Alto, Calif., Ch. 8.2; 1957.
[19] See for example the discussion of the Scalar wave equation in elliptic cylinder coordinates for $b \gg \lambda$ in P. Morse and H. Feshbach, "Methods of Theoretical Physics," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 1416–1420; 1953.
[20] E. Pinney, "Electromagnetic fields in a paraboloidal reflector," *J. Math. and Phys.*, vol. 26, pp. 42–55; 1947.
[21] J. R. Pierce, "Theory and Design of Electron Beams," D. Van Nostrand and Co., New York, N. Y.; 1954.
[22] ——, "Modes in sequences of lenses," *Nat'l. Acad. of Sci.*, vol. 47, pp. 1808–1813; November, 1961.
[23] The conventional use of r, R and z, Z and x, X in the following discussion to represent impedance and reactance should not be confused with the earlier use of the same symbols for coordinates.
[24] This particular application is discussed in C. G. Montgomery, R. H. Dicke, and E. M. Purcell, "Principles of Microwave Circuits," McGraw-Hill Book Company, Inc., New York, N. Y. sec. 7.10; 1948.
[25] E. Rosenblum, "Atmospheric absorption of 10-400 Gcs radiation," *Microwave J.*, vol. 4, pp. 91–96; March, 1961.
[26] R. Zimmerer, "New wavemeter for millimeter wavelengths," *Rev. Sci. Instr.*, vol. 33, pp. 858–859; August, 1962.

Reprinted from IEEE TRANSACTIONS
ON MICROWAVE THEORY AND TECHNIQUES
Volume MTT-11, Number 5, September, 1963

385-582 O - 70 - 28

New Wavemeter for Millimeter Wavelengths

ROBERT W. ZIMMERER
National Bureau of Standards, Boulder, Colorado
(Received May 18, 1962)

A new wavemeter of simple design is described. The principle of operation makes use of a new development in physical optics. The actual performance of the device was measured and compared with the theory.

THE application of physical optics to microwave circuits has brought about some new and useful microwave devices. One specific device is the confocal resonator described by Boyd and Gordon.[1] This was applied by Marcuse[2] in the design of his HCN maser to give a high Q resonant cavity at 88 Gc. An obvious application which comes to mind in view of the simplicity of the structure is its use as a wavemeter. The commercially available cavity wavemeters above 40 Gc leave much to be desired. They must be used with calibration curves and yield rather broad absorption dips in waveguide circuits.

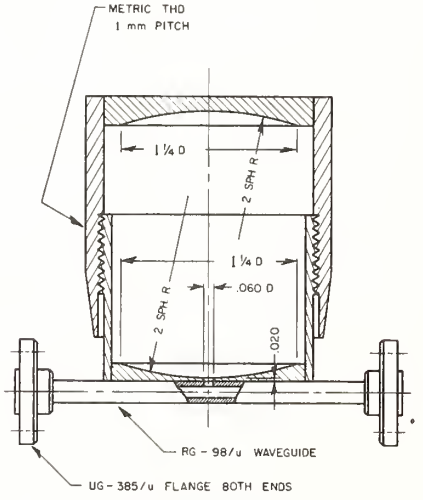


FIG. 1. Construction of microwave wavemeter using confocal resonator.

A small confocal resonator has been designed at our laboratory to provide a simple easily used wavemeter covering the 50- to 75-Gc band in RG-98/u waveguide. In operation one spherical mirror is moved back and forth with respect to the other by screwing the outside cylinder as shown in Fig. 1. The cavity resonances occur at intervals of $\lambda/2$.

By noting the distance the cylinder moves between successive absorption dips, as observed in the external microwave circuit, the wavelength can be directly measured. If the screw thread is metric the wavelength is given in millimeters without any conversion of units.

As with all cavity devices the wavelength within the cavity exceeds the free space wavelength by a small amount. From the theory as developed by Boyd and Gordon[1] or by Goubau and Schwering[3] one is led to expect a relation $\lambda_c=\lambda[1+\lambda/(4\pi b)]$ to hold, where λ_c is the wavelength in the cavity, λ is the free space wavelength, and b is the radius of curvature of the spherical mirrors. For the dimensions of the cavity this would indicate an increase of about 1% in the wavelength as measured over the free space wavelength. We have recorded an actual increase of 1.8% at 49.15 Gc, 1.6% at 59.10 Gc, and 1.4% at 68.95 Gc. This correction can be easily applied yielding at least three figure accuracy to wavelength measurements made with this instrument. The measured loaded Q value of this wavemeter is 15 000 at 59.1 Gc.

[1] G. D. Boyd and J. P. Gordon, Bell System Tech. J. **60**, 489 (1961).
[2] D. Marcuse, Proc. I.R.E. **49**, 1706 (1961).
[3] G. Goubau and F. Schwering, IRE Trans. on Antennas Propagation **AP-9**, 248 (1961).

The construction of the wavemeter shown in Fig. 1 is straightforward. The dimensions of the cavity were chosen to satisfy the relation $a^2=b\lambda$ where a is the mirror radius. The wavelength used for design purposes was 5 mm. The spherical surfaces of 2-in. radius were machined in brass and polished to a high finish. The inside cylinder wall between the spherical ends must be made quite lossy to suppress unwanted resonances. Aqua-Dag proved unsatisfactory but an insert made from microwave absorbant cloth served quite well. Other conducting paints may be suitable for this purpose. One broad wall on the RG-98/u waveguide was machined to a thickness of 0.010 in. and soldered into a slot in the brass blank before the spherical surface was machined. The spherical surface comes within 0.010 in. of the waveguide giving a total thickness of 0.020 in. which appears to be necessary for mechanical strength. The wavemeter was marked in the same manner as a micrometer head so that the indicated distance between successive absorption dips was twice the real distance giving the wavelength directly in millimeters. In operation one records the position of several successive dips and takes the average of their differences.

No doubt further refinements such as increasing the dimensions and using a precision screw thread would lead to significant improvement in accuracy of the wavelength measurements. Until this new type of wavemeter is manufactured commercially, this instrument in the rather simple design described here can be readily made in any instrument shop at small cost and provide the user of millimeter wave power with an accurate home made wavemeter.

Reprinted from The Review of Scientific Instruments, Vol. 33, No. 8, 858-859, August, 1962
Printed in U. S. A.

9.a. **Reflectors for a microwave Fabry-Perot interferometer,** W. Culshaw, IRE Trans. Microwave Theory & Tech. MTT-7, No. 2 (April 1959).

The advantages of microwave interferometers for wavelength and other measurements at millimeter wavelengths are indicated, and a microwave Fabry-Perot interferometer discussed in detail. Analogous to the cavity resonator, this requires reflectors of high reflectivity, small absorption, and adequate size. Stacked dielectric plates, and stacked planar or rod gratings are shown to be suitable forms of reflectors, and equations for the reflectivity, optimum spacing, and bandwidth of such structures are derived. A series of stacked metal plates with regularly spaced holes represents a good design of reflector for very small wavelengths. Fringes and wavelength measurements at 8-mm wavelength are given for one design of interferometer, these being accurate to 1 in 10^4 without any diffraction correction. For larger arpertures and reflectors in terms of the wavelength, errors due to diffraction will decrease.

9.b. **High resolution millimeter wave Fabry-Perot interferometer,** William Culshaw, IRE Trans. Microwave Theory & Tech. MTT-8, No. 2, 182-189 (March 1960).

The design and operation of a microwave Fabry-Perot interferometer at wavelengths around 6 mm is described. This uses reflectors which are simple, easy to make, and which are capable of scaling for operation at short wavelengths in the ultramicrowave region. With power reflection coefficients around 0.999, very sharp fringes and Q values around 100,000 were obtained on the interferometer. Effects of diffraction in the interferometer are considered, and wavelength measurements with this particular interferometer indicate that accuracies of 0.04 percent are obtained without any diffraction correction. Advantages of such an interferometer for ultramicrowaves are that the component parts are large compared with the wavelength, the effects of diffraction decrease with the wavelength, and the problem of maintaining a high Q with a single mode of propagation and a structure of adequate size is made much easier. Such an interferometer forms the cavity resonator for ultramicrowaves. It can thus be used for such conventional purposes as wavelength measurements, wavelength spectral analysis, dielectric constant, and loss measurements, or as the cavity resonator for frequency stabilization, or as the cavity resonator for a millimeter- or submillimeter-wavelength maser.

9.c. **Resonators for millimeter and submillimeter wavelengths,** William Culshaw, IRE Trans. Microwave Theory & Tech. MTT-9, No. 2, 135-144 (March 1961).

Further considerations on the mm-wave Fabry-Perot interferometer are presented. Computed Q values for parallel metal plate resonators indicate that at spacings around 2.5 cm, values ranging from 60,000 at 3 mm, to 300,000 at 0.1 mm wavelengths are possible. The plates must, however, be quite flat. These results are important for many investigations, and in particular for mm and sub-mm wave maser research. For the aperture per wavelength ratios possible here, diffraction effects should be small. Consideration is given to using curved reflectors or focused radiation in applications where the fields must be concentrated. For this purpose, re-entrant conical spherical resonators are treated in detail, as regards operation in the TEM mode at high orders of interference. Expressions for the Q and shunt impedance are given, and high values are possible at mm and sub-mm wavelengths. Quasi-optical methods of coupling into and out of such a resonator are proposed, and the higher modes possible in such a resonator are considered. Results indicate that it could have application to the mm-wave generation problem, and that it represents a good resonant cavity for solid-state research at mm and sub-mm wavelengths, and for maser applications in particular.

*9.d. **Experimental investigation of Fabry-Perot interferometers,** R. W. Zimmerer, Proc. IEEE 51, No. 3 (March 1963).

See also 3.b, 8.5.

* Private communication.

10. Components and Subsystems

* Private communication.

Precision Detector for Complex Insertion Ratio Measuring Systems

C. M. ALLRED, MEMBER, IEEE, AND R. A. LAWTON

Summary—**A precision, low-noise, highly sensitive, synchronous RF detector is described that permits simultaneous phase and level nulling in complex insertion ratio measuring systems. This insertion ratio is a measure of both the phase shift and magnitude change produced by a network when inserted in a transmission path.**

I. Introduction

SPECIAL DETECTORS are often needed in the attentuation and phase standards programs of the National Bureau of Standards. This paper describes a detector that was developed for this purpose.

The basic aspect of a 30-Mc system under development is portrayed in block-diagram form in Fig. 1. The two standard attenuators are waveguide-below-cutoff or piston attenuators operating in the TE_{11} mode and arranged so that their outputs are out of time phase by 90°. The vector sum of these two attenuators is a vector lying within a 90° sector and can null any voltage in an opposite quadrant. By rotating the pickup coils of the two standard attenuators, their outputs may be shifted by 180°. Hence a voltage lying in any quadrant may be nulled. The desired insertion ratio of the unknown networks is obtained from the two sets of readings from the two standard attenuators necessary to produce a null at the detector both before and after inserting the network. The insertion ratio obtained is a complex numeric giving both the magnitude and phase change produced by inserting the network. The junction is a special unit[1] that isolates the signals from the three channels and also presents matched impedances to the three channels.

The detector needed for this system should be a stable low-noise unit that provides the necessary information to enable the operator or servosystem to move each standard attenuator independently toward the null condition. Thus, the detector should indicate for each standard independently whether the attenuation should be increased or decreased and when zero output is obtained.

The other attenuation system[2] which has been in use in calibration work is shown in block-diagram form in

Manuscript received January 27, 1964.
The authors are with the Boulder Laboratories, National Bureau of Standards, Boulder, Colo.

[1] C. M. Allred and C. C. Cook, "A multiple isolated-input network with common output," *J. Res. Nat. Bur. Stand.* (USA), vol. 64C, pp. 225–228; July–September, 1960.
[2] C. M. Allred and C. C. Cook, "A precision RF attenuation calibration system," IRE TRANS. ON INSTRUMENTATION, vol. I-9, pp. 268–274; September. 1960.

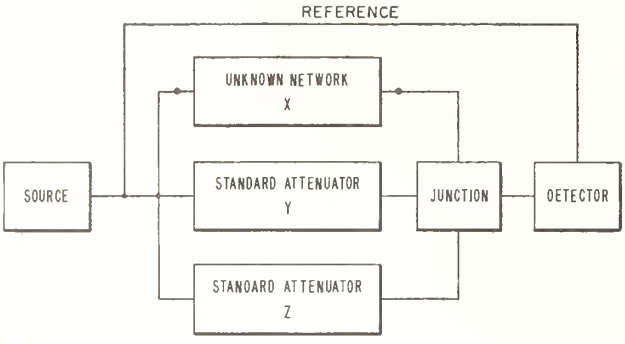


Fig. 1—Three-channel attenuation system.

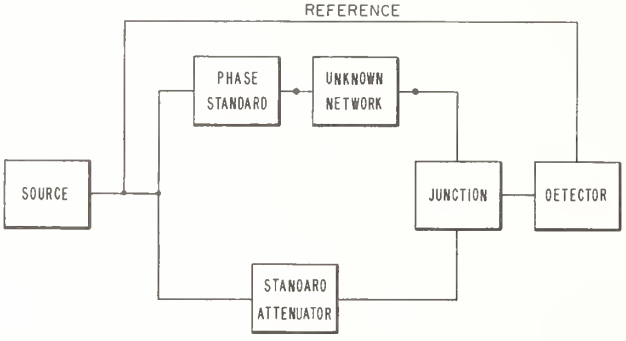


Fig. 2—Two-channel attenuation system.

Fig. 2. The two standards are adjusted for null condition, as indicated by the detector, before and after insertion of the network. The change in the phase standard gives the insertion phase angle of the network and the change in the attenuator provides the insertion loss of the unknown. The unknown network may also be inserted in the attenuator channel instead of the phase standard channel. The junction plays the same role as before. As in the previous system, it would be preferable to achieve a null condition by an essentially independent adjustment of the standards.

These systems are state-of-the-art systems where the research program is seeking accuracies in the order of a few ten-thousandths of a db per ten db, and the detectors must not be a deleterious factor.

II. Theory

In the phase and attenuation standards used in the above systems, except for the secondary effects, the phase standard is free of level changes and the attenuators are free of phase shifts.

For the three-channel system, Thevenin's theorem and the principle of superposition allows the output of the junction under a null condition to be written

$$V_y + jV_z + V_x e^{j\theta_x} = 0. \tag{1}$$

Thus,

$$V_y = -V_x \cos\theta_x \tag{2}$$

$$V_z = -V_x \sin\theta_x, \tag{3}$$

where V_y, V_z, and $V_x e^{j\theta_x}$ represent the respective outputs of the Y, Z, and X channels. Thus, the detector must determine two separate conditions independently, that is, when each attenuator nulls with its respective component of $V_x e^{j\theta_x}$. Phase or synchronous detectors, or perhaps more appropriately called here correlation detectors, have this property and have low noise producing characterisics. For this reason phase detectors were chosen for this application.

The basic principles of such detectors are represented by a simple multiplier which gives the arithmetic product of two signals followed by a filter or integrator which effectively takes the time average of the product. Thus, if $A\cos(\omega t+\psi)+n(t)$ represents an input signal plus noise, the output of the detector is

$$\overline{[A\cos(\omega t+\psi)+n(t)]B\cos(\omega t+\eta)} = \frac{AB}{2}\cos(\eta-\psi), \tag{4}$$

where the bar indicates a time average and the angle $(\eta-\psi)$ is the angle between the reference signal and the correlated input signal. The mean of $n(t)$, $\overline{n(t)}$ is assumed equal to zero. The output is zero for $\eta-\psi$ equal to $\pm\pi/2$ and is of maximum magnitude when $\eta-\psi$ is equal to zero or π. If two such units were used with one reference signal $B_1\cos(\omega t+\eta_1)$ in phase with the output of one attenuator and the other reference signal $B_2\cos(\omega t+\eta_2)$ in phase with the other attenuator and 90° out of phase with each other, they would be responsive only to adjustments of their corresponding attenuators.

If in the second system which uses the phase standard, two phase or correlation detectors as described above were used, one aligned with the attenuator and the other orthogonal to it, the output of the attenuator aligned detector would be proportional to

$$V_{ai} + V_\phi \cos\phi_i,$$

where V_{ai} and $V_\phi e^{j\phi_i}$, respectively, refer to the initial outputs of the attenuator and phase standard. The other detector output would be proportional to

$$V_\phi \sin\phi_i.$$

When a null is achieved

$$\phi_i = \pi \tag{5}$$

$$V_{ai} = +V_\phi. \tag{6}$$

Note that prior to null the attenuator setting is a function of ϕ_i. This means that the attenuator aligned detector would not be independent of the phase adjustment. However, the other detector would allow an independent phase adjustment, and if this adjustment were made first the attenuator adjustment would then be independent. If servosystems were used no problems would exist other than that the attenuator adjustment would not be final until the phase adjustment had been

completed. There is no essential difference in the nulling process after the unknown has been inserted in the phase standard channel.

However, if the unknown were inserted in the attenuator channel, the reference signals in the detectors would not, in general, be aligned properly, with the result that each adjustment would be dependent on the other. This is shown by the null condition equation.

$$V_{af}Ke^{j\theta} + V_{\phi}e^{j\phi_f} = 0 \tag{7}$$

or

$$V_{af}K\cos\theta + V_{\phi}\cos\phi_f = 0 \tag{8}$$

$$V_{af}K\sin\theta + V_{\phi}\sin\phi_f = 0, \tag{9}$$

where $Ke^{j\theta}$ is the insertion ratio of the network and the subscript f denotes the final measurement.

If, when the unknown network was inserted in the attenuator channel, the reference signals were arranged so that one was always aligned with the output signal of the phase standard and the other orthogonal to it (this is not too difficult to do), the first detector response would be proportional to

$$V_{af}K\cos(\theta_f - \phi) + V_{\phi}.$$

The other response would be proportional to

$$V_{af}K\sin(\theta_f - \phi),$$

and the same type of independence would exist as when the unknown was inserted in the phase standard channel.

The above discussion assumed perfection in the detectors and other components. The type of imperfections occurring in quality units merely degrades the degree of independence and makes it necessary to adjust the standards more than once. This is a rather mild inconvenience and with properly designed servosystems it should be insignificant.

Even with an ideal detector, one that provides the exact product between input and reference, noise is still present in the output because of the finite averaging time inherent in practical systems. For this discussion, the sensitivity is defined as the value of the input-signal amplitude that produces a mean deflection equal to the rms value of the random fluctuations in the output. If M is the meter reading, then

$$\overline{M} = [\overline{(M - \overline{M})^2}]^{1/2} \tag{10}$$

is the sensitivity criteria.

To evaluate A_m, the signal amplitude that produces such an $\overline{M}$, assume that an integration of time τ performs the filtering after detection. Thus,

$$M = \frac{1}{\tau}\int_0^{\tau}(A\cos\omega_0 t + n(t))B\cos\omega_0 t\,dt, \tag{11}$$

where τ is large compared to $2\pi/\omega_0$.

Thus,

$$\overline{M} = \frac{AB}{2}. \tag{12}$$

Now $n(t)$ (a narrow-band stationary Gaussian process) can be written as[3]

$$n(t) = X_c(t)\cos\omega_0 t - X_s(t)\sin\omega_0 t, \tag{13}$$

where $X_c(t)$ and $X_s(t)$ are random functions with zero means and

$$\overline{X_c^2} = \overline{X_s^2} = \overline{[n(t)]^2} = 2\int_0^{\infty} S_n(f)df \tag{14}$$

$$\overline{X_c(t)X_c(t+\sigma)} = \overline{X_s(t)X_s(t+\sigma)} = \int_{-\infty}^{\infty} S_n(f)e^{j2\pi(f-f_0)\sigma}df, \tag{15}$$

$S_n(f)$ = power spectral density, f = frequency.

Then, calculating $\overline{(M-\overline{M})^2}$,

$$\overline{(M-\overline{M})^2} = \frac{B^2}{4\tau^2}\int_0^{\tau}\int_0^{\tau}\overline{X_c(t_1)X_c(t_2)}\,dt_1dt_2 = \frac{B^2}{4\tau^2}\int_0^{\tau}dt_1\int_{-t_1}^{\tau-t_1}\overline{X_c(t_1)X_c(t_1+\sigma)}\,d\sigma, \tag{16}$$

where the change of variable, $\sigma = t_2 - t_1$ has been made. The limits of integration of the second integral may be extended to infinity since τ is much larger than σ_0. Here σ_0 is the value of σ for which $\overline{X_c(t)X_c(t+\sigma)}$ becomes essentially zero.[4] Thus,

$$\overline{(M+\overline{M})^2} = \frac{B^2}{4\tau}S_n(f_0) \tag{17}$$

and, therefore,

$$A_m = \sqrt{\frac{\overline{S_n(f_0)}}{\tau}}. \tag{18}$$

In this presentation $S_n(f)$ is spread over both positive and negative values of f and, hence, is half the value of some representations of noise spectral density.

When a simple RC filter of time constant τ is used, theory shows that

$$A_m = \sqrt{\frac{\overline{S_n(f_0)}}{2\tau}}. \tag{19}$$

If the multiplier or correlator were followed by a critically damped galvanometer of time constant τ, then

[3] W. B. Davenport and W. L. Root, "Random Signals and Noise," McGraw-Hill Book Company, Inc., New York, N. Y., pp. 158–162; 1958.

[4] A. van der Ziel, "Noise," Prentice-Hall, Inc., New York, N. Y., 1st ed., ch. 12 p. 313; 1954.

$$A_m = \sqrt{\frac{S_n(f_0)}{4\tau}}. \tag{20}$$

A more general representation of the above detection system might be made by a "black box" without memory consisting of two input ports and one output port. The "black box" is followed by a linear filter or integrator which provides the memory aspects.

Let the signal at one input be

$$x = A \cos \omega_0 t + n(t) \tag{21}$$

and the other input be the reference signal

$$y = B \cos \omega_0 t. \tag{22}$$

The output z of the black box can be written as a Maclaurin expansion for a function of two independent variables. Thus,

$$\begin{aligned} z = {} & c_{00} + c_{10}x + c_{01}y + c_{11}xy \\ & + c_{20}x^2 + c_{02}y^2 + c_{21}x^2y + c_{12}xy^2 \\ & + c_{30}x^3 + c_{03}y^3 + c_{22}x^2y^2 \\ & + c_{31}x^3y + c_{32}x^3y^2 + c_{33}x^3y^3 \\ & + c_{23}x^2y^3 + c_{13}xy^3 + \cdots . \end{aligned} \tag{23}$$

The predominant term should be $c_{11}xy$, the desired result, with the other terms small and decreasing in magnitude as the exponents increase. A brief discussion of the effects of some of the terms follows. The term c_{00} represents a dc output that is independent of the levels of the two inputs. If it is stable it can be balanced out. The terms $c_{10}x$ and $c_{01}y$ are directly proportional to the two inputs and are removed by the filter system following the correlator. The two undesirable terms $c_{20}x^2$ and $c_{02}y^2$ produce dc outputs that depend, respectively, on the signal plus noise and reference signal strength. This makes the dc output gain sensitive. In addition, the first term increases the noise level at the output. The effect of the second term may be balanced out if c_{02} is stable and y is of fixed level. The effects of the following four terms, $c_{21}x^2y$, $c_{12}xy^2$, $c_{30}x^3$, and $c_{03}y^3$ are removed by the filter. Succeeding terms have similar effects but normally decrease in magnitude.

III. The Detector

The detector consisted of a 30-Mc RF amplifier and two correlators or multipliers as shown in Fig. 3.

The RF amplifier was made up of a preamplifier with an input impedance of 17 Ω and a noise figure of 2 db. The preamplifier was followed by a band-pass filter with a center frequency of 30.0 Mc and a 3-db bandwidth of 210 kc which determined the bandwidth of the entire RF amplifier. After the filter came a postamplifier with a nominal output impedance of 50 Ω. The over-all gain of the RF amplifier could be varied from 128 db to 27 db.

A 6AR8 sheet beam vacuum tube was used as a correlator or multiplier with the circuit diagram shown in Fig. 4. The reference signal was fed in push pull to the deflector plates of the 6AR8 with a peak to peak voltage of 40 volts. The signal from the RF amplifier was fed through an emitter-follower circuit, employing a 2N2218 transistor, to the grid of the 6AR8. The emitter follower provided a low impedance for the grid of the 6AR8 which reduced the coupling between the deflector plates and grid. This coupling was further reduced by interpin shielding of the 6AR8 tube. Biasing of the deflectors for optimum sensitivity was provided by the 620-Ω resistor in the cathode circuit.

The arithmetic product of input signal and reference

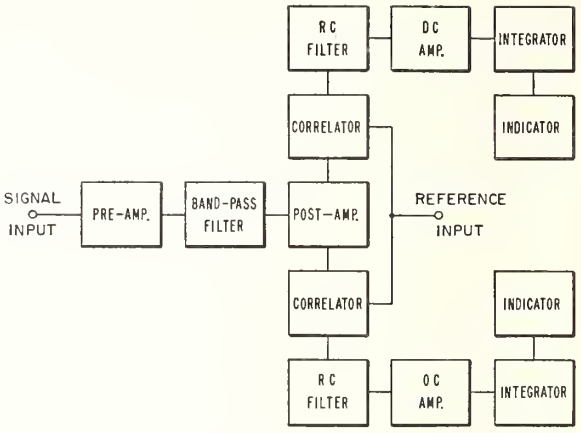


Fig. 3—The detector.

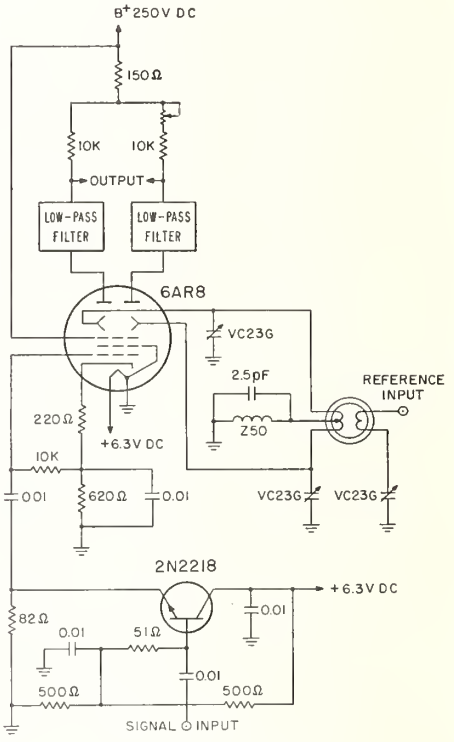


Fig. 4—The correlator.

signal then appeared as the voltage difference between the voltages across the two load resistors. This difference voltage was fed through an RC filter with a time constant of 1 sec to an available strip chart recorder which had as accessory equipment a ball and disc integrator. The RC filter was used to help prevent the noise peaks from overloading the recorder amplifier.

Signal and reference inputs to the correlator were matched to 50 Ω. Each of the correlators as well as the RF amplifier were well shielded electrostatically and, in addition, the correlators were magnetically shielded. The dc output and all supply lines were carefully filtered.

IV. Performance

The performance tests were made on the first insertion ratio measuring system described. This system and the detector are shown in Fig. 5.

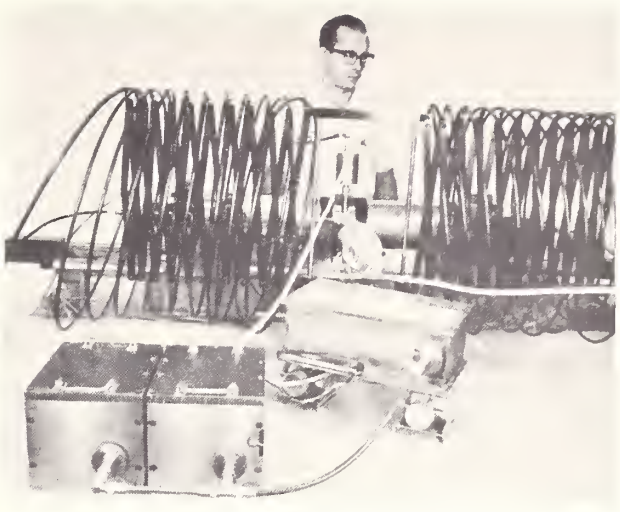


Fig. 5—Attenuation system and detector.

Cosine Dependence

A test was made of the response of one of the correlators to an input signal of constant amplitude but varying phase. The phase was varied by means of a line stretcher. The result is plotted in Fig. 6 and follows closely a cosine curve indicating the predominance of the xy term over other terms in the general theory.

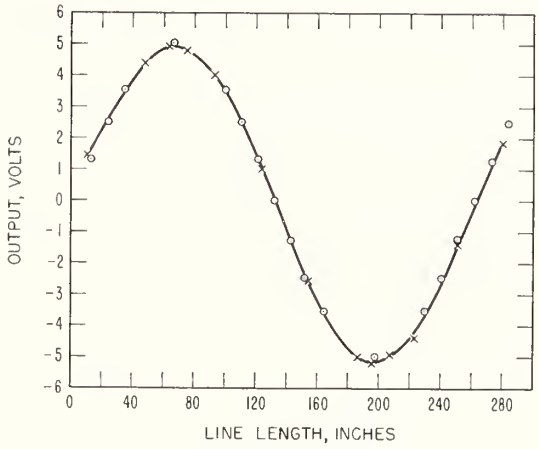


Fig. 6—Phase response of correlators for input amplitude of 600 mv peak.

Independence of Correlators

As a measure of the independence of the correlators to orthogonal or quadrature signals, it was found possible to have a ratio of 200 between the quadrature signals and the in-phase signals before a significant response was noted due to the quadrature signal.

Sensitivity

As stated in Section II, the theoretical value of the minimum detectable signal is

$$A_m = \sqrt{\frac{\overline{S_n(f_0)}}{\tau}}. \tag{18}$$

Here, $S_n(f_0)$, the open-circuited mean-square voltage per unit cycle per second bandwidth (spread over positive and negative frequencies), is given by

$$S_n(f_0) = 2FKTR, \tag{24}$$

where

$F = 2$-db noise figure
$K = 1.38 \times 10^{-23}$ joules per degree, is Boltzmann's constant
$T = 297$, is the room temperature in degrees Kelvin
$R = 16.5\ \Omega$, is the source impedance presented to the RF amplifier.

With an integration time τ of 30 sec, the value of A_m is

$A_m = 42.3 \times 10^{-12}$ volts peak
$= 42.3$ picovolts peak
$= 29.9$ picovolts rms.

TABLE I

Attenuation Reading (db-relative)	$\sqrt{\overline{(M-\overline{M})^2}}$	$\overline{M}$	$\sqrt{\dfrac{\overline{(M-\overline{M})^2}}{\overline{M}}}$	V_{min} (pico-volts)
47.6	1.38	1.00	1.38	42
50.0	1.73	1.89	.915	36
55.0	1.21	2.44	.495	37
60.0	1.25	5.13	.244	31

Experimentally the minimum detectable signal was found from various settings of one of the standard attenuators allowing no signal to pass through either of the other two channels. In Table I the rms fluctuations $\sqrt{\overline{(M-\overline{M})^2}}$ and mean signal deflections $\overline{M}$ of the integrator are tabulated for various settings of the standard piston attenuator, and a minimum detectable signal is extrapolated for each setting. The average minimum detectable signal for all the attenuator settings is 36 picovolts rms. The theoretical value is 17 per cent lower than this value. The agreement is considered

quite satisfactory, however, in view of the nature of the measurements, the uncertainties in the values of various quantities such as noise figure, absolute signal level, and imperfections in the integrator, together with the instabilities or drift in the dc ouput of the correlators with no noise present. This drift was 600 μv per hour and, of course, helps to increase the effective noise level. This drift amounts to about 1/3 of the minimum detectable signal for a 30 sec integration time. Plans have been formulated to further reduce this drift.

Leakage

The insertion ratio measuring device that was used is still under development, and shielding, though quite good, is inadequate as there were indications that leakage signals were present that produced a net indicator deflection of the order of the magnitude of the minimum detectable signal. Time did not allow for redesign of parts of the experimental system (nor is it always desirable to add elaborate shielding precautions to an experimental unit in the state of flux). An expedient was used of balancing out the mean deflection due to leakage. Various tests were then conducted to ascertain if the leakage varied during the operations necessary to measure the minimum detectable signal. The indications were that the "balanced mean" remained essentially constant and equal to zero.

V. Conclusions

A detecting system operating near theoretical limits has been constructed and performance tests indicate that it should significantly aid in the development of certain insertion ratio measuring systems with greatly increased range, convenience, and accuracy.

Reprinted from IEEE TRANSACTIONS
ON *INSTRUMENTATION AND MEASUREMENT*
Volume IM-13, Numbers 2 & 3, June-September, 1964
Pp. 76-81

*10.a. **A low input VSWR coaxial diode switch for the UHF band,** W. L. Ecklund, IEEE Trans. Microwave Theory & Tech. MTT-12, No. 3 (May 1964).

*10.b. **A method of improving isolation in multichannel waveguide systems,** G. F. Engen, IRE Trans. Microwave Theory & Tech. MTT-8, No. 4 (July 1960).

10.c. **Errors in dielectric measurements due to a sample insertion hole in a cavity,** A. J. Estin and H. E. Bussey, IRE Transactions Microwave Theory and Tech. MTT-8, No. 6, 650-653 (November 1960).

The measurement of complex permittivities of isotropic media at microwave frequencies is performed with high precision by means of cylindrical cavity resonators. However, a hole in the cavity wall for inserting the sample causes a frequency pulling of the resonator which in turn introduces an error in the measured dielectric constant. These effects are measured, and with perturbation theory as a guide, correction factors are developed.

10.d. **A versatile ratio instrument for the high ratio comparison of voltage or resistance,** Alfred E. Hess J. Res. NBS 70C (Eng. & Instr.), No. 3, 169-172 (September 1966).

A 9-dial resistance ratio instrument capable of precise high ratio comparison of voltage or resistance (from 1/1 to $10^7/1$) is described. Also described is the modification of a 6-dial universal ratio set to permit its additional use as a versatile ratio instrument. Paramount to the accuracy of these high ratio instruments is the carefully adjusted "common point" junction which is briefly discussed.

*10.e. **A high directivity, broadband coaxial coupler,** P. A. Hudson, IEEE Trans. Microwave Theory & Tech. MTT-14, No. 6, 293-294 (June 1966).

10.f. **Low-level low-frequency detection system,** Neil T. Larsen, Rev. Sci. Instr. 33, No. 11, 1200-1208 (November 1962).

An amplifier and detector having application in a null detection system or as a servo amplifier in low-frequency ac instruments is described. The design may be adapted to operate with a carrier frequency in the range from a few cps to a few kc. The unique and apparently unrecognized advantages of operating with a carrier frequency which is an even subharmonic of the principal coherent disturbing signal, such as the power line, are discussed in detail. Design considerations and performance data are presented, permitting an evaluation of the suitability of the device for a given application. For a particular input circuit configuration, it is conservatively estimated that a signal of less than 1.6×10^{-22} W from 100—Ω source may be detected with an observation time of less than 1 min. Experiments indicate that, with this device, the sensitivity is primarily limited by thermal noise in the input and by induced signals in the input leads.

The instrument takes the form of a sealed-off vacuum tube having a short section of waveguide as part of the tube. The ends of the waveguide may be sealed using known window techniques. The theory for the ideal case is presented and then means of correcting for the various perturbations present in an actual tube are discussed. The theory and the corrections are verified by experiment. The technique appears to have definite value for monitoring or measuring high- and medium-level cw or pulse-power flow. Theoretically the device is self-calibrating and therefore might make a good primary standard. Its suitability as a primary standard is under further investigation.

* Private communication.

Author Index

* Correspondence.

Subject Index

NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards [1] was established by an act of Congress March 3, 1901. Today, in addition to serving as the Nation's central measurement laboratory, the Bureau is a principal focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. To this end the Bureau conducts research and provides central national services in four broad program areas. These are: (1) basic measurements and standards, (2) materials measurements and standards, (3) technological measurements and standards, and (4) transfer of technology.

The Bureau comprises the Institute for Basic Standards, the Institute for Materials Research, the Institute for Applied Technology, the Center for Radiation Research, the Center for Computer Sciences and Technology, and the Office for Information Programs.

THE INSTITUTE FOR BASIC STANDARDS provides the central basis within the United States of a complete and consistent system of physical measurement; coordinates that system with measurement systems of other nations; and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. The Institute consists of an Office of Measurement Services and the following technical divisions:

Applied Mathematics—Electricity—Metrology—Mechanics—Heat—Atomic and Molecular Physics—Radio Physics [2]—Radio Engineering [2]—Time and Frequency [2]—Astrophysics [2]—Cryogenics.[2]

THE INSTITUTE FOR MATERIALS RESEARCH conducts materials research leading to improved methods of measurement standards, and data on the properties of well-characterized materials needed by industry, commerce, educational institutions, and Government; develops, produces, and distributes standard reference materials; relates the physical and chemical properties of materials to their behavior and their interaction with their environments; and provides advisory and research services to other Government agencies. The Institute consists of an Office of Standard Reference Materials and the following divisions:

Analytical Chemistry—Polymers—Metallurgy—Inorganic Materials—Physical Chemistry.

THE INSTITUTE FOR APPLIED TECHNOLOGY provides technical services to promote the use of available technology and to facilitate technological innovation in industry and Government; cooperates with public and private organizations in the development of technological standards, and test methodologies; and provides advisory and research services for Federal, state, and local government agencies. The Institute consists of the following technical divisions and offices:

Engineering Standards—Weights and Measures — Invention and Innovation — Vehicle Systems Research—Product Evaluation—Building Research—Instrument Shops—Measurement Engineering—Electronic Technology—Technical Analysis.

THE CENTER FOR RADIATION RESEARCH engages in research, measurement, and application of radiation to the solution of Bureau mission problems and the problems of other agencies and institutions. The Center consists of the following divisions:

Reactor Radiation—Linac Radiation—Nuclear Radiation—Applied Radiation.

THE CENTER FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides technical services designed to aid Government agencies in the selection, acquisition, and effective use of automatic data processing equipment; and serves as the principal focus for the development of Federal standards for automatic data processing equipment, techniques, and computer languages. The Center consists of the following offices and divisions:

Information Processing Standards—Computer Information — Computer Services — Systems Development—Information Processing Technology.

THE OFFICE FOR INFORMATION PROGRAMS promotes optimum dissemination and accessibility of scientific information generated within NBS and other agencies of the Federal Government; promotes the development of the National Standard Reference Data System and a system of information analysis centers dealing with the broader aspects of the National Measurement System, and provides appropriate services to ensure that the NBS staff has optimum accessibility to the scientific information of the world. The Office consists of the following organizational units:

Office of Standard Reference Data—Clearinghouse for Federal Scientific and Technical Information [3]—Office of Technical Information and Publications—Library—Office of Public Information—Office of International Relations.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D.C. 20234.
[2] Located at Boulder, Colorado 80302.
[3] Located at 5285 Port Royal Road, Springfield, Virginia 22151.

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH reports National Bureau of Standards research and development in physics, mathematics, chemistry, and engineering. Comprehensive scientific papers give complete details of the work, including laboratory data, experimental procedures, and theoretical and mathematical analyses. Illustrated with photographs, drawings, and charts.

Published in three sections, available separately:

• Physics and Chemistry

Papers of interest primarily to scientists working in these fields. This section covers a broad range of physical and chemical research, with major emphasis on standards of physical measurement, fundamental constants, and properties of matter. Issued six times a year. Annual subscription: Domestic, $9.50; foreign, $11.75*.

• Mathematical Sciences

Studies and compilations designed mainly for the mathematician and theoretical physicist. Topics in mathematical statistics, theory of experiment design, numerical analysis, theoretical physics and chemistry, logical design and programming of computers and computer systems. Short numerical tables. Issued quarterly. Annual subscription: Domestic, $5.00; foreign, $6.25*.

• Engineering and Instrumentation

Reporting results of interest chiefly to the engineer and the applied scientist. This section includes many of the new developments in instrumentation resulting from the Bureau's work in physical measurement, data processing, and development of test methods. It will also cover some of the work in acoustics, applied mechanics, building research, and cryogenic engineering. Issued quarterly. Annual subscription: Domestic, $5.00; foreign, $6.25*.

TECHNICAL NEWS BULLETIN

The best single source of information concerning the Bureau's research, developmental, cooperative and publication activities, this monthly publication is designed for the industry-oriented individual whose daily work involves intimate contact with science and technology—*for engineers, chemists, physicists, research managers, product-development managers, and company executives.* Annual subscription: Domestic, $3.00; foreign, $4.00*.

* Difference in price is due to extra cost of foreign mailing.

NONPERIODICALS

Applied Mathematics Series. Mathematical tables, manuals, and studies.

Building Science Series. Research results, test methods, and performance criteria of building materials, components, systems, and structures.

Handbooks. Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications. Proceedings of NBS conferences, bibliographies, annual reports, wall charts, pamphlets, etc.

Monographs. Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

National Standard Reference Data Series. NSRDS provides quantitive data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated.

Product Standards. Provide requirements for sizes, types, quality and methods for testing various industrial products. These standards are developed cooperatively with interested Government and industry groups and provide the basis for common understanding of product characteristics for both buyers and sellers. Their use is voluntary.

Technical Notes. This series consists of communications and reports (covering both other agency and NBS-sponsored work) of limited or transitory interest.

Federal Information Processing Standards Publications. This series is the official publication within the Federal Government for information on standards adopted and promulgated under the Public Law 89–306, and Bureau of the Budget Circular A–86 entitled, Standardization of Data Elements and Codes in Data Systems.

CLEARINGHOUSE

The Clearinghouse for Federal Scientific and Technical Information, operated by NBS, supplies unclassified information related to Government-generated science and technology in defense, space, atomic energy, and other national programs. For further information on Clearinghouse services, write:

Clearinghouse
U.S. Department of Commerce
Springfield, Virginia 22151

Order NBS publications from:

Superintendent of Documents
Government Printing Office
Washington, D.C. 20402

U. S. GOVERNMENT PRINTING OFFICE : 1970 O - 385-582

Announcement of New Volumes in the NBS Special Publication 300 Series Precision Measurement and Calibration

Superintendent of Documents,
Government Printing Office,
Washington, D. C. 20402

Dear Sir:

Please add my name to the announcement list of new volumes to be issued in the series: National Bureau of Standards Special Publication 300, Precision Measurement and Calibration.

Name .....

Company .....

Address .....

City ..... State ..... Zip Code .....

(Notification key N-353)

(cut here)

Official SI Unit Names and Symbols

For a complete statement of NBS practice, see NBS Tech. News Bull. Vol. 52, No. 6, June 1968.

Name	Symbol	Name	Symbol
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	mho	mho
degree Celsius [3]	°C	mole	mol
gram	g	siemens [4]	S

[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 Practice Guide", NBS Handbook 102. 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 O °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 $T_{68}=(t_F-32)/1.8$.
(20) curies to disintegrations per second, multipy by 3.7×10^{10} exactly.
(21) roentgens to coulombs per kilogram, multiply by $2.579\ 760 \times 10^{-4}$ exactly.

[5] Systeme International d'Unites (designated SI in all languages).