KEITHLEY A GREATER MEASURE OF CONFIDENCE

Parallel Parametric Test Methodologies

Randall Lee
Keithley Instruments, Inc.

Background

The production of many electronic devices begins with wafer processing. In addition to CMOS ICs, this can include such diverse devices as RF components based on III-V compounds and chemical detectors based on carbon nanotube (CNT) FETs. In both R&D and production applications, there is a great deal of effort devoted to increasing device test throughput in order to shorten the time to market and reduce costs.

One way of doing this is to run tests in parallel on wafer test elements (as opposed to testing devices sequentially) using automated or semi-automated wafer probers connected to parametric test systems. This reduces overhead time and increases throughput by using instruments that might otherwise sit idle, waiting for a test routine to call them into action. Two basic strategies exist for parallel parametric test. The relative strengths and weaknesses of these two strategies are compared below to help test engineers select the appropriate method when running parallel tests in production.

Fundamentals of Parallel Testing

The simplest definition of parallel parametric test is "an emerging strategy for wafer-level testing that involves concurrent execution of multiple tests on multiple scribe line test structures." This strategy can help today's highly automated, 24/7 fabs maximize the throughput of their existing parametric test hardware, reduce their cost of test, and lower ownership costs. Furthermore, parallel test offers fabs the flexibility to choose whether they want to increase their wafer test throughput dramatically, or use the saved time to acquire significantly more data, providing greater insight into production processes.

In most cases, the structures being tested in parallel are located within a single Test Element Group (TEG). Even among leading-edge IC manufacturers, very few have progressed to the point of testing structures in different TEGs simultaneously. Implementing parallel test involves using the parametric tester's controller to inter-leave execution of the multiple tests in a way that maximizes the use of processing time and test instrumentation capacity that would otherwise be standing idle. With appropriate test structures design, this *multi-threaded** approach to test sequencing can reduce the execution time for multiple tests on multiple structures to little more than the time needed to execute the longest test in the sequence.

** Thread: The context and code path in which program execution takes place, from start to finish, through a series of tasks.*

Parallel vs. Traditional Sequential Mode Testing

To illustrate the throughput advantage that parallel testing offers, it's helpful to contrast it with the traditional approach to parametric test, in which each test in the sequence must be completed before the next one begins. The total test time for an individual TEG is approximately the sum of the test times for the individual test devices, plus any delays due to switching latencies, which can be significant.

Today's parametric test systems can be equipped with up to eight source-measure units (SMUs), although most systems have fewer installed. Nevertheless, consider a tester equipped with eight SMUs operating in sequential mode for simple tests such as measuring a resistor, which requires one SMU for the two nodes. In this case, seven

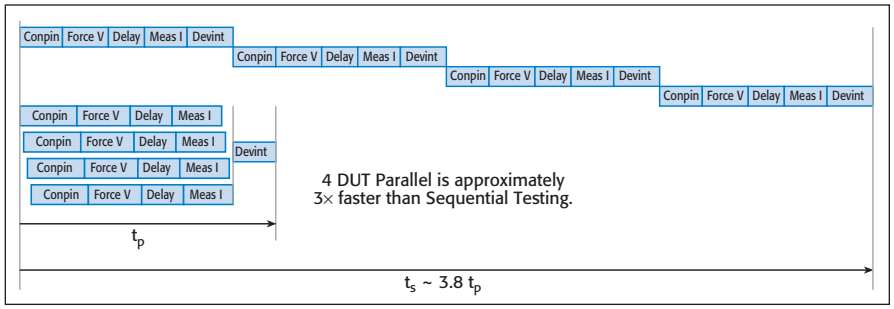

Figure 1. Comparison of elapsed times between sequential and parallel testing of four DUTs. The sequential test time (t_s) is approximately 3.8 times longer than the parallel test time (t_p).

SMUs would be sitting idle. Parallel testing increases utilization of the test cell and boosts throughput by using available tester resources to measure multiple devices simultaneously.

Figure 1 illustrates the difference in time required to perform a set of tests sequentially vs. the same tests performed in parallel. Ideally, parallel tests start simultaneously and chain together with no delays in each thread, but realistically there are slight delays between the start times of each test sequence due to latencies in the prober, controller, and parametric tester. In sequential mode, tests run consecutively, synchronized so that the start of the next test sequence begins upon conclusion of the prior sequence.

Devices tested in parallel may be all the same type (homogenous) or different types (heterogeneous). For example, two transistors, one resistor, and one diode could potentially be measured independently in parallel by performing different connect-force-measure sequences on all four devices simultaneously. ***Figures 2a*** and ***2b*** illustrate the difference between sequential mode and parallel mode testing within a TEG.

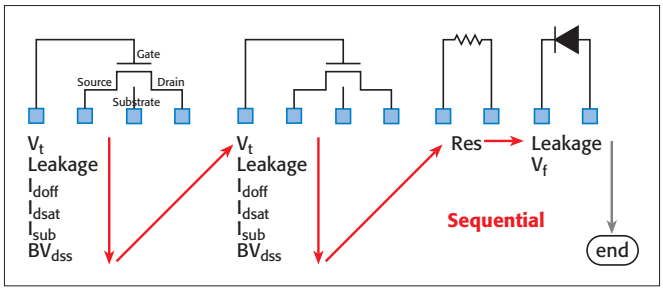

Figure 2a. Schematic of sequential mode testing.

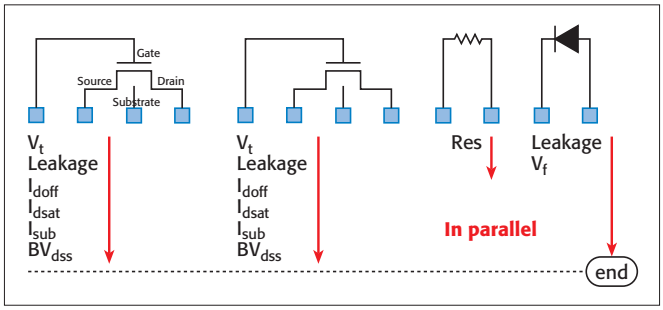

Figure 2b. Schematic of parallel mode testing.

Note in ***Figure 2*** how the parallel mode test sequence maximizes the use of available instrument resources (SMUs, signal generators, etc.). Parallel test has the potential for greatly reducing test times or allowing the collection of more data within a given time frame.

Parallel Test Strategies

The phrase 'independent and asynchronous measurements' is sometimes used to describe multiple instruments operating simultaneously (regardless of when they started) to perform different tests on different devices within a TEG. However, that's not the whole story in parallel testing. There are important issues to be considered in deciding when to assign instrument resources to another test or thread.

Parallel test can involve individual test routines that run concurrently and/or consecutively on multiple threads during overall test program execution. The term "Immediate Resource Recycling" can be used to describe a totally asynchronous parallel test strategy. In this strategy, as a test within a thread ends, the instrument resources it was using become immediately available for reuse in another test, in the same or a different thread (***Figure 3a***). However, users must carefully consider whether this strategy could lead to device interactions and faulty data, and therefore wrong conclusions as to wafer quality and acceptability. The second parallel test method, "Thread Synchronization", starts test sequences in different threads at one time (***Figure 3b***). All the tests within each thread finish, then resources and pins are reassigned before the next set of tests within each thread start.

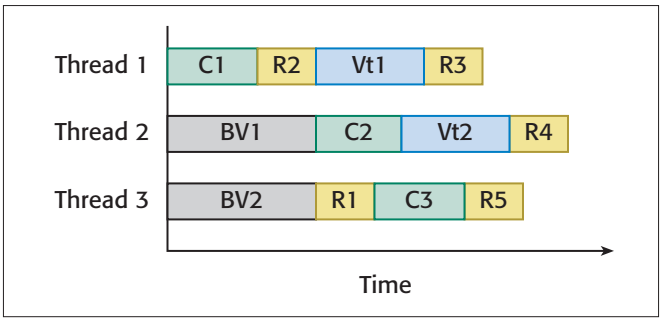

Figure 3a. "Immediate Resource Recycling" Parallel Test.

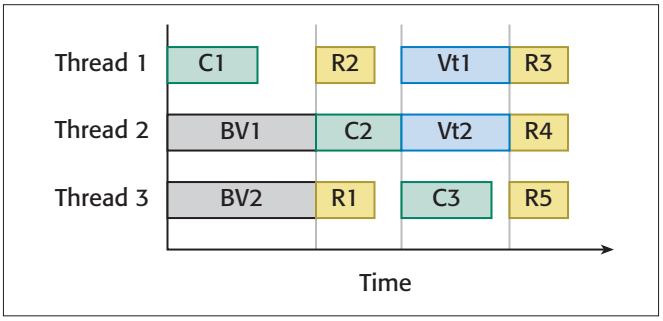

Figure 3b. "Thread Synchronization" Parallel Test.

"Thread Synchronization" is used to ensure that collected data is valid by taking into account device interactions that may take place during parallel testing. This is particularly important when developing a parallel test strategy for existing test structures and TEG layouts. Therefore, when first implementing parallel tests for an existing wafer, results should be correlated with those from the same tests when they ran sequentially.

From its years of experience in parallel testing, Keithley has found that device interactions on different threads can give different results depending on the sequence of tests or due to interactions with adjacent structures and tests. Therefore, some degree of thread synchronization is used by virtually all fabs to ensure data integrity, even if it results in slightly longer test times than using instruments in a totally asynchronous manner using an "Immediate Resource Recycling" strategy. The "Thread Synchronization" parallel test strategy also provides safeguards for tests that must run in a particular sequence.

For example, consider test BV1 in ***Figure 3a***. Typically, tests on failing devices take longer than on passing devices. If the BV1 test

takes longer on a failing device and uses an AC signal, or a substrate bias, or any other type of signal that interferes with the operation of the device used in Vt1, then the test accuracy of Vt1 will be affected as the end of BV1 extends into the beginning sections of Vt1. This type of potential device interaction can be extremely hard to predict during test programming, and almost impossible to evaluate and debug when test engineers are presented with anomalous data from a wafer test run, especially when destructive tests are involved. Lot disposition can become much more complicated. Thread Synchronization parallel test is simpler and more deterministic in implementation and for results analysis.

Another consideration is how parametric testers assign resources. In parallel tests, instruments and pins are essentially "owned" by the first thread that uses them. The tasks running in parallel can't share instruments that vary force conditions or measurements. Similarly, they can't share pins unless they are fixed bias or ground pins as set within the master test sequence. Once single tests complete in a thread, however, the instruments and pins are freed up to be claimed by the next thread and test that needs them.

Optimizing Results

By correlating the results of parallel vs. sequential mode test programs for existing device layouts, one can quickly determine if there are adverse interactions. To shorten this task, Keithley developed pt_execute, a test program characterization and optimization tool. It allows easy switching between sequential and parallel test modes to simplify analysis of throughput improvements and quickly detect sequential vs. parallel correlation issues. Now a feature of Keithley's KTE 5.2.0 and later releases of its parametric tester software, `pt _ execute` automates many of the decisions a test program developer would make. For instance, the software automatically detects the parametric tester hardware configuration and groups tests based on available instrument resources.

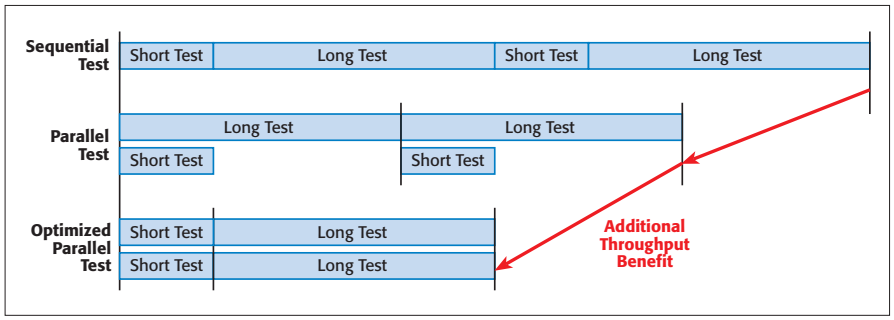

Figure 4. Group tests of similar lengths to gain the highest throughput benefit from "thread synchronization" parallel test.

While performing any portion of a group of tests in parallel will result in some throughput improvement, it may not achieve all the potential timesavings. For example, it's critical to group tests with similar test times to get the greatest benefit from parallel testing using the "thread synchronization" strategy (***Figure 4***). Of course, these tests must not interfere with each other.

Other factors affecting test time include device layout, TEG layout, and subsite test types. In addition, the parametric tester's data communication architecture must be considered. In the case of Keithley's Model S680 parametric tester, parallel testing currently supports a maximum of nine threads: eight VXI communication threads and one GPIB communication thread.

Conclusions

Parallel testing offers parametric test floor managers a tool to increase test cell throughput, increase the amount of data provided (or both) over traditional sequential testing. Still, users must choose a parallel test strategy that provides the best combination of throughput and data integrity. When optimized, parametric parallel testing offers a relatively inexpensive way to increase test throughput without jeopardizing data validity. Just as important, it addresses the growing need to perform more tests on the same structures in less time as device scaling increases the randomness of failures. Additional benefits include:

- Lower cost of ownership
- Lower cost of testing
- Potential for reducing the number of test cells required
- Additional test cell make-up capacity
- Reduced number of test operators and associated training
- Flexibility to test more extensively as desired
- Reduced consumables cost. KEITHLEY

About the Author

Randall Lee is a Senior Industry Market Manager in Keithley Instruments' Semiconductor Test Group where he is responsible for parametric tester product marketing and management. He received a BS in Electrical Engineering from Rensselaer Polytechnic Institute and has 25 years of experience in semiconductor lithography and wafer characterization and test.

KEITHLEY

A GREATER MEASURE OF CONFIDENCE

KEITHLEY INSTRUMENTS, INC. ■ 28775 AURORA ROAD ■ CLEVELAND, OHIO 44139-1891 ■ 440-248-0400 ■ Fax: 440-248-6168 ■ 1-888-KEITHLEY ■ www.keithley.com

BELGIUM
Sint-Pieters-Leeuw
Ph: 02-363 00 40
Fax: 02-363 00 64
www.keithley.nl

CHINA
Beijing
Ph: 8610-82255010
Fax: 8610-82255018
www.keithley.com.cn

FINLAND
Espoo
Ph: 09-88171661
Fax: 09-88171662
www.keithley.com

FRANCE
Saint-Aubin
Ph: 01-64 53 20 20
Fax: 01-60-11-77-26
www.keithley.fr

GERMANY
Germering
Ph: 089-84 93 07-40
Fax: 089-84 93 07-34
www.keithley.de

INDIA
Bangalore
Ph: 080 22 12 80-27/28/29
Fax: 080 22 12 80 05
www.keithley.com

ITALY
Milano
Ph: 02-553842.1
Fax: 02-55384228
www.keithley.it

JAPAN
Tokyo
Ph: 81-3-5733-7555
Fax: 81-3-5733-7556
www.keithley.jp

KOREA
Seoul
Ph: 82-2-574-7778
Fax: 82-2-574-7838
www.keithley.co.kr

MALAYSIA
Kuala Lumpur
Ph: 60-3-4041-0899
Fax: 60-3-4042-0899
www.keithley.com

NETHERLANDS
Gorinchem
Ph: 0183-63 53 33
Fax: 0183-63 08 21
www.keithley.nl

SINGAPORE
Singapore
Ph: 65-6747-9077
Fax: 65-6747-2991
www.keithley.com.sg

SWEDEN
Solna
Ph: 08-50 90 46 00
Fax: 08-655 26 10
www.keithley.com

SWITZERLAND
Zürich
Ph: 044-821 94 44
Fax: 41-44-820 30 81
www.keithley.ch

TAIWAN
Hsinchu
Ph: 886-3-572-9077
Fax: 886-3-572-9031
www.keithley.com.tw

UNITED KINGDOM
Theale
Ph: 0118-929 75 00
Fax: 0118-929 75 19
www.keithley.co.uk

 Printed in the U.S.A. No. 0000 0707