KEITHLEY

WHITE PAPER

Reducing parametric test costs with faster, smarter parallel test techniques

Jeff Kuo, Steven Weinzierl, Keithley Instruments
Glenn Alers, Gregory Harm, Novellus Systems

Introduction

The 1999 SIA roadmap included an ominous prediction—that by about 2012, the cost of test (COT) per transistor would surpass the cost of fabrication per transistor, as indicated in ***Figure 1*** [1].

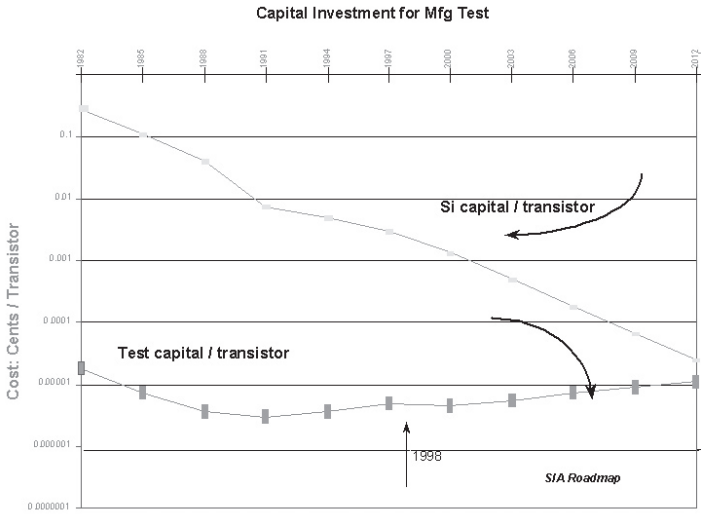

Figure 1. Fabrication and test cost trends.

Although the validity of extrapolating a decade into the future in an industry that delivers a new technology generation of every 2–3 years is debatable, the trend line is nonetheless alarming, particularly for a manufacturing step that, while vitally important, is too often undervalued. Typical strategies for decreasing cost of test include testing less, testing more efficiently, testing differently, and reducing the cost of the testers used [2].

Keithley Instruments, Inc.
28775 Aurora Road
Cleveland, Ohio 44139
(440) 248-0400
Fax: (440) 248-6168
www.keithley.com

Implications for parametric cost of test

Although a wafer's overall cost of test is dominated by the cost of functional testing, many fabs use a common organizational and reporting structure for both parametric and functional testing. This common structure sometimes colors perceptions of the economics associated with parametric test. However, the economics of parametric test differ significantly from those associated with functional test:

- Parametric test uses a sampling strategy, rather than measuring every die on every wafer.
- Parametric test results are used for process control and yield improvement, not for binning finished integrated circuits (ICs).
- Depending on the supplier, equipment in the parametric test cell can often be reused extensively– in fact, a recent analysis indicates it's possible to achieve up to 85% capital equipment reuse over five or more process nodes.
- Parametric test involves measuring a wide array of signal types, ranging from femtoamp DC leakage to 40GHz RF s-parameters.

Applying a typical cost of ownership model to parametric test as used in volume production, then performing sensitivity analysis, reveals that while a 50% decrease in initial capital equipment cost decreases the cost of test per wafer by only 15%, a 50% test time reduction (TTR) delivers nearly a 50% decrease in cost of test per wafer. Ongoing TTR is dependent on choosing a system with a robust, flexible software and hardware architecture that allows performing field upgrades cost-effectively. These upgrades make possible an on-going entitlement to TTR (and therefore COT reduction) in the range of 10% per year. The high value associated with TTR has led us to focus this work on the "test more efficiently" strategy for reducing parametric COT.

More efficient parametric test with parallel test

Although parallel testing is a well-accepted technique for TTR in functional test, it has only recently become available on parametric testers. This is due in part to the complexity of managing the wide range of measurements involved in parametric test, which requires measuring the electrical parameters of the key devices that form the basic building blocks of all integrated circuits: resistors, capacitors, diodes, transistors, inductors, varactors, etc. Measurements are performed on specially designed test structures, typically located in the scribe lines of product wafers. Force-measure sequences can be programmed for single bias

points or as multiple bias points swept in time. For the DC portion of the test suite, a typical set of requirements for measurement resources such as source-measure units (SMUs) might resemble those outlined in ***Table 1***:

Device	Type of Test	Measurement level/ method	Number of SMUs needed
CMOS transistor	Threshold voltage, V_t	Max. g_m of 20 steps	3-4
	Leakage	1pA	2
	I_{doff}	100fA	2
	Saturation current, I_{dsat}	µA to mA	2
	I_{sub}	nA to 10µA, sweep 50 steps	3
	Drain-source breakdown, BVdss	V	1
Resistor	Resistance	µA	1
Diode	Leakage	pA	1
	Forward junction voltage, Vf	force single current	1

Table 1. Example SMU requirements.

A schematic example of sequential mode testing of the devices within a test site might look like ***Figure 2***.

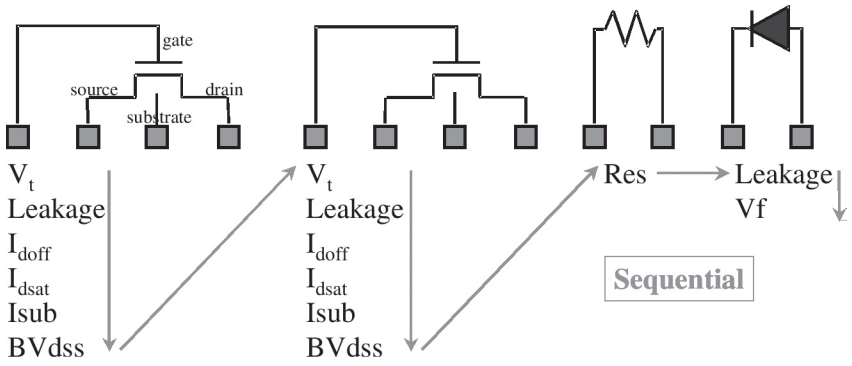

Figure 2. Example of sequential mode testing of devices within a single test site.

A modern parametric test system can have up to eight identical high resolution and high power SMUs. That means that in a sequential test mode, when a resistor is being measured (requiring one SMU), then up to seven SMUs are sitting idle. By measuring multiple mixed types of devices simultaneously within a single probe touchdown and thereby increasing utilization of both the tester and the prober, parallel test delivers higher throughput. For example, two resistors, one diode, and one transistor could possibly be measured

simultaneously by independently and asynchronously performing different connect-force-measure sequences on all four devices at the same time (***Figure 3***).

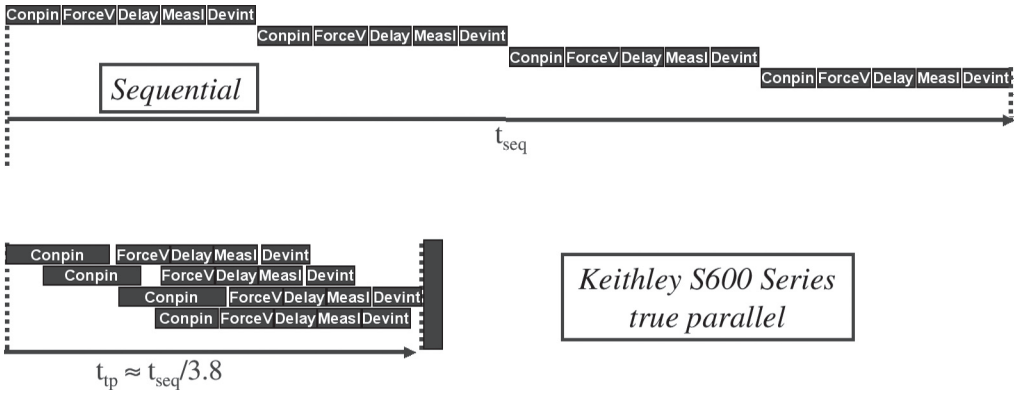

Figure 3.

Figure 4 illustrates a schematic example of parallel mode testing of the devices within a test site that maximizes instrumentation resources.

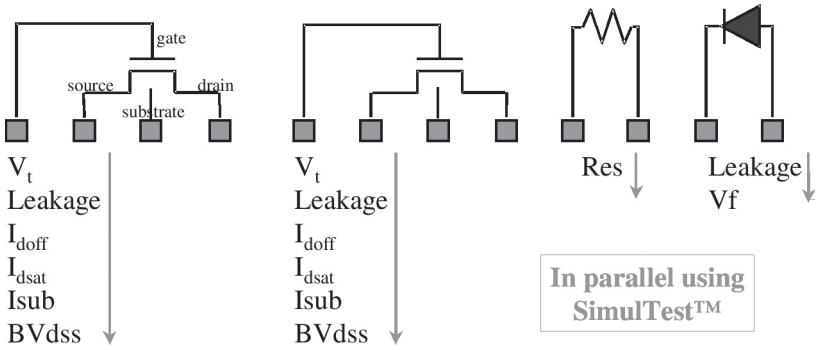

Figure 4.

A parametric test system must have several fundamental hardware and software capabilities to support parallel test:

- Identical, redundant measurement paths, all with lab-grade resolution to support fast test setup and high code reuse, and minimize contention between hardware resources.
- Source and measurement hardware able to run independently. For example, each piece of hardware needs to have its own high precision A/D converters, as well as its own embedded real-time logic processors and communications channels.
- The test execution environment must be multi-threaded.

Test structure design is another important consideration when attempting to get the maximum benefit of parametric parallel test. Depending on the company or fab's philosophy, test structures are typically designed to optimize one of these aspects: minimizing wafer area (lots of shared pads) and/or maximizing the quality of test results (little or no sharing of critical pads). Structures designed to maximize the quality of results typically allow a higher degree of parallel testing. Fabs whose philosophy for sequential test structures leads them to minimize area typically achieve appreciable throughput improvement initially, then can realize additional TTR by making small changes to their test structure designs as they intercept future mask changes.

Parametric test is used primarily during the process development, process integration, and volume fabrication portions of an IC's lifecycle (shown in ***Figure 5***).

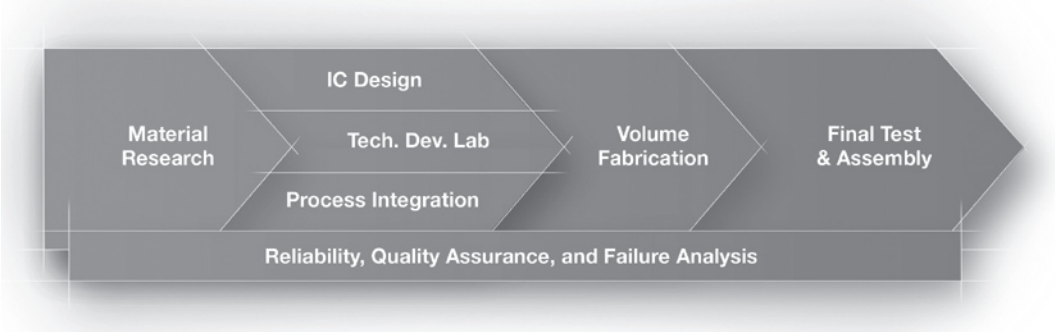

Figure 5.

Throughout the process ramp, the amount of parametric testing performed per wafer ranges widely and constantly, with roughly 100 times more parameters measured during process integration (when the learning curve is steepest) than during volume fabrication (when throughput and incremental yield improvement are more important). Therefore, parametric parallel test offers complementary benefits:

1. The ability to acquire the same amount of data in significantly less time during volume fabrication.
2. The ability to acquire more data in the same amount of test time during process development.

Volume production parallel parametric test—same amount of data in less time

One volume production logic IC manufacturer performs 300 parametric tests per site on the usual variety of device types. Fast integration (17msec) for signal averaging is

used, and the fab's philosophy dictates optimizing test structures for high data integrity—the test devices share few probe contact pads and the scribe line test insert isn't optimized for minimum area. This case allows a high degree of parallelism using existing test structures and probe cards, and the fab achieved 1.7 times higher throughput in measurements at the sites overall, not including prober indexing (wafer movement) time between sites (***Table 2***).

Test mode	Test time per site (sec)
Sequential test	98 sec
Parallel test	56 sec
Test time reduction	42%
Throughput improvement	1.7x

Table 2.

Process development parallel parametric test—more data in the same amount of time

The less obvious benefit of parametric parallel test is the ability to acquire more data in the same amount of time. This use case can occur, for example, during process development, when the learning curve on new materials and devices is the steepest and the opportunity to shorten time-to-market is the greatest. Time-to-market is a primary profitability metric for any IC product. During process development, fabs need to obtain an enormous amount of data quickly for statistical analysis to determine process sensitivity and variability, for verification of process and device models, and for performing corner testing to produce initial process control limits.

Voltage-ramped breakdown (VRB) is one of the reliability tests used during process development to characterize gate capacitors and inter-level dielectrics (ILDs). The test is a very common check of damage to the gate dielectric from a poorly formed oxide or from damage induced by processing. In the case of inter-level dielectrics and the copper damascene process, this test is an important indicator of the integrity of the copper diffusion barrier layer and capping layer interface. The growing use of low κ dielectrics makes this test even more important because of their lower intrinsic breakdown fields and lower interfacial adhesion strengths. The typical test structure for inter-level dielectric reliability (***Figure 6***) in a copper/low κ process is an inter-digitated metal-dielectric comb structure comprised of two parallel metal lines with dielectric between the lines:

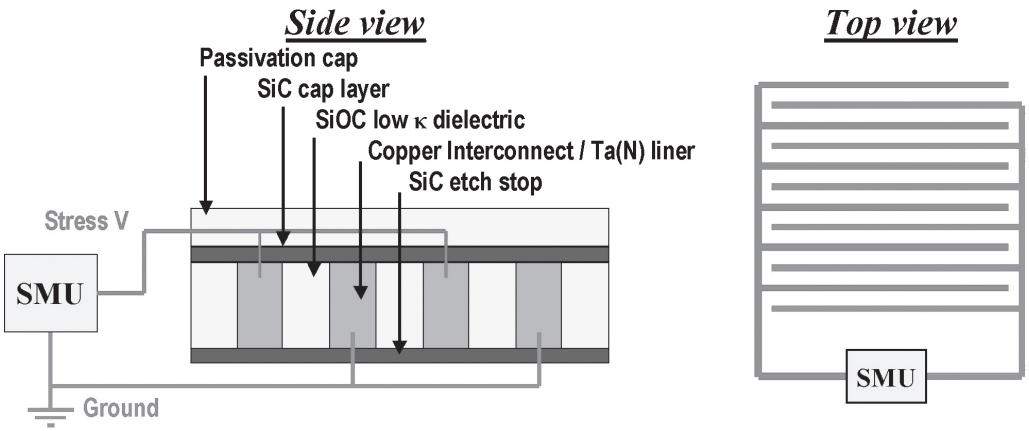

Figure 6.

VRB is a destructive test where voltage across the dielectric is ramped from 0V to as high as 100V, and leakage current is monitored. An abrupt increase in the measured leakage current from one voltage bias point to the next indicates the dielectric has catastrophically broken down and the voltage immediately before breakdown occurred is recorded. Due to the statistical nature of the failure mechanisms, many die are measured across the wafer, and cumulative probability of breakdown voltages is compared between different processes. Test time depends more on the voltage at which the dielectric fails (with good devices taking longer to test) and less on whether multiple DUTs are tested in parallel. A typical ramp rate for the VRB test of comb capacitors with low κ dielectric might be 1V/s. This ramp rate is slow relative to other breakdown tests because the voltages can be quite high and the leakage currents can be transient for low κ dielectrics. If breakdown occurs at 5MV/cm with a 0.2μm dielectric spacing, then it would take 100s to get to the 100V breakdown voltage. Faster ramp rates would result in even higher breakdown voltages because the effective time at a voltage is shorter. Such high breakdown voltages might exceed the voltage limit of the tester or change the failure mechanism in the low κ comb structure.

Given the relatively long time needed to perform this test, the number of die tested must be limited to obtain reasonable test throughput. A standard parametric test sampling strategy might be to measure only 16 out of the 121 die available on the wafer, and only one of the 12 structures available within a die (***Figure 7***).

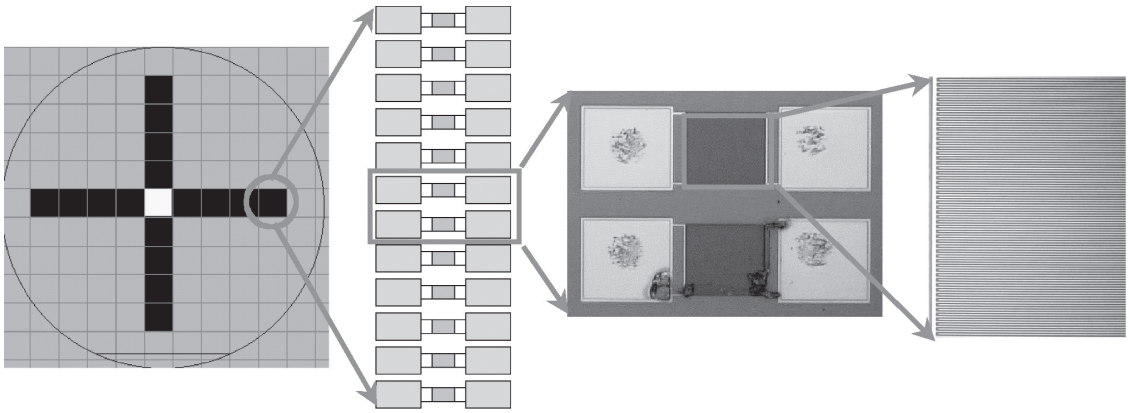

Figure 7.

This strategy was believed to provide an optimal tradeoff between test time and amount of data. Process effects such as dielectric erosion and other phenomena were always observed to occur on spatial scales consistent with the chosen die sampling that spanned the wafer, so it was not believed that measuring more structures in closer proximity (more than one device per die) would provide any additional process information.

The test time for 16 die was approximately one hour. Because it was modeled that measuring four DUT in parallel within the same die would not increase the test time and in the interest of discovering new processing phenomena, three more DUTs (#2, 3, 4) were measured in each site (***Figure 8***).

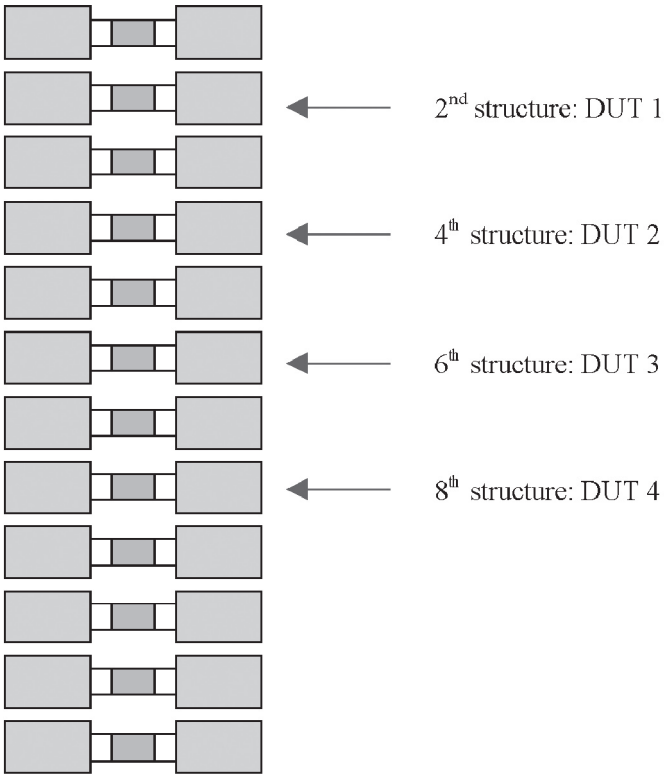

Figure 8.

Figure 9 is the resulting Cumulative Probability Plot (CPP) of VRB test results from the test wafer.

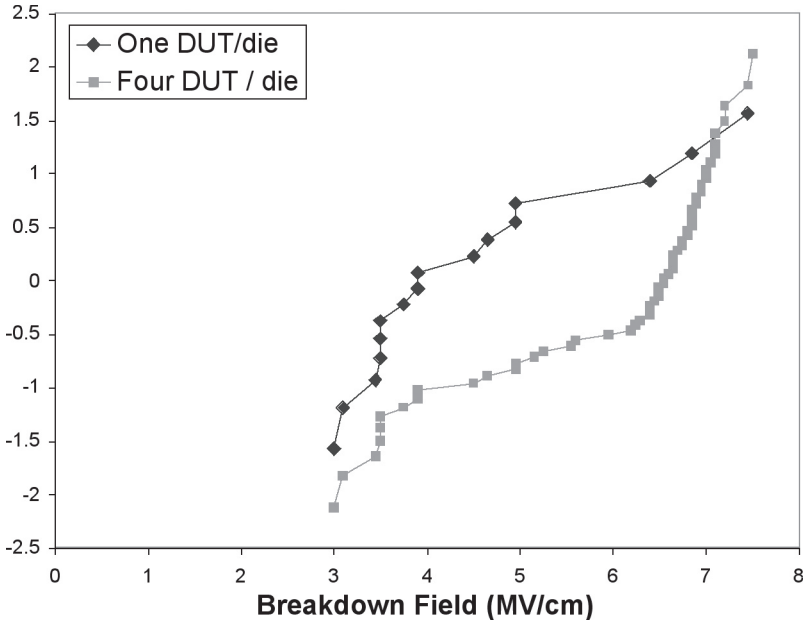

Figure 9.

When only one DUT per die was tested, the median breakdown field was ~4MV/cm and the distribution was a very broad Gaussian, with no sign of multimode failures. Based on this data set, one might conclude the integrity of the low κ dielectric layer was compromised across the whole wafer, so the wafer and the process it represents should be rejected. However, the curve for the four DUTs combined that was acquired in nominally the same test time shows that the median breakdown field was 50% higher at 6MV/cm, and the distribution appears to be bimodal. A bimodal distribution indicates there might be a very local process issue affecting the integrity of the low κ dielectric, but the general integrity of the low κ dielectric is good. This conclusion is significantly different than the one drawn from the one-DUT-per-die curve. Failure analysis of the failed die showed localized cracking of the dielectric passivation layer in the neighborhood of the die during test (***Figure 10***).

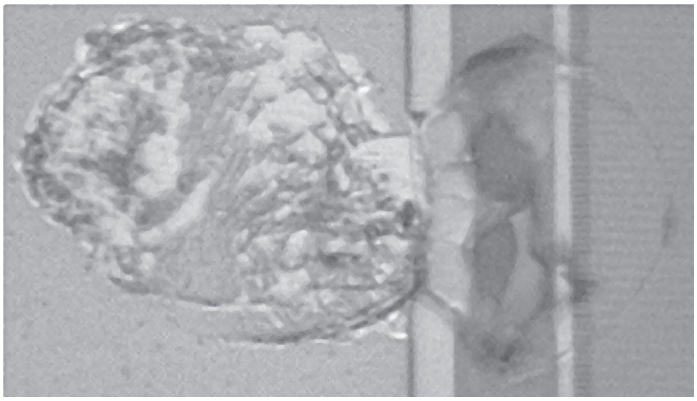

Figure 10.

This cracking of the passivation layer was sufficient to degrade the breakdown properties of the low κ dielectric.

Summary and conclusions

The advent of parallel testing capability takes modern parametric test systems out of the extrapolation that predicts wafer test cost will exceed wafer fabrication cost in three technology nodes. Parallel parametric test has been shown to deliver the same data in substantially less test time in the volume production use case. It was also shown to deliver substantially more data (and learning) in the same test time during process development, with the discovery of a new Cu/low κ process integration effect. By coordinating the development of test structures for parallel test with scheduled mask changes, parallel parametric test also provides ongoing opportunities for decreasing parametric cost of test in volume production.

Acknowledgments

We would like to thank Peter Griffiths and Carl Scharrer for their contributions to this article.

References

1. Sengupta, Sanjay, et al. "Defect-Based Test: A Key Enabler for Successful Migration to Structural Test." *Intel Technology Journal*. 1st Quarter 1999. <http://www.intel.com/technology/itj/q11999/articles/art_6c.htm> (29 Mar. 2004).
2. Carlson, Steve. "ATE struggles to Keep pace with VLSI." *EE Times*. December 13, 2001. <http://www.us.design-reuse.com/articles/article2278.html> (29 Mar. 2004).

About the authors

Jeff Kuo is a Senior Applications Engineer at Keithley Instruments, Inc. He received his Ph.D. in Materials Science and Engineering from the University of Texas at Austin. Dr. Kuo can be reached at jkuo@keithley.com.

Steven Weinzierl is a Product Marketer in the business development group at Keithley Instruments. He received his Ph.D. in 1992 from Cornell University, and has also held a variety of technical and marketing positions at KLA-Tencor and Solid State Measurements. His focus has been electrical and optical measurements of advanced materials, and can be reached at sweinzierl@keithley.com.

Glenn Alers is Senior Process Manager in the integration group at Novellus Systems, Inc. He is responsible for reliability characterization of materials for copper/low κ interconnect structures. He received his Ph.D. in physics in 1991 from the University of Illinois, Urbana-Champaign.

Gregory Harm is a Test Engineer at Novellus Systems, Inc. He graduated from the University of California at Santa Cruz with a B.S. in Physics.

KEITHLEY

Keithley Instruments, Inc. 28775 Aurora Road • Cleveland, Ohio 44139 • 440-248-0400 • Fax: 440-248-6168
1-888-KEITHLEY (534-8453) • www.keithley.com

No. 2499
4042KGW