KEITHLEY A GREATER MEASURE OF CONFIDENCE

Reducing Parametric Test Costs With Faster, Smarter Parallel Test Techniques

Jeff Kuo, Steven Weinzierl, Keithley Instruments

Glenn Alers, Gregory Harm, Novellus Systems

The 1999 SIA roadmap predicted a transistor's cost of test (COT) would exceed its fabrication cost by 2012. COT reduction strategies include testing less, testing more efficiently, testing differently, and reducing the cost of the testers used [1]. Applying a typical cost of ownership model to parametric test in volume production, then performing sensitivity analysis reveals that, while cutting initial capital equipment cost by 50% decreases COT per wafer by only 15%, a 50% test time reduction delivers nearly a 50% COT decrease per wafer. Obviously, "testing more efficiently" is the more effective strategy for reducing parametric COT.

Parallel parametric testing

Today's parametric testers can have up to eight source-measure units (SMUs). When measuring a resistor sequentially (requiring one SMU), up to seven SMUs are idle. By measuring multiple device types simultaneously within a single probe touchdown and increasing hardware utilization, parallel test increases throughput significantly. For example, two resistors, one diode, and one transistor could be measured simultaneously by asynchronously performing different connect-force-measure sequences on all four devices at once. Details on parallel test implementation and test structures are available elsewhere [2].

Parametric parallel test offers different benefits depending on when in a given process node's lifecycle it is being used. It allows acquiring more data in the same test time during process development, or the same amount of data in less time during volume fabrication.

Volume production—same amount of data, less time

One logic IC manufacturer performs 300 parametric tests/site on a variety of devices. Fast integration (17ms) for signal averaging is used, and the fab's philosophy dictates optimizing test structures for data integrity. Test devices share few probe contact pads and the scribe line test insert isn't optimized for minimum area, allowing significant parallelism with existing test structures and probe cards. Parallel testing let this fab achieve 1.7× higher throughput in measurements at the sites overall, excluding wafer movement time between sites (***Table 1***).

Table 1.

Test mode	Test time per site (seconds)
Sequential test	98 s
Parallel test	56 s
Test time reduction	42%
Throughput improvement	1.7 ×

Process development—more data, same time

Acquiring more data in the same time is invaluable during process development, when the learning curve is steepest and the opportunity to shorten time-to-market is greatest.

Voltage-ramped breakdown (VRB) is a reliability test for characterizing gate capacitors and inter-level dielectrics (ILDs). For the copper damascene process, it's an important indicator of copper diffusion barrier layer and capping layer interface integrity. The typical test structure for ILD reliability in a copper/low-κ process is an inter-digitated metal-dielectric comb structure. In this destructive test, voltage across the dielectric is ramped from 0V to as high as 100V, while leakage current is monitored. An abrupt leakage current increase indicates the dielectric has catastrophically broken down; the voltage bias immediately before breakdown is recorded.

The statistical nature of the failure mechanisms requires measuring many die across the wafer, with cumulative probability of breakdown voltages compared between different processes. Test time depends more on the voltage at which the dielectric fails (good devices take longer to test) and less on whether multiple devices under test (DUTs) are measured in parallel. A typical ramp rate

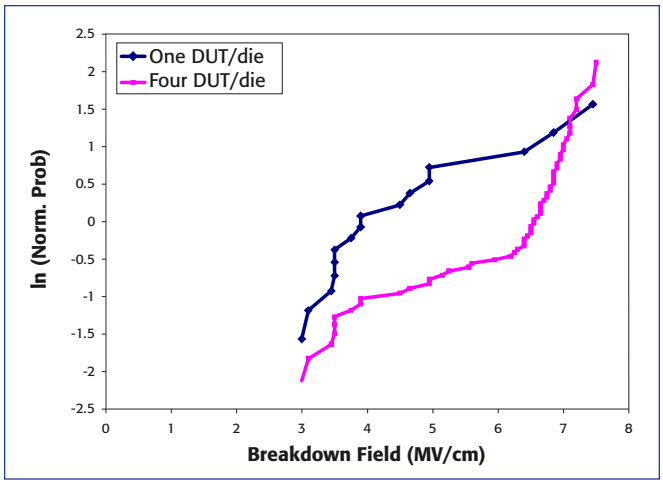

Figure 1.

for the VRB test of comb capacitors with low-k dielectric might be 1V/s, which 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 0.2mm dielectric spacing, it would take 100s to reach the 100V breakdown voltage. Faster ramp rates would produce even higher breakdown voltages (potentially exceeding the tester's voltage limit or changing the comb structure's failure mechanism) because the effective time at a voltage is less.

This test's length requires limiting the number of die tested to obtain reasonable throughput. A standard wafer sampling strategy might be to measure 16 of the 121 dies available, and only one of the 12 structures within a die. Process effects like dielectric erosion and other phenomena always occurred on spatial scales consistent with the chosen die sampling that spanned the wafer, so measuring more structures in closer proximity (more than one device per die) wasn't thought to provide additional process information.

The test time for 16 dies was approximately one hour. Measuring four DUTs in parallel within the same die wouldn't increase the test time, so three more DUTs were measured in parallel at each site in the interest of discovering new processing phenomena. ***Figure 1*** is the resulting cumulative probability plot of VRB test results.

When testing only one DUT per die, the median breakdown field was ~4MV/cm and the distribution was a very broad Gaussian, with no sign of multimode failures. One might conclude the dielectric layer's integrity was compromised across the wafer, so it and the process it represents should be rejected. However, the curve for the four DUTs combined, acquired in nominally the same test time, showed the median breakdown field was 50% higher at 6MV/cm, and the distribution appeared bimodal, indicating there might be a localized process issue affecting the dielectric, but its general integrity was good. This conclusion differs significantly from the one drawn from the one-DUT-per-die curve. Failure analysis showed localized cracking of the dielectric passivation layer near the die during test, sufficient to locally degrade the dielectric's breakdown properties.

Conclusions

Parallel parametric test delivers the same data in substantially less time in volume production and substantially more data (and learning) in the same time during process development. When test structure development for parallel test is coordinated with scheduled mask changes, there are ongoing opportunities for decreasing parametric COT in volume production. KEITHLEY

References

1. 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).
2. Jeff Kuo, Steven Weinzierl, Glenn Alers, Gregory Harm, "Reducing parametric test costs with faster, smarter parallel test techniques." http://www.keithley.com/servlet/Data?id=15590.

About the Authors

Jeff Kuo (jkuo@keithley.com) is a Senior Applications Engineer at Keithley. He received his Ph.D. in Materials Science and Engineering from the University of Texas at Austin.

Steven Weinzierl (sweinzierl@keithley.com) is a Product Marketer in the business development group at Keithley. He received his Ph.D. in 1992 from Cornell University.

Glenn Alers is Senior Process Manager in the integration group at Novellus Systems. He's responsible for reliability characterization of materials for copper/low-k 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. 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. 2530
0704