# TECHNICAL NOTE

TN-107

# IMD and Other High Voltage Testing

## Introduction

Trying to use standard dc parametric test systems for high voltage testing leads to many problems. So many, in fact, that such testing is seldom even attempted. Instead, high voltage testing has historically been done manually with curve tracers [1] or bench instruments. Even when it was justifiable as a 100% screening test, adding high voltage testing to an automatic parametric tester seldom proves effective or practical.

DC parametric test systems typically span a 200V range, but historically that has not been much of a limitation. That is because most transistor, resistance, and oxide measurements require only a few tens of volts for characterization and monitoring. Furthermore, the trend is toward even lower voltage processes for high volume digital integrated circuits.

Nevertheless, process testing of thick dielectrics requires much more than 200V in order to assure adequate quality and reliability. Additionally, monitoring of power IC manufacturing often requires much higher voltages.

In order to automate such applications, Reedholm has developed a high voltage testing capability useful to 1500V. Originally developed for a power IC customer, the high voltage capability has been extended from providing continuous I/V characteristics on one pin to measuring snap-back breakdown of inter-metal dielectric (IMD) test structures on any of four device pins.

## Types of Breakdown

From a test viewpoint, two distinctive types of breakdown occur in thick oxide or IMD structures: continuous and snap-back. Of the two, snap-back testing is far harder to accurately determine.

### Continuous Breakdown

High voltage transistors and diodes typically exhibit continuous, single valued current-voltage characteristics when voltages approach avalanche breakdown levels. If the current available is kept low enough, electron multiplication [2] during avalanche does not lead to device damage. Thus, degradation of the operating characteristic is not affected by operation in the breakdown region, and structure testing produces consistent, reproducible results.

Figure 1 illustrates device behavior in the low current region for both snap-back and continuous breakdown.

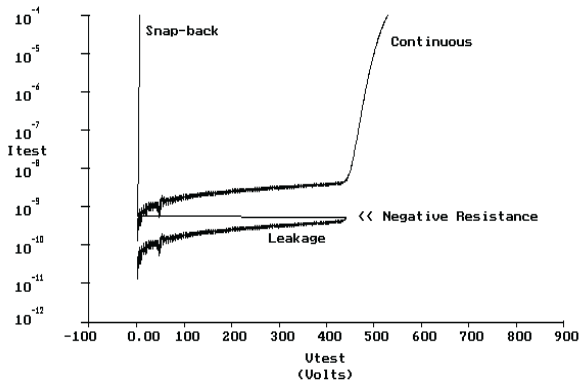


Figure 1 - Breakdown Characteristics

Few special or difficult requirements are placed upon a parametric test system if structures exhibit continuous conductivity throughout the range of test. As power levels can be kept low, it is only necessary to provide high voltage switching and good current sensitivity.

### Parametric and Power Device Testing

That is not to say that high speed, high power testing of power devices is simple. For instance, Reedholm parametric testers are extremely fast for the dc parametric test industry. With Reedholm systems, precise measurements are made in less than 100$\mu$sec while delivering over 50 watts of power. Yet those times are one to three orders of magnitude too slow, and the power level one to two orders of magnitude too low, for functional testing of power devices. For those reasons, testing of power devices is not covered further in this note.

### Snap-back Breakdown

Unlike the continuous breakdown exhibited by junctions, dielectrics and thick oxides exhibit discontinuous, snap-back breakdown which is destructive. Testing is difficult because the conversion from insulator to conductor at breakdown is extremely fast and because a test can't be repeated once a device has failed.

To complicate matters further, the rapid drop in voltage caused by breakdown draws so much energy from test system cable and instrumentation capacitance that the test structure is usually destroyed by localized heating. Even when attempts are made to limit power delivered to the device under test after snap-back, the energy stored in the fixturing capacitance is often enough to produce catastrophic breakdown and complete destruction of the device.

Trying to use low field tunneling current measurements for detection of dielectric thinning has proven fruitless because the disruptive breakdown processes [3], whatever their causes, typically occur at lower fields than would induce measurable tunneling current. That is why high voltage ramps are almost exclusively used to detect effective thinning.

## Why Dielectric Testing is Done

As difficult as it is, benefits of automatically testing dielectric strength have justified many attempts at solution, both visual and electrical. Unfortunately, most dielectric defects between metal layers or poly to metal layers can't be viewed using optical techniques. Electrical tests are often the only way to find dielectric quality problems and then prove their elimination.

While some problems can be severe enough to reduce yield, their presence is usually not felt until field failures result [4]. As there should seldom be any thinning in a controlled process, the presence of it beyond the deposition control limits would indicate a tail in the failure distribution. In these days of customer demands for 10 FIT reliability levels, such a tail might prove disastrous.

Some of these problems arise even with seemingly stable processes, so it isn't enough to find and eliminate problems only in the development phase or as a one time fix. Well after characterization of a new or fixed dielectric, adequate sampling must be maintained to assure that neither generalized nor patterned problems occur.

- Hillocks and other asperities in conducting layers can produce significant thinning in overlying glass which, in turn, places less compressive stress on that protrusion than on the rest of a metal line. Too much thinning can lead to premature wearout from shorts between conducting layers or to opens in the lower metal lines. Wearout happens under operating conditions since material accumulation due to electromigration will happen more quickly in a region of lower physical stress.
- Dielectric voids from incomplete dielectric coverage also permit excessive metal migration (electro- or stress) and subsequent early wearout due to inter-layer shorts or opens in metal lines.
- Large area dielectric thinning doesn't normally lead to wearout failures, but could affect sensitivity to ESD [5]. Since thinning can result in breakdown at significantly lower fields than implied by dielectric strength, higher on-chip transient voltages during ESD may occur than was designed for.
- Dielectric cracks can grow in cross section due to low level current flow causing excessive conduction between layers and eventual loss of proper circuit operation.

## Limits of Tunneling Measurements

The difficulty of making high voltage measurements has led some to the intuitive approach of using the highest voltage available (typically 100 to 200V) in a test system and measuring the resultant leakage. Inherent in such an approach are assumptions that:

- Thinning of 1$\mu$m dielectrics will be detected at 200V.
- Problems not detectable at 200V aren't a concern.

The flaw in the assumptions can be seen by considering the current/field relationship of Fowler-Nordheim (FN) tunneling shown in Figure 2. There just isn't enough current flowing at 2MV/cm (i.e., 2V at a thickness of 10nm) to be measured. Although the FN equation seems to hold over a wide range of currents, there are practical limitations to finding the current associated with each voltage. Note the limits that probe cards place on current sensitivity in practical automatic systems.

Even with the most sensitive current measurement instrumentation, stress voltages need to be a significant portion of the dielectric breakdown in order to cause measurable tunneling current at fields below 5MV/cm.

While oxide trapping [7] can affect the shape of the curve at high currents, tunneling current falls off rapidly at field strengths well below dielectric breakdown. A few equations are derived below for estimation of current

sensitivity test requirements. In the FN equation, current density (J) in amps/cm$^2$ is related to the electric field (E) in MV/cm and the constants A in amps/MV$^2$and B in MV/cm.

$$J = AE^2 e^{-B/E} \qquad (1)$$

While use of electric field and current densities are appropriate when describing physical behavior, from a test viewpoint only the voltage and current levels are important.

By introducing current (I) and area (a), Equation 1) can be re-written as:

$$I = aJ = aAE^2 e^{-B/E} \qquad (2)$$

Converting Equation 2) by taking its logarithm:

$$\log(I) = \log(aA) + 2\log(E) - \left(\frac{B}{E}\right)\log(e) \qquad (3)$$

In this note, values for constants A (1.25E+6 A/MV$^2$) and B (233.5MV/cm) were taken from the Nissan-Cohen et al. [8] paper which assumed field emission from the silicon surface.

For a structure area of 10,000$\mu$2 the current vs field relationship becomes:

$$\log(I) = 2.194 = 2\log - \left(\frac{101.4}{E}\right) \qquad (4)$$

The shape of the FN graph of log (I) is dominated by the last term in Equation 4), so except for an area dependent offset, the log of the tunneling current changes inversely with applied field. The second term is not much of a factor for fields below 5MV/cm, but its effect increases the value of I by two decades when the field reaches 10MV/cm.

## Measurement Sensitivity Considerations

In practical terms, a 500 volt bias on a 1$\mu$m thick oxide with an area of 10,000$\mu$m$^2$ produces a 5MV/cm field and a predicted tunneling current of around 0.2fA. The capacitor area would have to be five orders of magnitude larger, or 10cm$^2$, to produce a predicted current in the picoampere region.

For volume testing of oxides using an automatic prober, it is practical to make measurements down to the 100fA level, but even that level would require a 1cm x 1cm structure for detection if no thinning defects were present.

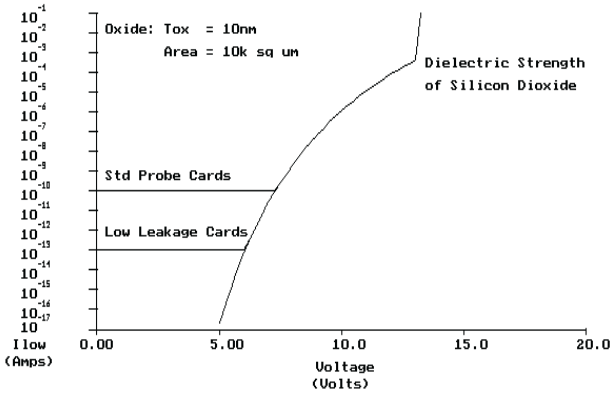


Figure 2 - Idealized FN I-V Graph

## Impact of Thinning on Tunneling Current

The preceding comments illustrate that tunneling current measurements are only practical at relatively high fields. Thus, an insulator that insulates near zero volts yet conducts current at voltages well below intrinsic breakdown would be characterized as severely thinned.

If it were possible to generate tunneling currents in a 1$\mu$m dielectric, the plot shown in Figure 2 would apply for a 10,000$\mu$m$^2$ device except that the voltage axis would be scaled by 100:1. Such a device would require 700V of bias to produce a 10pA current. Thus, one could say that a detectable current of 10pA at 700V would imply a dielectric thickness of 1$\mu$m. Similarly, a detectable current of 10pA at 200V would imply a thickness of (200/700) x 1$\mu$m.

If the stress supply were limited to typical test system limit of 200V, the current sensitivity were 10pA, and the dielectric strength were 10MV/cm, it would not be possible to detect thinning below 71%, or 500V/700V. Such a limit would be intolerable for hillock detection since it wouldn't detect thinning until virtually the entire dielectric was compromised.

## Considerations with HV Ramps

It is unfortunate that tunneling current measurements fail to work effectively with thick dielectrics. Since they do not, one is faced with three basic problems in high voltage breakdown testing:

- Delivery of voltages >1000V to the test structure
- Detection of rapid breakdown
- Proper system operation after breakdown

### Matrix Switching

DC parametric test systems use crosspoint matrix switching systems to connect system instrumentation to the device under test. Relays in the switching system are usually capable of switching higher voltages than the instrumentation can deliver, but the upper limit is typically five to six hundred volts.

Reliable, higher voltage relays are available, but are not of the same quality and compactness. For instance, each crosspoint shown in Figure 3 would have all the relays (force, sense, and guard) housed in one package for standard dc systems. At higher voltages, each relay would likely be placed in separate package.

Saying that the relays need to switch high potentials is not really correct. In order for relays to have reasonable lifetimes, prevention of switching with voltages applied is done in software by dc parametric test systems. Thus, it is not required that high voltages be switched, but that relays programmed "open" stay that way and do not break down and conduct current in the presence of high voltage.

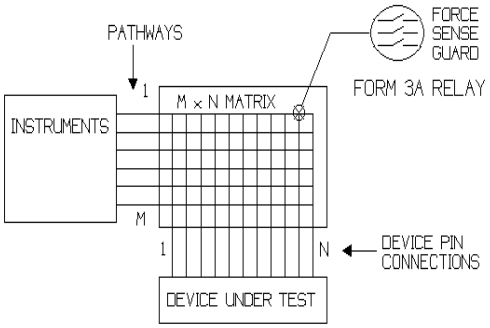


Figure 3 - Matrix Block Diagram

### Need for Rapid Detection

While the exact nature of dielectric breakdown isn't nearly as well characterized or modeled as semiconductor junction breakdown, it is extremely fast compared to response times of dc parametric test systems and associated cabling. Thus a breakdown event can be modeled as a time step function relative to a dc test system as in Figure 4.

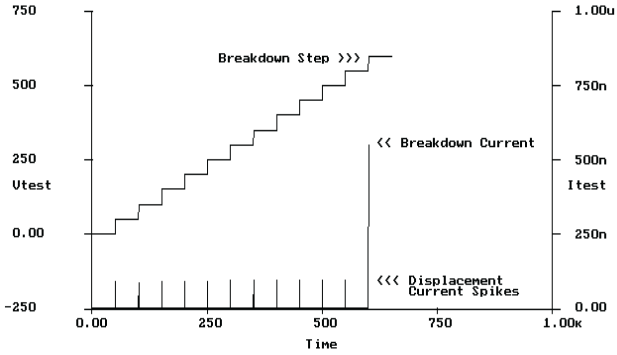


Figure 4 - Simple Model Timing Relationship

### Uncontrolled Charge

Another consideration is the uncontrolled delivery of charge during breakdown. Being quite thick, even fairly large area test structures have relatively little capacitance compared to the cabling leading to the device under test. As breakdown is effectively a shorting out of the structure, it causes virtually instantaneous discharge of energy stored in the system instrumentation and associated cabling. Figure 5 is a simple model of destructive dielectric in which closure of switch $S_1$ represents breakdown.

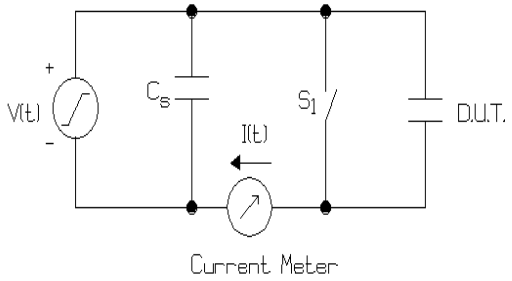


Figure 5 - Simple Model: S1 closes when V(t) = Vbk

Although the shield capacitance can be effectively reduced to zero through active guarding, the guard circuit amplifiers have several orders less bandwidth than would be required to eliminate system capacitance during breakdown. DC parametric testers can have 500pF of system capacitance per connected device pin, so energy discharge can be substantial for a micro-electronic test structure.

That charge is usually large enough to lead to a permanent conductive path between plates of a test structure. However, if the test system operation is not sufficiently fast and well controlled, excess energy flowing into the structure for even a few milliseconds [9] can cause extremely high temperatures in the defect region. In some cases, temperatures can get high enough to blow a crater in the top plate and overlying passivation of the test structure.

When structures are robust or when large area asperities are the cause of failure, cratering may not occur, but sufficiently slow instrumentation can supply enough current to open the lines connecting the structure to the probing pads, thereby causing the structure to appear electrically "open."

Unless the failures are to be analyzed, physical destruction of the structure is unimportant. However, automated testing requires failure be confirmed electrically, and a catastrophic opening of the failure through cratering or opening of a line masks the failure by measuring "open" at low voltages after the breakdown event has occurred.

## Practical Stress Supplies

Very fast (for dc parametric test) instrumentation is required to detect breakdown before a structure is destroyed through excess charge dumping. While IEEE-488 controlled instrumentation could possibly be used for this type of testing, the detection, interpretation, and stress shutdown all have to occur in one to two milliseconds. Typical IEEE-488 applications require 10 to 20 milliseconds for each bus action, so it is not unusual for over 100 milliseconds to elapse before a breakdown ramp can be halted under software control using IEEE-488 instruments.

A supply with current limiting might seem to be a way to limit charge flow. However, the delay before current limiting starts produces enough charge (which adds to the charge stored in the supply's output capacitance) to destroy the test structure. A solution is needed which avoids those inherent power supply limitations.

## Proper Operation After Breakdown

Since dielectric breakdown at high voltage is a very fast event compared to the time constants of the cabling leading to the test structure, it acts as a step function to the cabling. That produces a problem as the cabling for dc parametric testing has very high Q due to low series and high shunt resistance. Resonance in the cabling occurs in the 5 - 20 MHz range depending on the overall signal path.

A breakdown pulse of 1000V could produce many times that voltage in a damped sinusoid if other elements in the path did not reduce the peak potential. Nevertheless, the breakdown and resultant amplitude multiplication act as a spark transmitter with the un-shielded DUT leads acting as the broadcast antenna. Nearby equipment could be affected by the noise transmission even without direct connection to the test system.

Immediately after breakdown (represented in Figure 6 by closure of $S_1$ with $V_R$ at the breakdown potential), the current flow is limited only by the RLC networks used to model the shielded cables that connect the device under test to the test system.

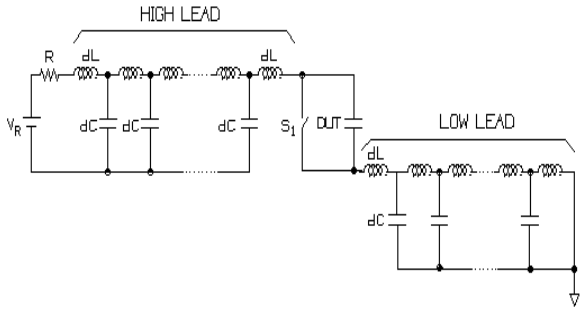


Figure 6 - Cabling Shown as Transmission Line

If the device doesn't rupture (and completely open) or if $V_R$ isn't reduced to zero volts, the current flow would be limited only by the total series resistance of the high and low cables. That resistance is < 1 ohm in a typical system. Thus the available instantaneous current is practically unlimited with a short time constant as shown in Figure 7.

Of far greater effect is the conducted current which flows through signal paths back to earth ground. The path through the prober can be great enough to cause digital control regis-

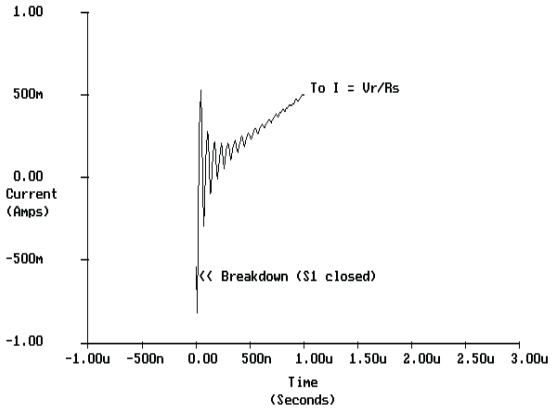


Figure 7 - Current Response at Breakdown

ters inside the prober to reset or re-program. Those who do high voltage functional testing regularly address such problems in order to achieve reliable prober operation.

Similarly, the ground current can cause enough voltage drop due to inductance in the digital grounds that the digital circuits which control the matrix relays (and other instrument actions) can be falsely set. Thus, inappropriate pathways can be connected, instrumentation ranges changed, etc. as a result of breakdown. If detection of the breakdown event is too slow, changes in the instrumentation could alter the test set up so much that the breakdown event would be missed entirely.

## Automating Dielectric Testing

Because of the numerous problems associated with configuring a traditional dc parametric matrix, instrumentation, and software for dielectric testing, Reedholm chose an approach which, by design, addresses the problems described and maximizes data quality.

A four-pin high voltage matrix section, high voltage current-controlled supply, and associated cabling are used in conjunction with the basic Reedholm dc parametric test instrumentation to provide dielectric and device breakdown testing up to 1500V. Some of the characteristics include:

- Detection of breakdown within 75$\mu$s
- Voltage shutdown within 50$\mu$s of event detection
- Programmable ramp step times as short as 1ms
- Cabling and instrument capacitance < 10pF
- Leakage current measured prior to start of ramp
- Mode selection between continuous and snap back
- Use of result categories similar to JEDEC Vramp
- Generation of error categories for trouble shooting
- Full integration into Reedholm application software

With high speed detection and stepping, breakdown ramps can be run from 0 to 1000V in approximately 100ms to a resolution of 10V. Testing to better resolution is seldom required, but would take proportionally longer. Conversely, the test could be made faster with larger steps.

### Operating Modes

Whether in continuous or snap-back operating mode, standard Reedholm instrumentation is augmented with a high voltage power supply configured as an inverting operational amplifier (see Figures 8 & 9). Under software control, PS#1 is used to sink current which can only pass through the 2MΩ feedback resistor of the high voltage amplifier. The operational gain of the amplifier keeps the inverting input at virtual ground, so any voltage developed across the feedback resistor appears at the supply output. Thus, setting PS#1 to sink 500$\mu$A results in an output voltage of +1000V (500$\mu$A x 2MΩ).

The operating modes use different measurements. For snap back, the delivered voltage is monitored; for continuous mode, current through the device is measured.

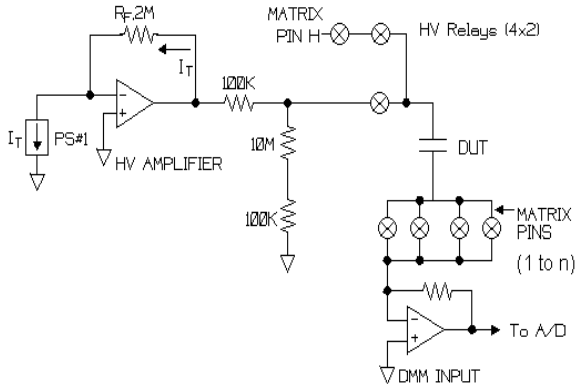


Figure 8 - Continuous Mode Operation

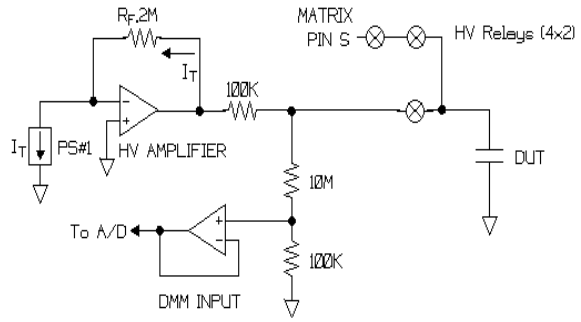


Figure 9 - Snap-back Mode Operation

### Test Operation

An energy limiting output resistor (100k), a 101:1 voltage divider (10M & 100k), plus eight high voltage relays are placed in a small box mounted close to the probe card on the prober analog cable. Test voltages and/or currents can be routed to the device under test via the regular instrument matrix or the high voltage supply matrix.

During the test, an initial current measurement is made at the starting voltage. If it is too high, the test is terminated and a value of 9.0E+21 returned. If the DUT isn't shorted, the ramp continues with the voltage step specified by the user. At each step, as many measurements as possible are made during the step time specified

by the user. The test terminates either when the ramp reaches compliance, when the current exceeds the user specified amount, or when the measured voltage drops by the user specified amount.

When the ramp terminates, high voltage supply current is immediately reduced to zero and all relays connecting the high voltage are opened. As a result, the high voltage output returns to zero volts within fifty microseconds.

### Maximizing Detection Speed

Fast, low level software procedures are used to minimize the possibility of missing a breakdown. For instance, an 80386 based PC running at 20MHz takes around 500$\mu$s to complete the standard Reedholm current forcing command needed for each high voltage step. For most applications, the 500$\mu$s is virtually instantaneous and doesn't need to be faster. However, during that time the computer's CPU is busy executing code, and the breakdown event might be missed. For that reason, a fast "AMPS" command was written to provide current steps within 20$\mu$s on the same computer. This allowed a reduction in measurement (or detection time) from 700$\mu$s to 70$\mu$s. At those speeds, breakdown can always be detected.

Since device connections are sometimes opened by the breakdown event, it is necessary that event detection as well as re-establishment of hardware control occur before the 400$\mu$s it takes for relays to start opening and/or closing.

### Eliminating External Compromises

Asynchronous computer events that out prioritize the high voltage software operation need to be avoided. Thus the keyboard has to be locked out, and any networking interrogation suspended for the duration of high voltage testing. Obviously, a dedicated processor is needed to execute test code with such control, so a multi-tasking computer is unsuitable for this class of testing. Additionally, probe card design must be capable of withstanding repeated testing at the test limits without degradation.

Because the probe card can be the limiting factor in high voltage testing, characterization of cards in a "probes up" condition needs to be done. Fortunately, the high voltage test code can be used for that characterization.

## Summary

Extensive evaluation has been done with continuous mode operation from 0V to 1500V with no anomalies. Numerous power transistors and diodes have been tested.

Providing reliable operation under dielectric breakdown presented considerable design problems which were overcome. Characterization of the design was done using fast-acting spark gaps, frequently used as surge protectors, with 600V and 1000V nominal breakdown voltages. Such devices can be used thousands of times with little alteration in breakdown potential, so Reedholm has been able to perform extensive system characterization to complement the experience of users testing product wafers.

The high voltage testing option has been released as an option for use within the Reedholm application software. Thus, customers can set up tests using a data base menu and gather data using a variety of automatic probers. Testing that might be done only in development or under crisis mode can now be routinely done as a process control monitor.

At this time, several users are automatically gathering data at voltages up to 700V for first to second metal and from poly to metal. They have characterized their probe cards as capable of handling 1200V even at 4mil spacing between probe tips.

## References


1) S. Garrard, “Wafer Level Reliability Intermetal Dielectric Integrity,” 1990 International Wafer Level Reliability Workshop, p. 165

2) A.S. Grove, Physics and Technology of Semiconductor Devices, John Wiley & Sons, Inc., p. 194 (1967)

3) Agaroff & Brophy, Electronic Processes in Materials, McGraw Hill, p. 368 (1963)

4) A.J. Franklin, “Wafer Level J-Ramp Analysis of Thick Gate-Field Oxides, 1991 International Wafer Level Reliability Workshop,” p. 29

5) Ibid., p. 27

6) K. Boyko & J. Klema, “Summary—Insulator Reliability,” 1990 International Wafer Level Reliability Workshop, p. 251

7) Y. Nissan-Cohen et al., “Measurement of Fowler Nordheim Tunneling Currents in MOS Structures Under Charge Trapping Conditions,” Solid State Electronics, Vol. 28, No. 7, 1985, p. 720

8) Ibid., p. 27

9) J.S. Suehle & M. Gaitan, “Application of CMOS-Compatible Micro-Hotplates for In-situ Process Monitors,” 1992 International Wafer Level Reliability Workshop, p. 122