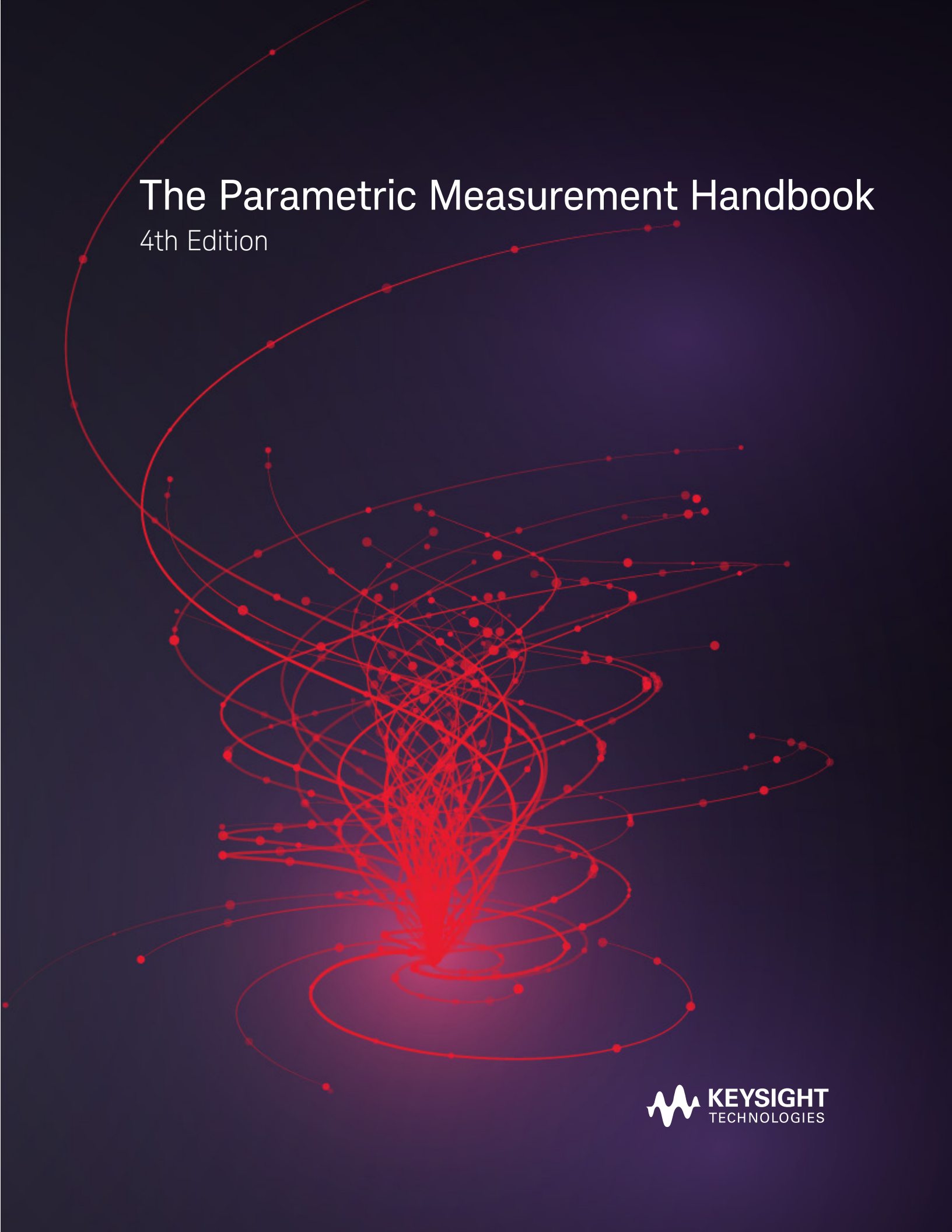
The Parametric Measurement Handbook

4th Edition



KEYSIGHT TECHNOLOGIES

Table of Contents

Table of Contents *(continued)*

Table of Contents *(continued)*

Table of Contents *(continued)*

Preface

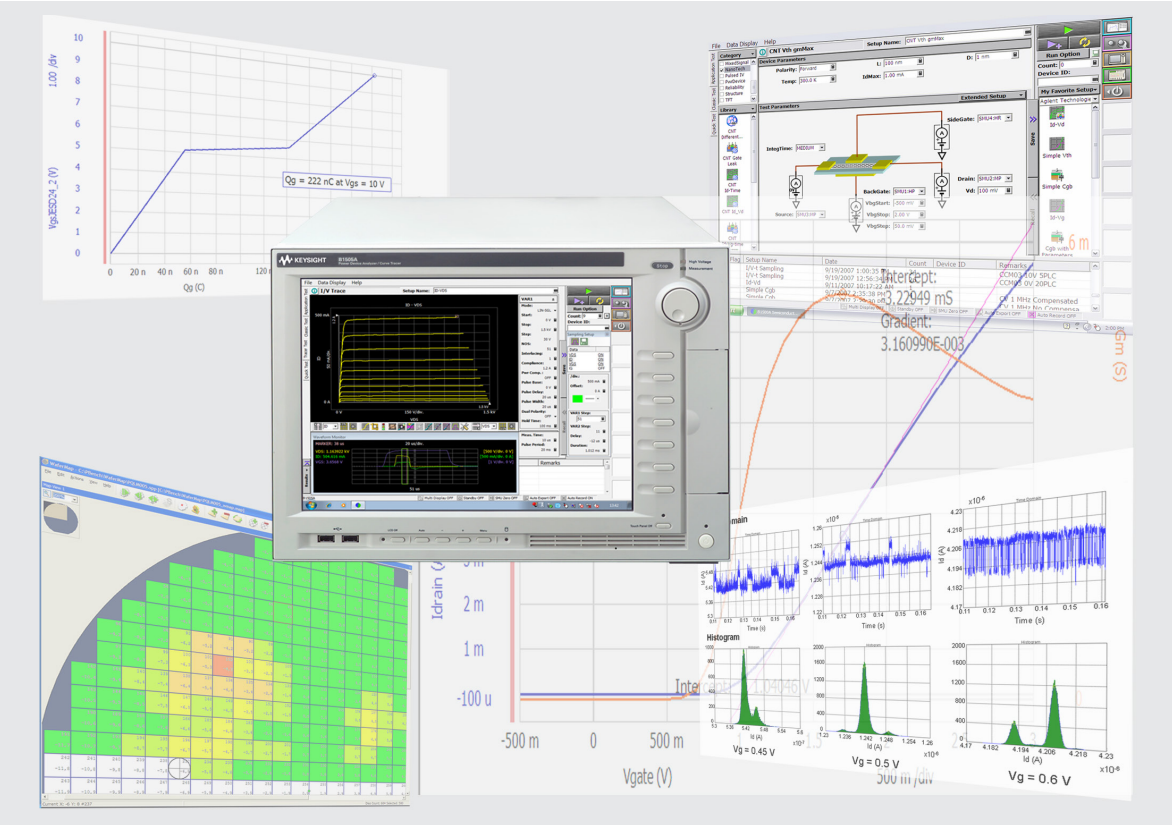
"An expert is a man who has made all the mistakes, which can be made, in a very narrow field" – Niels Bohr

It has been seven years since we published the first edition of this handbook, and in the interim many things have changed. Fully recovered from the shock (in 1999) of transitioning from Hewlett-Packard to Agilent Technologies, in 2014 we once again changed names as the electronic test and measurement portion of Agilent became Keysight Technologies ("unlocking measurement insights"). Despite our two name changes, Keysight remains true to its roots as HP's original business. We continue to work at supplying our customers with the best electronic test solutions available.

If you have an earlier version of this handbook, then you probably want to know: What is new? The main thing you will notice about this edition versus earlier ones is the addition of any entire new chapter (chapter 9) devoted to power device test. This was necessary as Keysight has come out with many new solutions to test power devices since the handbook was first published. Also, power device testing is a complex enough task that it warrants having a chapter devoted exclusively to it. The other changes have been interwoven into existing chapters, and they include updated information on enhancements to existing products as well as information on some of our newer products.

Many people helped with the content and review of this updated handbook, but in particular I would like to acknowledge the contributions of Yasushi Okawa for his many excellent presentations and training materials on power device test. Without those chapter 9 would have been much more difficult to write. Also, thanks go to Helen Amato, Biow-Huei Sim and Gwen Soo for their extensive work to update the handbook layout and images. Overall I think that there is enough new content in this edition to enable engineers and researchers to make accurate parametric measurements on low and high power devices both now and for many years into the future.

Alan Wadsworth
January 2018



Chapter 1 Keysight Technologies' Parametric Measurement Solutions

"The central activity of engineering, as distinguished from science, is the design of new devices, processes and systems." – Myron Tribus

What is a parametric test?

The question as to what constitutes parametric testing is an interesting one and is possibly open to some debate. Nevertheless, in general parametric testing involves the electrical testing and characterization of four main types of semiconductor devices: resistors, diodes, transistors, and capacitors. This is not to say that parametric tests never involve the testing of other device types; however, the vast majority of parametric test structures can be classified into one of these categories or considered to be a combination of these categories.

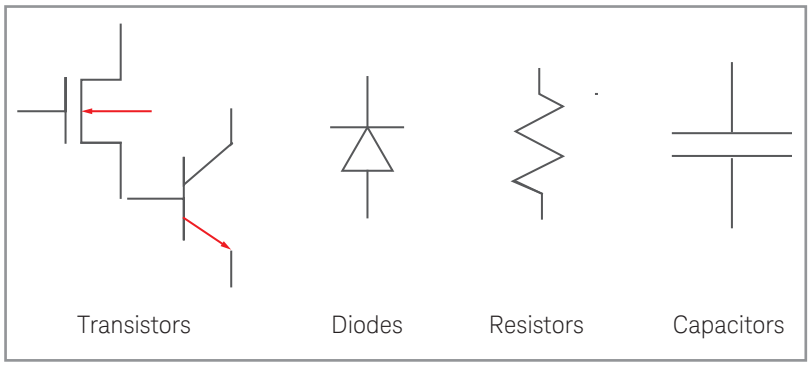


Figure 1.1. Parametric tests involve the testing of these four basic device types.

The vast majority of parametric testing involves either current versus voltage (IV) or capacitance versus voltage (CV) measurements.

To many people, parametric testing means "DC" testing, but this is not an accurate description. Of course, it can take source/monitor units (SMUs) anywhere from milliseconds to seconds to make a measurement, which is certainly "slow" by the standards of functional testers (which typically perform measurements in the nanosecond or picosecond range). However, in recent years, the need to perform extremely fast parametric measurements (1 µs spot measurements with data sampling rates in the nanosecond range) has greatly increased. This has required the creation of new measurement module types (such as the waveform generator/fast measurement unit or WGFMU) to meet this need. Extremely fast IV and pulsed IV measurements will continue to take on increased importance in the future, as transistor lithographies continue to shrink and more exotic materials are incorporated into semiconductor processes.

What is a parametric test? *(continued)*

One major subcategory of parametric testing is reliability testing. Reliability testing relies heavily upon the well-known Arrhenius equation that expresses the rate constant (k) of a chemical reaction as follows:

$$k = Ae^{-\frac{E_a}{RT}}$$

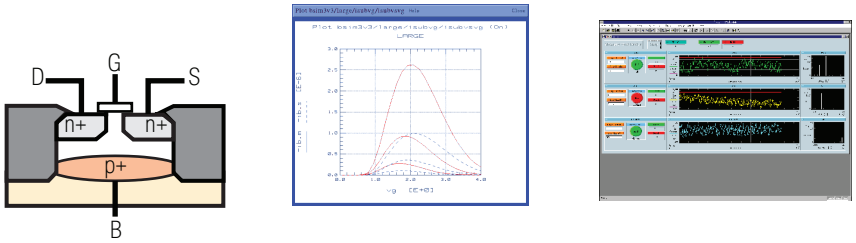
Where:

A is the pre-exponential factor
E_a is the activation energy
R is the ideal gas constant
T is the temperature in degree Kelvin

In reliability testing, devices are typically stressed by the application of large currents and/or voltages (larger than that experienced by the devices under normal operation) to lower the value of the activation energy and thereby increase the rate of the failure mechanism. Temperature is also often increased to achieve this same purpose. Once the failure mechanism occurs, then the expected failure rate under normal operating conditions can then be extrapolated using a variety of mathematical and statistical techniques.

Why is parametric testing performed?

The purpose of parametric testing is to determine the characteristics of a semiconductor manufacturing process. Broadly speaking, parametric testing covers three main areas: process development, device modeling, and process control.



Process Development | Device Modeling | Process Control

Figure 1.2. Parametric testing is widely used in three main areas: process development, device modeling, and process control.

Why is parametric testing performed? *(continued)*

The first two of these areas are performed in a laboratory, R&D or pre-production environment, while the last one is performed in a manufacturing environment. The parametric equipment used in these disparate environments obviously has different requirements, with benchtop instruments being used for process development and process modeling, and high-throughput testers being used for process production.

Figure 1.3. A production parametric tester is designed to optimize throughput

It is important to understand that parametric test is, almost never performed on final products. Instead it is performed on special structures that are designed to yield information about the process itself. Parametric tests are also generally performed directly on semiconductor wafers. In production test, the parametric test structures are sometimes located in the scribe lanes or "streets" of the wafer to minimize the wafer area taken up by these devices. However, for process development and reliability testing entire wafers of nothing but parametric test structures are often fabricated.

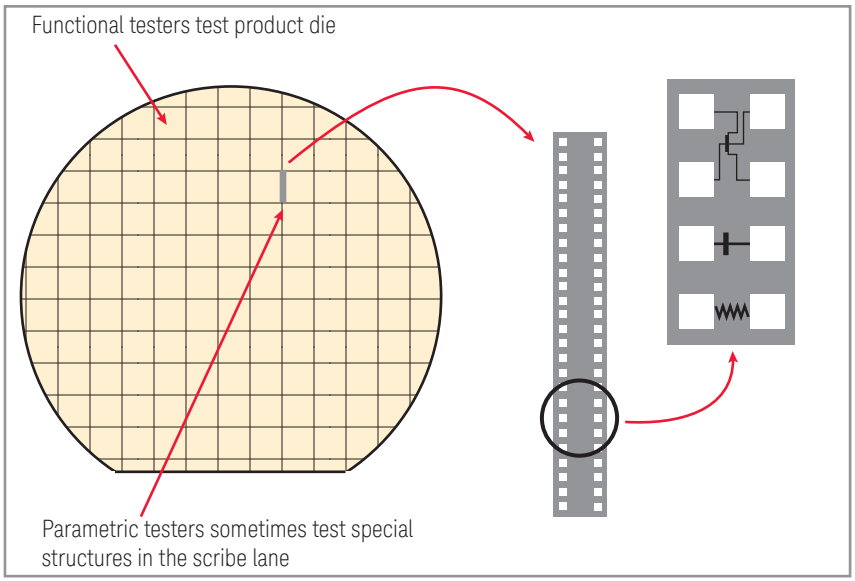


Figure 1.4. To conserve valuable wafer area, parametric test structures are sometimes placed in the wafer's scribe lines (or "streets").

Where is parametric testing done?

In production, parametric tests are typically performed on wafers after they have completed the wafer fabrication process (i.e. after passivation has been applied) but before electrical sort (E-sort) on the functional product dice.

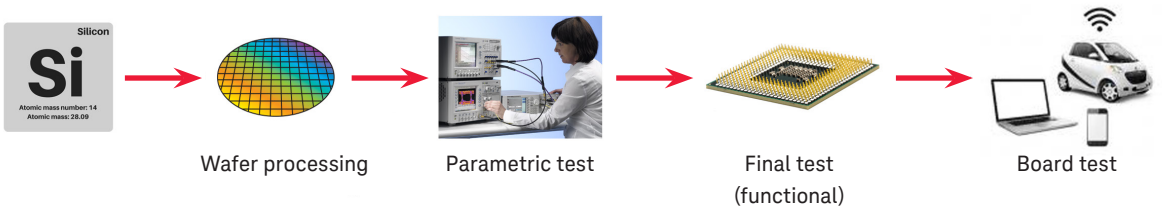


Figure 1.5. Parametric testing is performed after wafer fabrication is complete but before functional product verification.

Each wafer from every lot is tested and the data is stored into a database. As the amount of data is quite massive, various software tools are employed to manipulate the data into a variety of different formats. One popular format is the wafer map, where a scalar quantity is plotted across a wafer using different colors for different ranges of the data value.

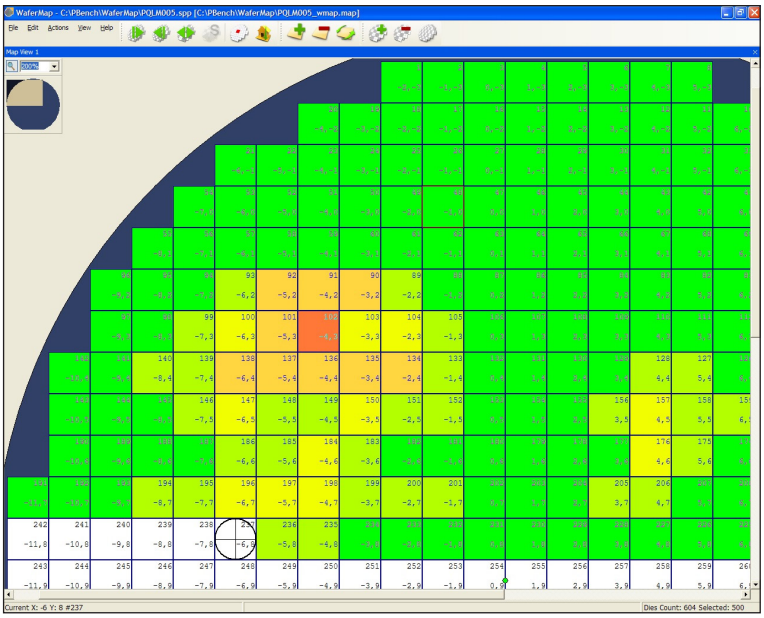


Figure 1.6. Wafer map example.

Where is parametric testing done? *(continued)*

For advanced processes, conventional test structures placed in the scribe lines or even drop-in test die placed around the wafer may not be sufficient to adequately characterize the process. Advanced processes typically require much more testing due to their innate complexity, and it is sometimes difficult to fit all of the necessary test structures into the available area. Unfortunately, the tips of the probe card have physical limitations on how small they can be made and still maintain reasonable probe card lifetimes, which in turn creates physical limits on the minimum size of the probe pads. This means that the pads cannot be physically scaled down as devices scale down with each new process generation.

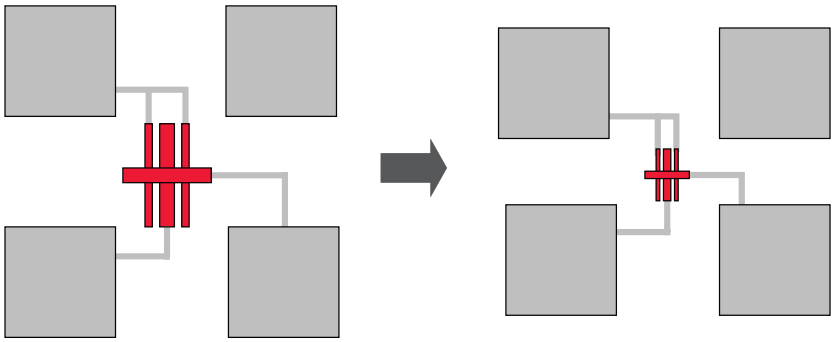


Figure 1.7. Pad sizes cannot be scaled down with changing design rules, which limit the number of conventional test structures that can be placed on a wafer.

One solution to this issue is to use arrays, since an array allows test devices to share pads and thereby improve the test device to pad ratio. An example of this scheme is shown below.

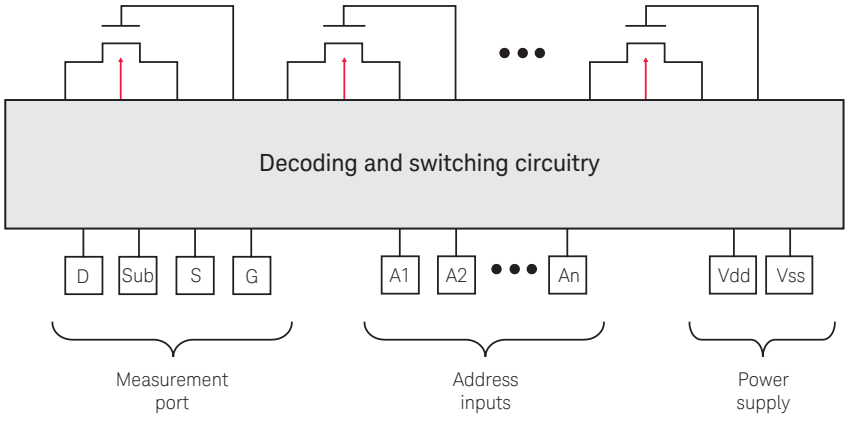


Figure 1.8. An example of an addressable array scheme used for production parametric testing.

Addressable arrays can offer significant throughput improvement over conventional test device arrangements, although achieving substantial test time reduction requires a complete re-engineering of all of the parametric test processes.

Parametric instrument history

The first instruments that could perform anything resembling parametric measurement were analog curve tracers. However, these instruments had a number of limitations, not the least of which was that the “data” was a display on a CRT screen. The only way to save this data was to take a Polaroid picture of the display, which still did not supply you with actual numeric information.

The 1980’s

In the early 1980’s, Keysight Technologies (then part of Hewlett Packard) introduced the world’s first digital parameter analyzer, the 4145A. The 4145A was the first instrument to incorporate four source/monitor units (to be discussed further in chapter 3) in a single box along with the software necessary to integrate all of these resources. The 4145A produced plots similar to that of a curve tracer, but its output consisted of discrete points of digital data that could be transferred into other software for storage and analysis. This product was something revolutionary for the semiconductor industry, and the 4145A quickly supplanted the curve tracer except for a few specialized applications. Later in the 1980’s, an improved version of this product, the 4145B, was also introduced.

Figure 1.9. The 4145B Semiconductor Parameter Analyzer

The 1990’s

In the early 1990’s, Keysight Technologies (then part of Hewlett Packard) introduced the 4155 and 4156 semiconductor parameter analyzers. These instruments built upon the 4145A/B by adding new capabilities including a color display and keyboard, automatic data analysis capability, a pulsed sweep capability, a thinned-out sampling mode, DC and AC stressing and a standby mode, just to name a few. There were several versions of these products (the “A”, “B” and “C” models), and with each subsequent version more features and capabilities were added. In particular, quasi-static capacitance versus voltage (QSCV) measurement capability was added to the "C" version along with vastly improved voltage monitor unit (VMU) measurement capability.

Figure 1.10. The 4156C Precision Semiconductor Parameter Analyzer

The 21st century

At the beginning of the 21st century, it became increasingly clear that parametric testing was becoming more complicated and that a complete solution needed to be able to perform more than just IV measurements. A solution that could perform both CV and IV measurements was required, and this solution also had to have the flexibility to add additional measurement resources in the future as testing needs evolved. To meet these challenges, Keysight Technologies (then part of Agilent Technologies) introduced the B1500A semiconductor device analyzer with Keysight EasyEXPERT software in 2005. The B1500A supports all aspects of parametric testing, from basic manual measurements to test automation across a wafer in conjunction with a semiautomatic wafer prober. Because the B1500A utilizes the Microsoft® Windows® operating system, it integrates easily into PC-based work environments. In addition, the familiar Windows graphical user interface (GUI) and convenient online help menus minimize the need for instrument training. The B1500A has already supplanted the 4155C and 4156C as the tool of choice for state-of-the-art parametric measurement needs.

Figure 1.11. The B1500A Semiconductor Device Analyzer

The B1500A is a modular instrument, and it supports a variety of module types. Modularity is an extremely important feature since parametric testing continues to be an extremely dynamic area where new types of testing are constantly being required. Parametric testing requires more than simple source/monitor unit resources, and increasingly complex tests requiring capacitance testing, high-speed testing and fast pulsed measurement are becoming common. Modularity ensures that measurement equipment does not become outdated and that you have the flexibility of adding new test capabilities to your parametric measurement equipment as your measurement needs change in the future.

The 21st century *(continued)*

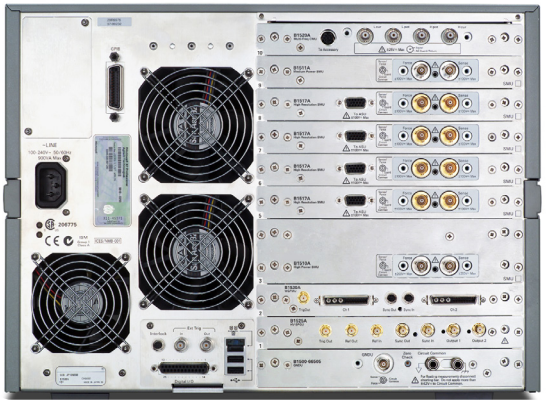


Figure 1.12. The ten-slot B1500A supports a variety of types of source/monitor units as well as a capacitance measurement module, a high-voltage pulse generator module and a waveform generator/fast measurement unit module.

In addition to the B1500A, Keysight introduced the B2900A family of precision benchtop SMUs. The B2900A SMU family has an integrated GUI that supports measurement and graphing right from the front panel, and it is available in both one and two channel versions.

	B2901A	B2902A	B2911A	B2912A
Number of Channels	1	2	1	2
Maximum V	±210 V			
Maximum I (DC/Pulsed)	±3 A (DC) / ±10.5 A (Pulsed)			
Minimum Pulse Width	50 µs			
Measurement Resolution (I)	100 fA		10 fA	
Measurement Resolution (V)	100 nV		100 nV	
Minimum Sampling Interval	20 µs		10 µs	

Figure 1.13. Table showing key measurement capabilities of the B2900A benchtop SMU series.

The B2900A SMUs are supported in a variety of different software platforms (including BenchVue), but for parametric test, the key feature is that they are supported in EasyEXPERT group+. In fact, Keysight can supply a cable (N1294A-032 Digital I/O Trigger Cable for Multiple Unit Control) that allows you to control up to four two-channel B2900A SMUs (eight SMU channels) and maintain complete synchronization between all eight channels. If you require only basic IV measurements functions, it is very cost effective to use one or more B2900A SMUs with EasyEXPERT group+. Also, the B2900A SMUs can supply more current and voltage than B1500A SMUs can. Thus they can provide a good intermediate solution between the B1500A and the B1505A and B1506A (which will be discussed next).

Figure 1.14. B2902A two-channel SMU being controlled with EasyEXPERT group+ software using an external PC.

To meet the needs of the expanding high-power market and to supply an alternative to conventional analog curve tracers (which are no longer being made by major manufacturers), Keysight introduced the B1505A Power Device Analyzer/Curve Tracer in 2009. The B1505A can supply up to ±1500 A and ±10 kV, and it can perform 10 µs pulsed measurements. In addition, the B1505A comes with a tracer test mode that supports a knob sweep capability and an oscilloscope view that enables you to verify current and voltage pulses. The B1505A also supports automated CV measurements at up to 3 kV of DC bias, and gate charge measurements at up to 1100 A and 3 kV.

Figure 1.15. The B1505A power device analyzer/curve tracer with its ultra-high current and ultra-high voltage modules.

Unlike analog curve tracers, the B1505A is PC-based and uses the same user-friendly Keysight EasyEXPERT software as the B1500A.

Several new features in the B1505A's tracer test mode provide dramatic usability improvements over traditional curve tracers. A snapshot feature allows you to save and display multiple data traces so that you can easily compare them with data from the current measurement. A stoplight feature allows you to graphically define forbidden regions (either voltage or current based) such that the measurement immediately ceases if the trace enters the forbidden area. Best of all, an auto-record feature keeps a running record of the most recent trace changes so that you can replay and save measurement trace data even if the device is inadvertently damaged or destroyed. Taken together, these improvements represent a truly revolutionary advancement in curve tracer design that can significantly reduce device characterization cycle times.

A sister product to the B1505A is the B1506A Power Device Analyzer for Circuit Design. The B1506A has many of the same capabilities as does the B1505A, but it is focused on the needs of power semiconductor device users and circuit designers rather than on the needs of power device manufacturers. The B1506A can supply up to ±1500 A and ±3 kV, and it can make 10 µs pulsed measurements. Just like the B1505A, the B1506A supports automated CV measurements at up to 3 kV of DC bias and gate charge measurements at up to 1100 A and 3 kV. However, the B1506A does not support on-wafer CV or gate charge measurements, the 10 kV expander unit, or the GaN current collapse switch.

Figure 1.16. The B1506A power device analyzer for circuit design.

The B1506A also has its own unique software interface, Easy Test Navigator (although it also supports EasyEXPERT group+). For engineers who are not test experts and whose only goal is to create a device data sheet, Easy Test Navigator provides an intuitive interface that can automatically generate a complete device data sheet.

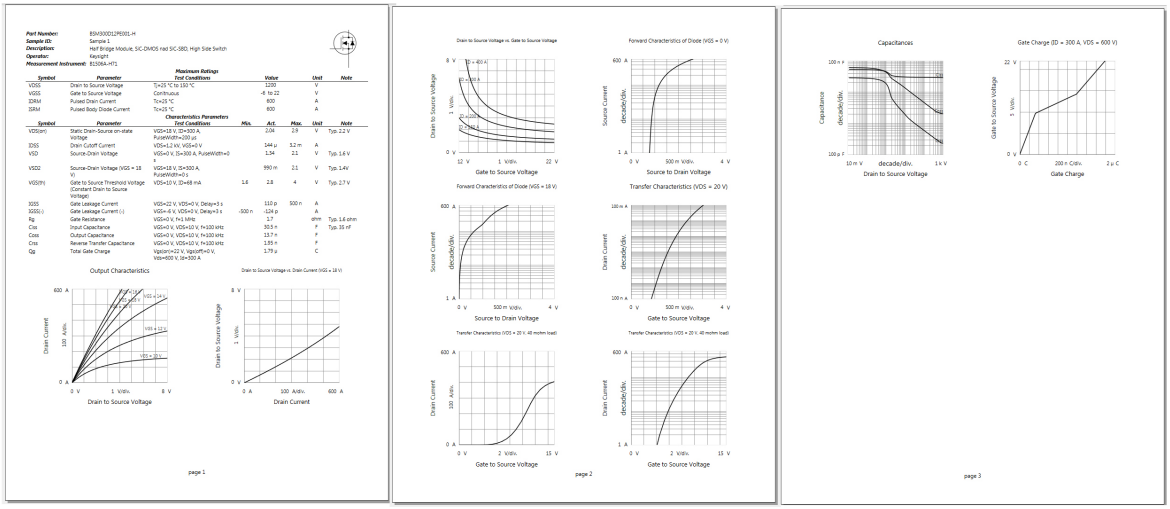


Figure 1.17. Easy Test Navigator can automatically generate a customized datasheet.

Power device performance is inextricably tied-in to package thermal performance, so the ability to measure device operation across different temperatures is extremely important. For this reason, both the B1505A and B1506A support packaged device thermal testing. If you only need to characterize device performance at room and hot temperature, then a thermal plate option is available (sold by inTEST Thermal Solutions). The thermal plate temperature is controllable via both EasyEXPERT group+ and Easy Test Navigator software, and the N1265A and B1506A test fixtures come with two k-type thermocouple inputs so that both the ambient and DUT temperature can be monitored at the same time.

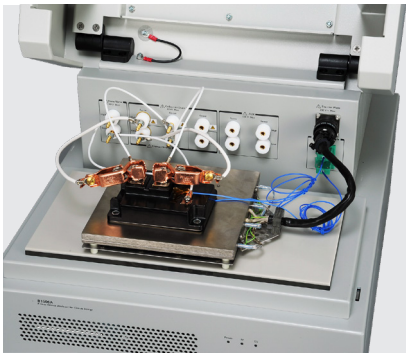


Figure 1.18. The B1505A/B1506A thermal plate can support temperature measurements from room temperature up to 250°C.

If both cold and hot temperature measurements are necessary, then Keysight can supply thermal test adapters for both the N1265A and B1506A test fixtures that support the use of inTEST temperature forcing systems.

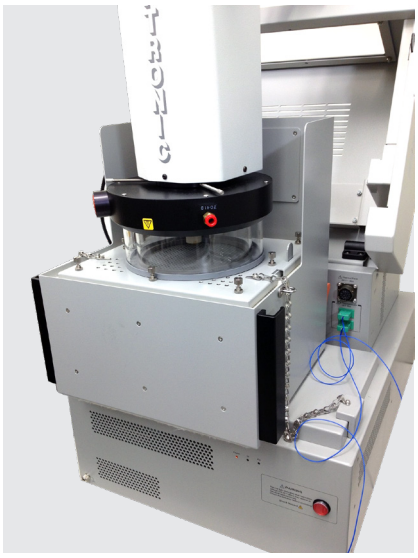


Figure 1.19. The B1505A/B1506A can support the use of inTEST temperature forcing systems for temperature measurements from -50°C to 250°C.

Both EasyEXPERT and Easy Test Navigator software can support temperature triggered measurements (using measured data from the thermocouple inputs). This allows for the automated measurement of device parameters across temperature.

Chapter 2 Parametric Measurement Basics

"I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it ..."

"But when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be."

— William Thomson (Lord Kelvin)

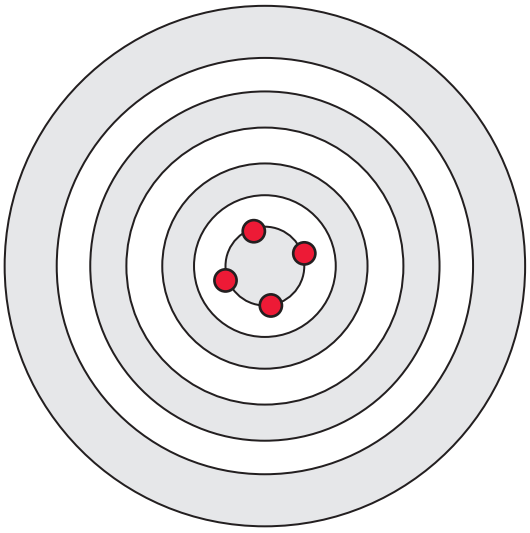
Measurement terminology

Before you can begin to collect measurement data, you need to know some basic measurement terminology. In particular, understanding the precise meaning of the terms accuracy, repeatability and resolution is essential to understanding parametric test and the measurement capabilities of a parametric measurement resource.

Accuracy and repeatability

Accuracy and repeatability are related but different. Accuracy is the degree of conformity of a measured or calculated quantity to its actual (true) value. Repeatability (also known as precision) is the degree to which repeated measurements or calculations show the same or similar results. Accuracy and repeatability do not have any innate correlation. A measurement can have high accuracy and high repeatability, high accuracy and low repeatability, low accuracy and high repeatability, or both low accuracy and low repeatability.

The visual analogy that is often used to explain these terms is a bull's-eye target. A group of points that are close to the center of a target but spaced far apart from one another (as shown below) have high accuracy but low repeatability.

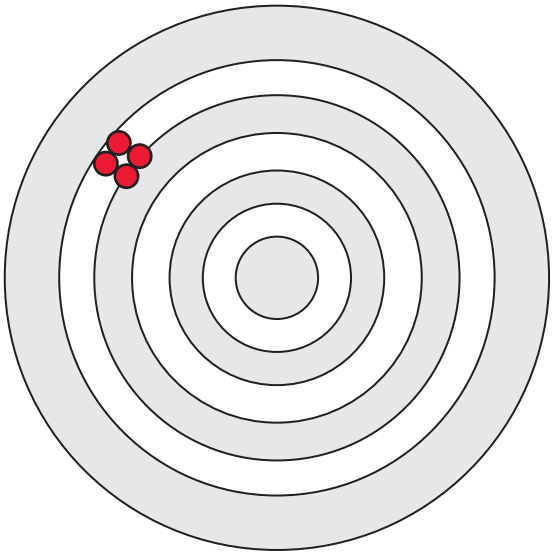


High accuracy, low repeatability

Figure 2.1. High accuracy and low repeatability.

Accuracy and repeatability *(continued)*

A group of points that are far from the center of a target but spaced close together (as shown below) have low accuracy but high repeatability.



Low accuracy, high repeatability

Figure 2.2. Low accuracy and high repeatability.

The analogies for the two other cases should be fairly obvious. For production parametric test, repeatability is usually more important than absolute accuracy. This is because parametric testing is concerned with monitoring trends, which makes having highly repeatable measurements extremely important.

Resolution

In a parametric test, resolution is what allows us to gauge accuracy and repeatability. In this sense, it is the most important specification for parametric testing. Resolution is the smallest change in data that an instrument can display. The easiest way to think about this is in terms of an analog-to-digital converter (ADC) circuit. The figure below shows a simplified version of an ADC:

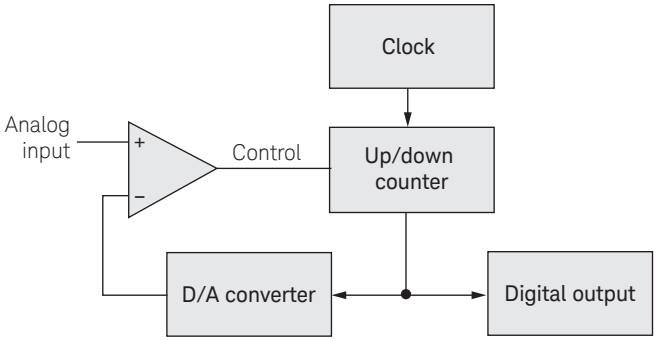


Figure 2.3. Simplified analog-to-digital converter (ADC) circuit.

Resolution *(continued)*

As this illustrates, the number of bits available to the digital-to-analog converter (DAC) will determine the fineness of the measurement detail that can be distinguished, which is the resolution. For example, a resolution of 20 bits represents the ability to distinguish one part in 2^{20} or 1,048,576 (basically one part in a million).

To express the measurement resolution in terms of current or voltage, you need to know the measurement range in which you are operating. For example, in a 5 V measurement range, the ADC measures from –5 V to +5 V, which (assuming a 20-bit ADC) yields a readable or resolvable measurement resolution of:

$$\frac{(5+5)\text{V}}{1{,}048{,}576} \approx 10\ \mu\text{V}$$

However, due to noise and other factors, the readable resolution is not the same as the resolution specified in a data sheet. The resolution specified in a data sheet takes into account factors such as thermal noise and amplifier offset and noise, which introduce a stochastic element into the data sheet specification. Typically, the specified resolution is 1 to 2 orders of magnitude larger than the readable resolution.

Traceability

Traceability is defined as the property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons, all having stated uncertainties. It is important for any company that is manufacturing measurement equipment to establish traceability up to national and international standards, in order to guarantee that measurements are being made correctly.

There are a number of National Metrology Institutes (NMIs) that provide standard reference materials of the highest quality and metrological value. The intent of all NMIs is to realize SI units (such as volts, amperes, ohms, etc.) as closely as possible. A partial list of NMIs is shown below.

Name	Acronym	Country
National Institute of Standards and Technology	NIST	USA
National Metrology Institute of Japan	NMIJ	Japan
National Physical Laboratory	NPL	UK
Physikalisch Technische Bundesanstalt	PTB	Germany

Keysight Technologies obtains standards from these NMIs to use as calibration references. Keysight's measurement products include a certificate of calibration when they are originally shipped from the factory or after a product calibration has been performed. This document is in compliance with the traceability requirements from these NMIs.

Shielding and guarding

Beginners sometimes confuse shielding with guarding, although they are quite different and have totally different purposes. The purpose of shielding is to prevent electrostatic noise from interfering with a measurement, while the purpose of guarding is to prevent leakage of currents and enable low-current measurements (in the femtoamp range). A shield does not act as a guard, and a guard does not necessarily provide good shielding. As a general rule of thumb, when making measurements below 1 nanoamp (10^{-9} A) you should use guarding; when making measurements below 1 picoamp (10^{-12} A) you should use both guarding and shielding.

Shielding: Maintaining a low noise floor

Shielding involves surrounding the measurement environment with conductive material so as to eliminate electrostatic noise. Basic electromagnetic field theory shows that surrounding a volume with conductive material creates a barrier across which electromagnetic signals cannot cross. This is also often referred to as creating a "Faraday cage." Typically, for parametric tests, this is accomplished using some sort of shielding box as shown below:

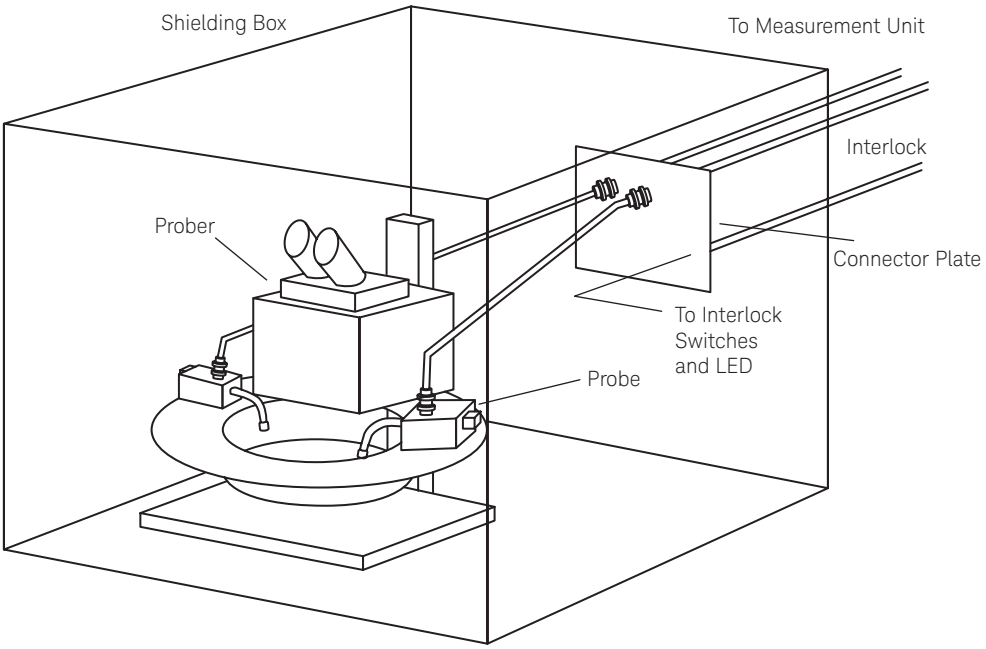


Figure 2.4. A well-shielded parametric measurement environment.

There are a number of best practices that one should keep in mind when trying to create a well-shielded measurement environment:

- Enclose the measurement area with conductive material and measurement cables with a metal shield or braid.
- Electrically connect shielding enclosure and cable shields to the test instrument common (shield) and/or to earth ground.
- To reduce capacitance and vibration effects, keep the shielding and test circuit as far apart as possible.
- Eliminate "light leaks" through gaps around doors and hinges, tubing and cable entry points, connectors and connector panels, and seams/joints between panels.
- Fill seams/gaps/joints with conductive caulking or gaskets.

Shielding: Maintaining a low noise floor *(continued)*

A reasonable question to ask is: Since a proper shield completely surrounds the measurement environment, how do you connect anything to it? The solution is to use a connector plate that is specifically designed for this purpose.

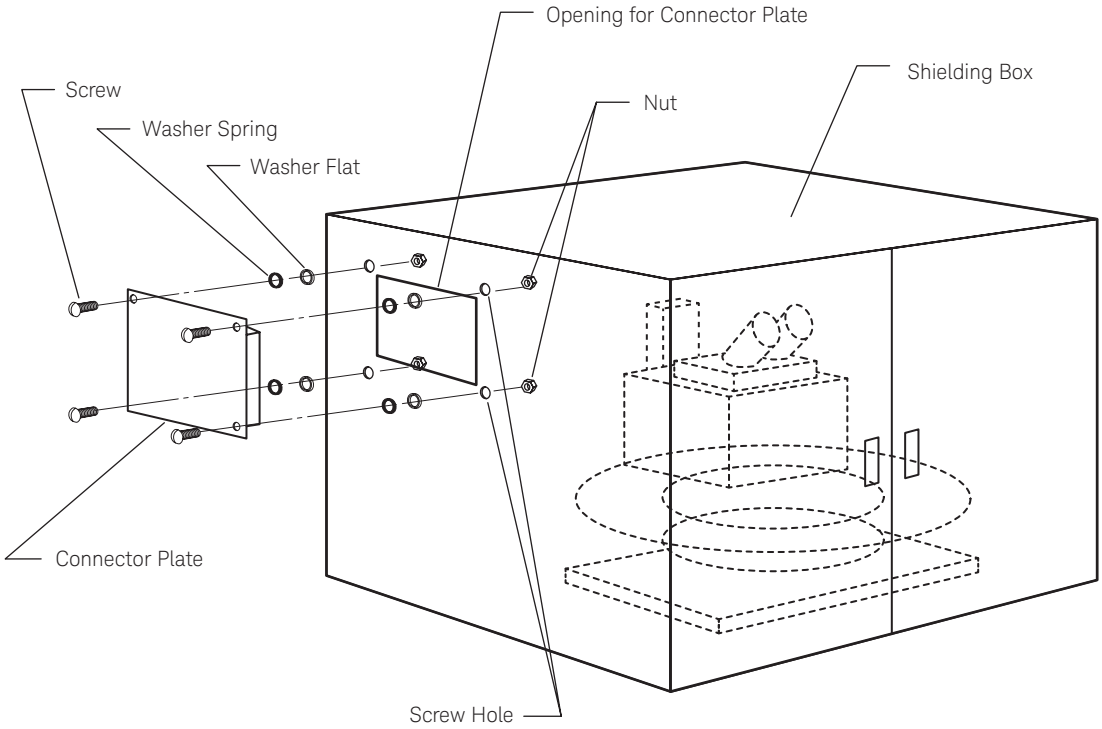


Figure 2.5. Using a connector plate to connect through a shielding box.

Keysight Technologies can supply a variety of connector plates with both triaxial and BNC connectors specifically for this purpose.

In summary, the following are the key points to keep in mind when constructing a well-shielded measurement environment:

- Keep all charged objects (including people) and conductors away from the measurement area.
- Use highly conductive materials instead of insulating materials near the test circuit.
- Avoid movement and vibration near the measurement area (air currents can cause movement and/or vibration).
- When measuring currents <1 pA, shield the measurement area with a conductive (metal) enclosure and connect the enclosure to the test instrument common (shield) and/or to earth ground.
- Minimize the capacitance between the shielding enclosure and the test circuit.

Avoiding ground loops

One common problem when using shields that has not yet been mentioned is creating ground loops. If all shields are tied to ground and you have multiple shielded instruments and cables in a given measurement environment, then it is almost a certainty that the grounds to which these are tied are not at the same potential. Consider the simple case of two conductive planes both tied to ground:

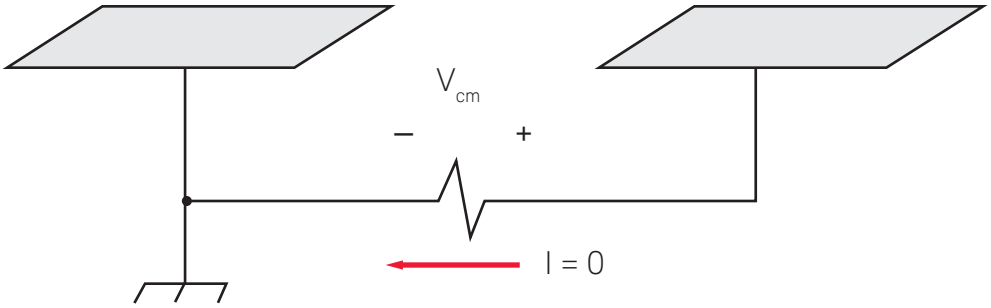


Figure 2.6. Conductive planes connected at only one point cannot have any current flow between them.

As long as these conductive planes are only connected together at one point, any difference in the common mode voltage (V_{cm}) of these planes is not an issue because no current can flow between the planes. However, if the ground planes are strapped together in more than one location then a loop path for current is created that can cause significant current to flow through the conductive planes. This common mode voltage can seriously impact parametric measurement results:

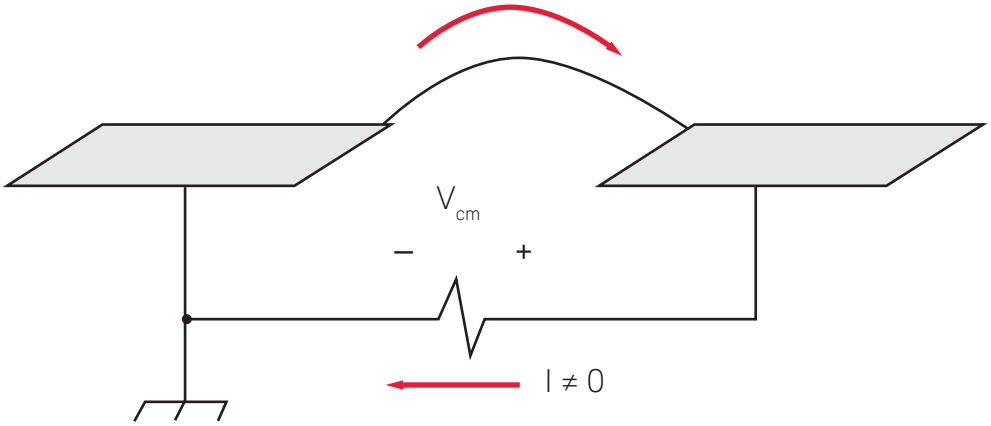


Figure 2.7. Conductive planes connected at multiple points creates a loop for current (a condition to be avoided).

Thus, although it may be counter-intuitive, the best strategy when making parametric measurements is to avoid excessive ground connections and instead connect everything together through a single ground point.

Note: Since instruments almost always have their ground tied to chassis ground, which is in turn tied to earth ground, this is already taken care of for you. If you suspect that you are having a grounding issue, then the simplest solution is to make sure that the power cords of all of your equipment are tied together to a common power strip.

Guarding: Isolation for low-current measurement

Guarding involves surrounding a signal line with an actively driven conductor maintained at the same voltage potential as that of the signal to eliminate leakage currents. Good guarding can only be achieved using triaxial connectors and cabling. In a triaxial cable, the signal line is surrounded by the guard line (separated of course by insulating material), which is in turn surrounded by a grounded shield line (also separated by insulating material). The following illustration shows a cut-away view of a triaxial cable.

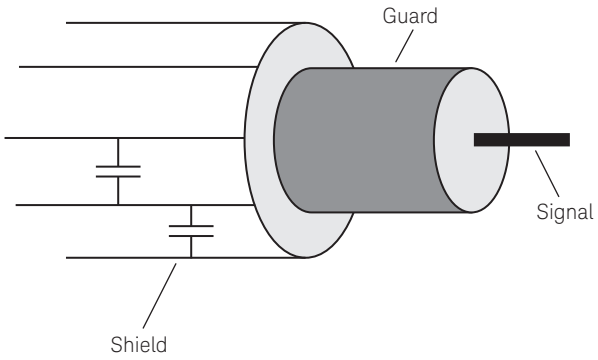


Figure 2.8. Cut-away view of a triaxial cable, showing parasitic capacitance between shield to guard and guard to signal.

Even though there is parasitic capacitance in this cable, the driven guard effectively eliminates the effects of this parasitic capacitance by isolating the signal line from the shield. The diagram below shows the structure of the circuitry driving the guarded (triaxial) output.

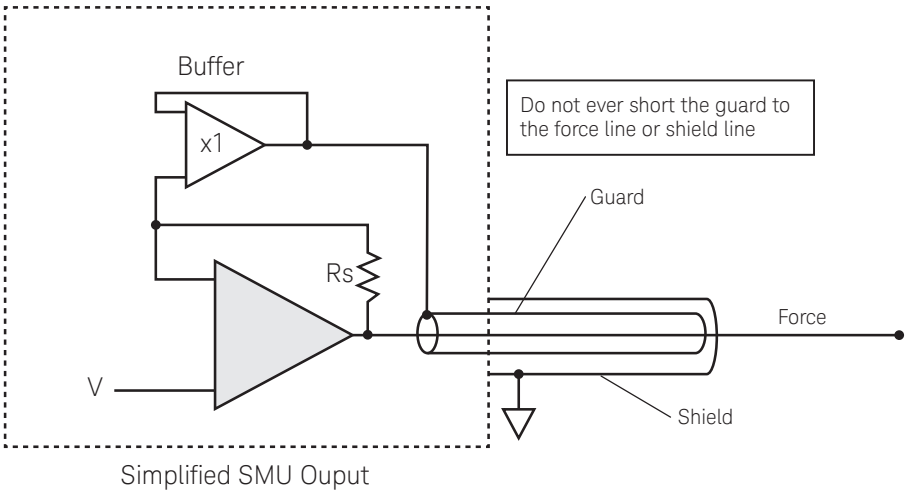


Figure 2.9. Circuitry to provide a guarded (triaxial) output.

The guard line is actively driven by a buffer circuit to maintain the same voltage potential as the center conductor (force) line. Obviously, if you short the guard line to either the force or shield lines then you risk damaging the output circuitry.

Guarding: Isolation for low-current measurement *(continued)*

The inevitable question gets asked: Why do we need triaxial cables to measure low currents? After all, BNC cables use insulating material, so why should we expect any leakage current? Some simple calculations help to illustrate the need for triaxial cables when measuring ultra-low currents. Assume that the insulation material has a resistance of 1 GΩ and that we apply 100 V to the center conductor:

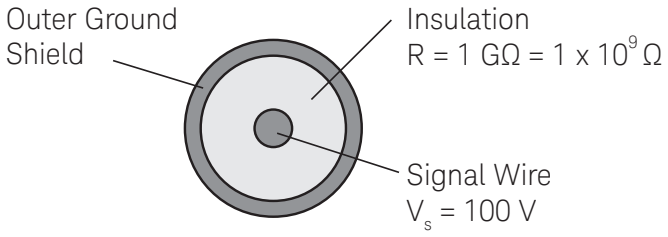


Figure 2.10. Leakage current when using a BNC (coaxial) cable.

A simple calculation shows that the leakage current will be:

$$i_{leak} = \frac{100\ V}{1\times10^{9}\ \Omega} = 1\times10^{-7}\ A = 100\ nA$$

Obviously, if we are trying to measure currents in the femtoamp (10^{-15} A) current range, we cannot use cables with 100 nA of leakage current.

Now let us examine the same case using a triaxial cable. In this case we assume that the driven guard can track the applied (signal) voltage to within one part in a million (which is a reasonable and conservative assumption):

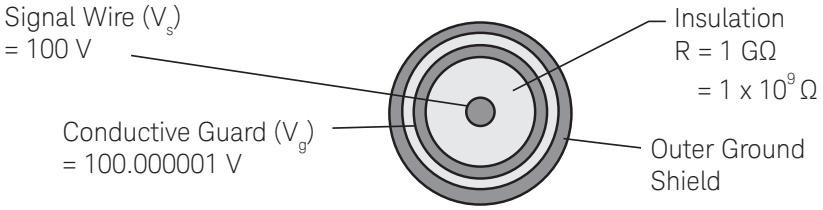


Figure 2.11. Leakage current when using a triaxial cable.

Thus, the leakage current will be the difference between the voltages of the signal and guard divided by the effective resistance:

$$i_{leak} = \frac{(100{,}000001\ V - 100\ V)}{1\times10^{9}\ \Omega} = \frac{1\times10^{-6}\ V}{1\times10^{9}\ \Omega} = 1\times10^{-15}\ A = 1\ fA$$

By using triaxial cables with a driven guard, we have reduced the leakage current to 1 fA, or by a factor of one-hundred million (10^{8}) as compared to using a BNC cable. It is for this reason that all low-current parametric measurements are made using triaxial cables.

Connecting triaxial and BNC cables

Inevitably, you will be faced with the issue of connecting BNC and triaxial cables. The three key questions that require answers are:

1. How do I connect up the driven guard?
2. How does this affect my low-current measurement capability?
3. Where do I get the necessary connectors?

We will address each of these concerns in turn.

The most important point to understand when connecting up triaxial and BNC connectors is how to connect the driven guard. Remember that the driven guard exists to allow for sub-nanoamp measurements. We can state as a general rule of thumb that for current measurements above one nanoamp, we can simply connect the force/sense line to the center conductor of the BNC cable and the shield to the outer ground shield of the BNC cable as shown below.

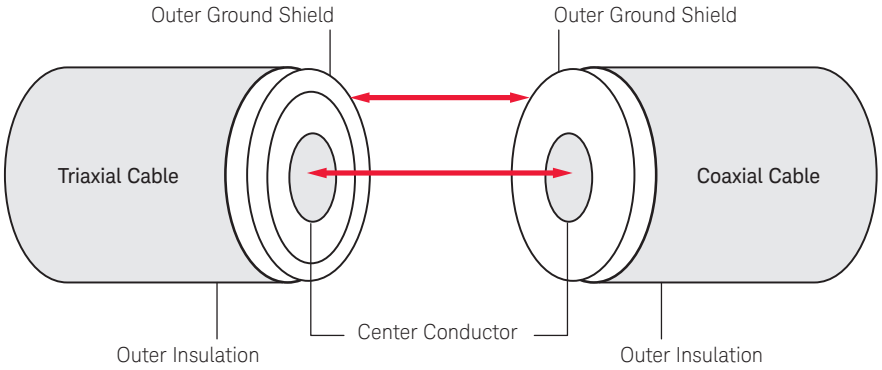


Figure 2.12. Triaxial to BNC connection when not making low-current measurements (1 nA and above).

The advantage of this configuration is its simplicity. We do not need to take any special precautions with the driven guard.

However, the only way that we can make a low-current (below one nanoamp) measurement with a BNC cable is to connect the driven guard to the outer shield of the BNC cable as shown below.

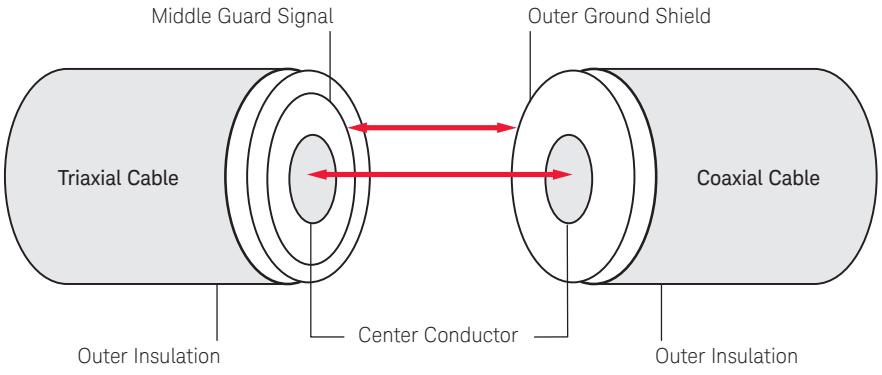


Figure 2.13. Triaxial to BNC connection when making low-current measurements.

Connecting triaxial and BNC cables *(continued)*

This configuration presents a couple of significant challenges:

1. The outer shield of the BNC must be isolated from ground.
2. The outer shield of the BNC can reach dangerous voltages (>40 V), therefore precautions must be taken to ensure that the user is protected from receiving a fatal electrical shock.

There are analytical wafer prober companies with ready-made solutions to isolate the BNC shield from ground to prevent these types of issues from occurring.

You must be very careful when selecting triaxial to BNC adapters to make sure that you understand which of the above two schemes the adapter is using. If you are uncertain about your connector, then it is strongly advised that you use a multi-meter to check the connections. You can purchase triaxial to BNC connectors from Keysight as well as a variety of third-party suppliers.

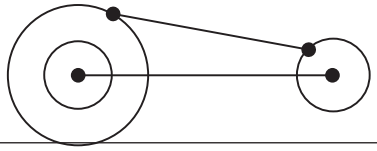
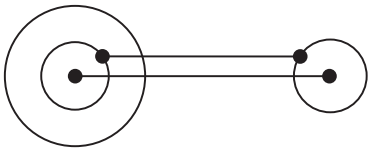
Part number	Description		
1250-2652	Triaxial (F) -BNC(M)		Safe. Not suitable for low-current measurements
1250-2653	Triaxial (M) - BNC(F)		
1250-2650	Triaxial (M) - BNC(F)		WARNING! Shock Hazard! Required for low-current measurements.
1250-2651	Triaxial (F) - BNC(M)		
1250-1830	Triaxial (F) - BNC(F)		

Figure 2.14. Table showing the different types of triaxial to BNC connectors supplied by Keysight Technologies.

Kelvin (4-wire) measurements

No wire or cable is a perfect conductor. Everything that you use to connect your instruments to the device under test (DUT) has some innate resistance (even though the resistance may be very small). The important question is: When do I need to be concerned about this cable resistance? The corollary to this question is: How do I eliminate the cable resistance effects? In this section, we will answer both of these questions.

Every cable can be thought of as a resistor. The resistance of a cable or wire typically used in parametric tests is very small (on the order of milliohms). Obviously, if you are trying to measure a test structure with a resistance that is many orders of magnitude greater than 0.1 Ω, you can safely ignore the cable resistance effects. However, suppose you are trying to measure a structure with only a few Ohms of resistance. In this case, the error introduced by the cable resistance can significantly skew your measurement results.

In order to remove the effects of the cable resistance, we need to eliminate the resistive voltage drop caused by the current flow through the cable and only measure the resistive effects associated with the test structure. We can do this by making what is known as a 4-wire measurement. This technique requires two separate lines for each terminal on the structure that we want to measure. One pair of lines is used to force current to the DUT, and the other pair of lines is used to sense the voltage measurement. The key point of this measurement is that we separate out the lines used to supply current to the DUT from the lines used to measure the voltage drop across the device. Since the sense lines making the voltage measurement are not conducting any current, there is no voltage drop due to cable resistance. Therefore, the cable resistance effects are eliminated. This technique is often described as a Kelvin measurement, after the famous British mathematician and physicist Lord Kelvin who invented the technique (and who is quoted at the beginning of this chapter).

Kelvin (4-wire) measurements *(continued)*

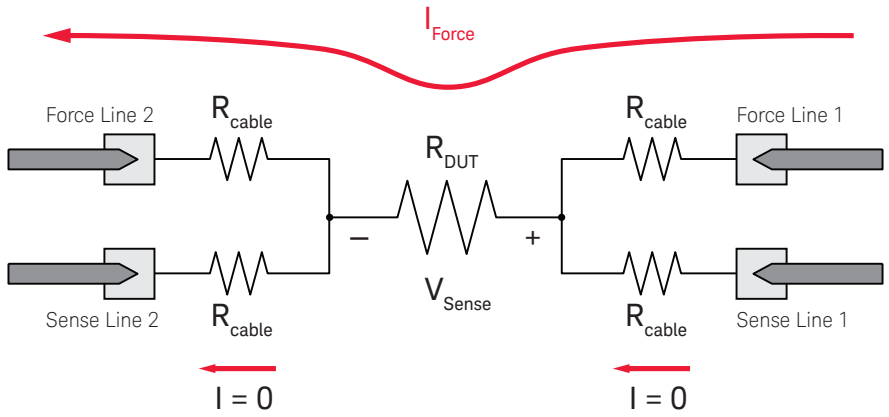


Figure 2.15. A Kelvin (or 4-wire) voltage measurement eliminates the effects of cable resistance by separating out the lines carrying the force current from the lines sensing the voltage.

To be effective, a Kelvin measurement requires that the force and sense lines be shorted together as close to the DUT as possible. Many analytical wafer probing companies make Kelvin triaxial positioners that short the force and sense lines together right at the test pad.

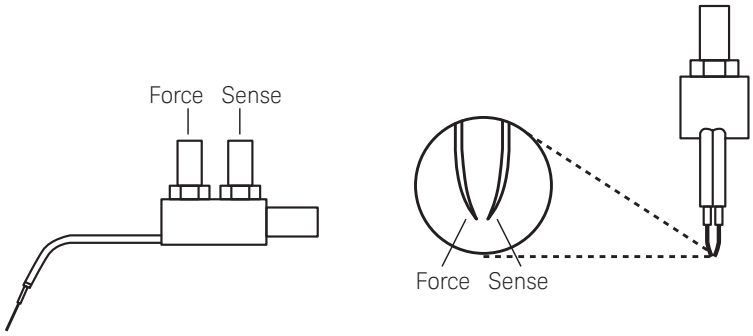


Figure 2.16. Kelvin triaxial positioners (shown above) are ideal since they terminate the force and sense lines at the DUT pad.

Kelvin triaxial cables

Keysight Technologies can supply Kelvin triaxial cables. Each Kelvin triaxial cable can replace two standard triaxial cables. The internal configuration of the Kelvin triaxial cable is as shown below:

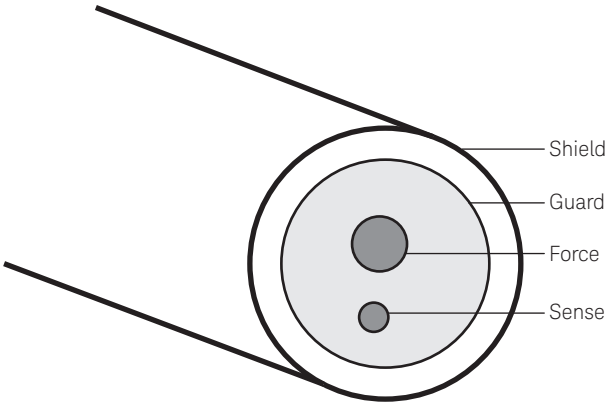


Figure 2.17. The Kelvin triaxial cable combines both the force and sense lines into a single cable.

The Kelvin triaxial cable generally provides better low-noise performance than using two standard triaxial cables. In addition, the connector on the Kelvin triaxial cable makes it impossible to connect the force and sense lines incorrectly.

Note: You should only order this cable if you are certain that whatever equipment you plan on using it with (wafer prober, switching matrix, etc.) supports this connector.

Figure 2.18. The Kelvin triaxial cable mates with Keysight SMUs and many analytical wafer prober connector plates.

Noise in electrical measurements

Noise is an inescapable part of electrical measurement that arises from the fact that all electronic charge is discrete. It is impossible to give adequate coverage to all of the possible types of noise that can be encountered in parametric testing in this short section. Instead, we will give a brief overview of the most common types of electrical noise.

Shot noise

Shot noise is the noise associated with electrical current. It is the result of the fact that an electrical current consists of a flow of electrons and that the electrons have a discrete charge. The value that we define as the electrical current is the average value of the number of electrons passing through a given cross section per second. This is given by:

$$I = nq$$

Where q is the value of the electron charge and n is the average number of electrons passing through the given cross section per unit time. The true current, i(t), is time dependent and fluctuates around this average value depending upon the actual number of electrons passing over the given cross section at any instant in time. The shot noise can therefore be written as:

$$i_{shot}(t) = i(t) - i = i(t) - nq$$

By making certain idealized assumptions, we can show that the spectral density function of the shot noise is proportional only to the electronic charge and the value of the average current:

$$S_i(f) \alpha qI$$

The most important feature of this function is that the noise spectrum is constant at all frequencies. Spectra with this type of frequency independence are referred to as white noise. Shot noise can have a particularly strong impact on parametric measurements when you are trying to measure a very small signal in a DUT that is being biased by a much larger DC current.

It should be pointed out that the above results are only an approximation. In reality, one can easily see that the spectral density function cannot have a constant value across all frequencies. If we were to calculate the mean square current by integrating the spectral density function across all frequencies (from zero to infinity), then we would end up with an infinite value. This does not make any physical sense, since no real current can supply an infinite noise power. There has to be an upper frequency limit at which the spectral density function drops off. However, for most situation of interest in parametric tests, this frequency is so high that we can make the simplifying assumption that shot noise is constant across frequency.

Thermal noise

Resistors that are not carrying any current can exhibit voltage fluctuations due to the random thermal motion of electrons. This thermal noise (also known as Johnson noise) results in voltage fluctuations that have a zero time average value and a spectral density function described by the equation shown below:

$$S_v(f) = 4kTR$$

In this equation, k is the Boltzmann's constant, T is the temperature in degree Kelvin and R is the resistance. Since this spectral density function is independent of frequency, it is also a form of white noise. It should be noted that the actual noise observed across a resistor will be larger than that predicted by the above equation due to other thermal processes (such as the continuous generation and recombination of carriers) that are occurring in addition to electron thermal motion. Sometimes this additional noise is accounted for by substituting a "noise temperature" into the above equation, and this noise temperature can be much higher than the actual temperature of the resistor.

Flicker or 1/f noise

Many semiconductor devices exhibit a type of noise that is much greater at low frequencies than that expected by shot or thermal noise. This type of noise has a spectral density function that can be described by the equation shown below:

$$S_i(f) \, \alpha \frac{1}{f^n}, \text{ where } 0 < n < 2$$

In most cases, n is very close to 1. Various types of noise with a 1/f dependency occur in nature, and they are actively studied in many scientific fields. In electronics, 1/f noise is typically referred to as flicker noise, and it is caused by the presence of a direct electric current. You may also see 1/f noise referred to as pink noise, to emphasize its intermediate position between white noise (which has no frequency dependence) and red or Brownian noise (which has $1/f^2$ frequency dependence).

Random Telegraph Noise (RTN)

Random telegraph noise (also known as burst noise or popcorn noise) manifests itself as fluctuations in the drain current of a MOSFET device under active bias. The root cause of RTN is attributed to the trapping and detrapping of electrons, and when RTN occurs, it usually dominates all other low-frequency noise components. Until recently, RTN was mainly the concern of CMOS image sensor manufacturers, where the RTN created white spots in what should have been dark areas. However, as lithographies have continued to shrink, RTN has become of concern to more device manufacturers, since it can sometimes cause an SRAM cell to erroneously flip to its opposite state. Methods and techniques to characterize RTN will be discussed later in this handbook.

Power line cycle noise

One of the most common sources of noise in parametric testing is power line cycle noise. The AC voltage that measurement instrumentation uses for power typically has a frequency of either 50 or 60 Hz (depending on a given country's standards), and this AC voltage creates powerful electromagnetic fields that can superimpose noise on the quantity under measurement. The effects of power line cycle noise can be mitigated through proper shielding; however, as the measurement range decreases, the effects of power line cycle noise become more pronounced. In this case, the only effective means to eliminate power line cycle noise is to average the measurement over one or more power line cycles. Virtually all parametric measurement instruments have this capability. Careful reading of most instrument specifications reveals that their lowest level of measurement resolution actually requires integration over 16 power line cycles. Power line cycle integration will be discussed further in chapter 3.

Chapter 3 Source/Monitor Unit (SMU) Fundamentals

"You do not really understand something unless you can explain it to your grandmother." — Albert Einstein

SMU overview

Introduction

The primary measurement resource for parametric test is the source/monitor unit or SMU. This is sometimes also referred to as a source/measurement unit, although the acronym ends up the same. The SMU can force voltage or current and simultaneously measure voltage and/or current. The diagram below shows a simplified SMU equivalent circuit:

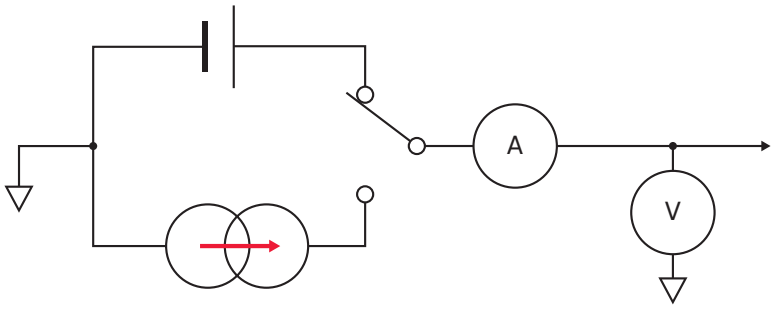


Figure 3.1. Simplified SMU schematic.

Since SMUs must measure very low currents (1 fA or less), they always have triaxial outputs for the reasons discussed in the previous chapter.

There are different types of SMUs. The most common SMU is the medium-power SMU (MPSMU); as the name implies this SMU can supply moderate levels of voltage and current (±100 V and ±100 mA). It also has current measurement resolution down to 10 fA. For precision measurements there is the high-resolution SMU (HRSMU); this SMU can supply the same current and voltage as the MPSMU, but it provides current measurement resolution of 1 fA or less and voltage measurement resolution of 0.5 µV. The high-power SMU (HPSMU) is for situations requiring larger currents and voltages than that supplied by the MPSMU or HRSMU; this SMU can supply current up to ±1 A and voltages up to ±200 V, and the measurement resolution capability is similar to that of a MPSMU.

In addition to the basic SMU types just mentioned, the B1500A MPSMU and HRSMU support an additional module that enables them to achieve current measurement resolutions of 0.1 femtoamps (100 attoamps). In order to achieve this level of current measurement resolution, the actual measurement unit has to be placed in close proximity to the device under test (DUT). This means that this module has to be mounted onto the wafer prober and connected back to an SMU installed in the parameter analyzer mainframe via some sort of cabling arrangement. In the case of the Keysight B1500A, this module is called the atto-sense and switch unit (ASU). Besides being able to provide 0.1 fA current measurement resolution, the ASU also has some switching capabilities that will be discussed further in chapter 4.

Introduction *(continued)*

Module	HPSMU	MPSMU	HRSMU	ASU
Maximum Force Voltage	±200 V	±100 V	±100 V	±100 V
Maximum Force Current	±1 A	±100 mA	±100 mA	±100 mA
Voltage Measurement Resolution	2 µV	0.5 µV	0.5 µV	0.5 µV
Current Measurement Resolution	10 fA	10 fA	1 fA	0.1 fA

Figure 3.2. The key specifications of the basic SMU module types.

The B1505A and B1506A support two unique types of internal SMU modules. The high-current SMU (HCSMU) operates in two modes: DC and pulsed. In DC mode the HCSMU can source up to ±1 A at 40 V; in pulsed mode the HCSMU can source up to ±20 A at 20 V. The HCSMU has a unique output configuration and cabling requirements that will be discussed later in this chapter. The high-voltage SMU (HVSMU) can source up to 3000 V at 4 mA. Due to its high output voltage capability, the HVSMU requires a high-voltage triaxial cable that has a special screw-on triaxial connector. This ensures that a standard triaxial cable cannot accidentally be used with this module.

Module	HCSMU	HVSMU
Maximum force voltage	±40 V (DC) ±20 V (Pulsed)	±3000 V
Maximum force current	±1 A (DC) ±20 A (Pulsed)	±8 mA at ±1500 V ±4 mA at ±3000 V
Voltage measurement resolution	200 nV	200 µV
Current measurement resolution	10 pA	10 fA

Figure 3.3. The key specifications of the HCSMU and HVSMU (B1505A and B1506A only).

Note: For the HVSMU module, the output voltage and current must be of the same polarity.

The B1500A, B1505A and B1506A support one additional special SMU type, which is the medium current SMU (MCSMU). Although the MCSMU is supported in both the B1500A and B1505A/B1506A mainframes, its purpose for these products is different. In the B1500A, the MCSMU functions as a standalone module, and it provides two main benefits. The first benefit is that it supports 50 µs SMU pulsing, which is 10 times narrower than the pulses that can be generated by the B1500A MPSMU, HPSMU and HRSMU modules. The second benefit is that it supports the Oscilloscope View feature in the Tracer Test Mode, which allows the user to verify pulsed measurement waveforms right on the front panel of the B1500A. In contrast, in the B1505A and B1506A the MCSMU module is not intended to be used by itself but rather it is used as a control SMU for other modules. All of these features will be explained in greater detail in later sections of this handbook.

Except for the HCSMU and MCSMU, SMUs are single-ended devices with one end always tied to a common internal reference point. The SMU reference level is normally tied to chassis ground via an external shorting bar, but this shorting bar can be removed and the SMU reference can be tied to an external voltage (up to 42 V maximum) using various types of connectors.

Introduction *(continued)*

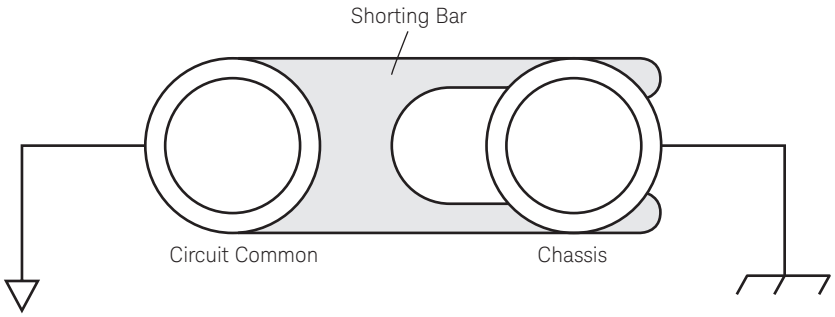


Figure 3.4. You can remove the shorting bar from the chassis ground to float the SMU circuit common.

Note: In the case of the B2900A, the SMU channels are tied to chassis ground via an internal relay. This relay is normally closed, but it can be opened using a software command to allow for floating SMU measurements.

All SMUs have some pulsing capability during sweep measurements (used to prevent heating on thermally sensitive devices), and SMUs can also make time sampling measurements. However, the pulsing and time sampling capabilities of SMUs are relatively slow (in the microsecond range). It is important to understand when you can use an SMU to make pulsed measurements, and when you need the pulsing capabilities of a semiconductor pulse generator unit (SPGU) or waveform generator/fast measurement unit (WGFMU) (both of which have nanosecond pulsing capabilities). These different solutions will be covered in detail when we discuss making high-speed measurements in chapter 5. Because of their relatively slow pulsing capabilities, triaxial cables and DC probes can be used with SMUs when performing pulsed and time sampling measurements. However, as will be discussed in subsequent chapters, pulsing and fast measurements made with other types of modules (such as the HV-SPGU and WGFMU modules) require specialized cables and RF probes for optimal measurement results.

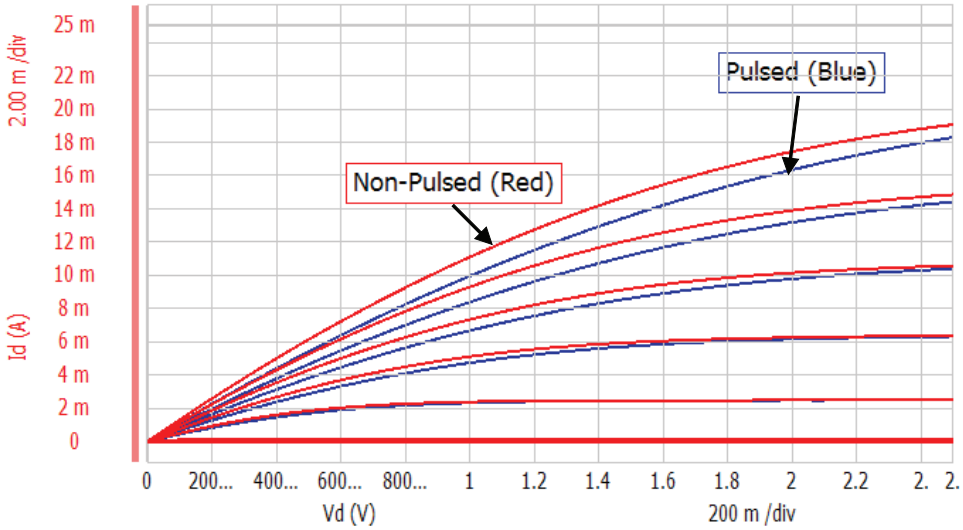


Figure 3.5. SMUs can be pulsed during sweep measurements to eliminate device self-heating effects that can distort measurement results as shown in the above MOSFET current drain versus voltage drain plots.

SMU operation modes and settings

SMUs have three basic modes of operation: voltage source, current source and common. In common mode, the SMU acts as a 0 V voltage source, it cannot perform any measurement, and the current compliance is automatically set to the SMU's maximum value. In addition, for sweep measurements, you can set SMUs in voltage pulse or current pulse mode to prevent device self-heating on thermally sensitive devices.

SMUs have the ability to specify a compliance setting. The compliance setting is always opposite to that of the source setting of the SMU (that is, current compliance when the SMU is in voltage source mode and voltage compliance when the SMU is in current source mode). When an SMU reaches compliance, it acts as a constant voltage or current source. The compliance feature prevents inadvertent device damage by not allowing the measured quantity to exceed the specified compliance value. In addition, on swept sources, it is also possible to specify power compliance. The power compliance prevents the total power output by the SMU from exceeding the specified power compliance value. If both standard and power compliance are specified, then the SMU will never exceed whatever is the lower of these two settings.

On the "Measurement Setup" tab in Keysight EasyEXPERT software you can specify the various compliance settings for each SMU. In addition, there is a "Sweep status" selection menu that allows you to select either "CONTINUE AT ANY" or "STOP AT ANY ABNORMAL". The "continue at any setting" will continue making measurements regardless of any abnormal conditions that may occur (such as a measurement error, reaching compliance, etc.). Conversely, the "stop at any abnormal" setting will immediately halt testing when any of these conditions occurs. Especially in the case of automated measurements (where you are not monitoring the status of the instrument as it measures), the stop at any abnormal feature can be valuable in reducing needless measurement time. An example of this is shown below.

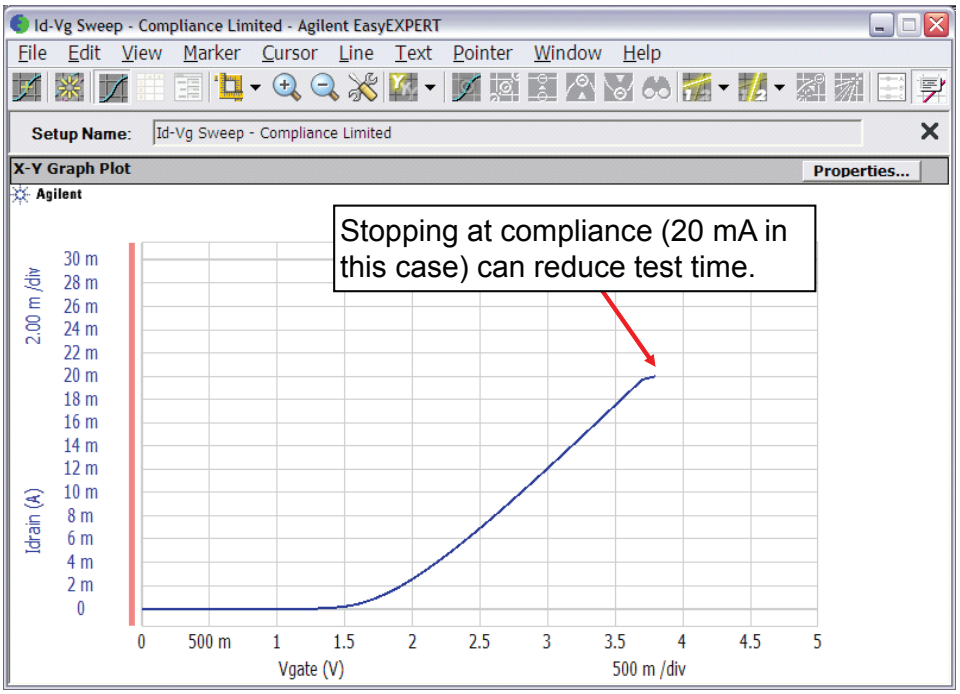


Figure 3.6. By using the **stop at any abnormal** setting you can automatically halt testing when compliance is reached and avoid needlessly continuing a measurement.

SMU force and sense outputs

The situations requiring Kelvin measurements and the basic theory of Kelvin measurement have already been discussed. By separating the force and sense lines you can eliminate the effect of cable resistance from your parametric measurement. To make this task easier, modern SMUs are designed with both force and sense outputs. The following illustration shows the output configuration of a Kelvin SMU.

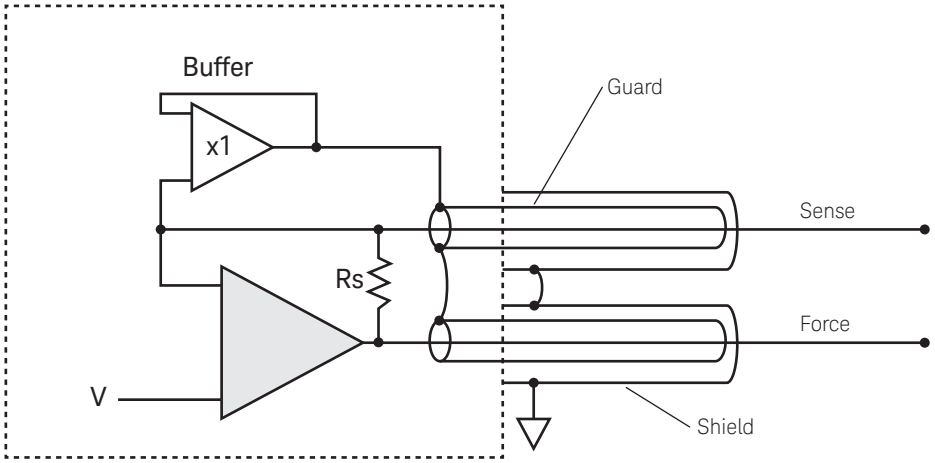


Figure 3.7. Simplified Kelvin SMU output circuit.

The great benefit of a Kelvin SMU configuration is that you only need two SMUs to perform a Kelvin measurement as shown below.

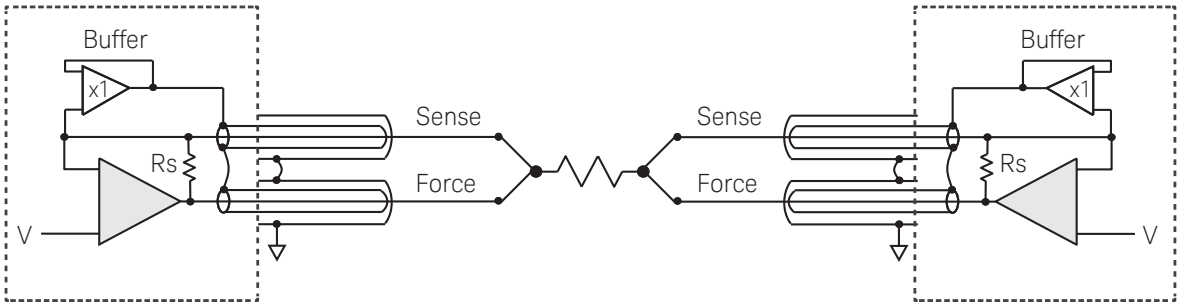


Figure 3.8. Making a Kelvin measurement on a resistor with two Kelvin SMUs.

It is important to point out that the force and sense lines should be shorted together as close to the DUT as possible (for example by using Kelvin triaxial positioners), since the effects of any additional resistance beyond the shorting point will not be eliminated from the measurement.

SMU force and sense outputs *(continued)*

Many situations do not require Kelvin measurements, so it is important to understand which SMU output to use when not making Kelvin measurements. If not making Kelvin measurements and using only one output of the SMU then you must always use the **force** output. The force and sense lines are internally connected via a resistor and this internal sense point is a high-impedance node, so the SMU has no problems with monitoring the current and voltage on the force output when only using the force line. However, consider what happens if you use only the sense line. In this case the entire current flowing out (or in) through the sense output has to pass through the internal resistor as shown below.

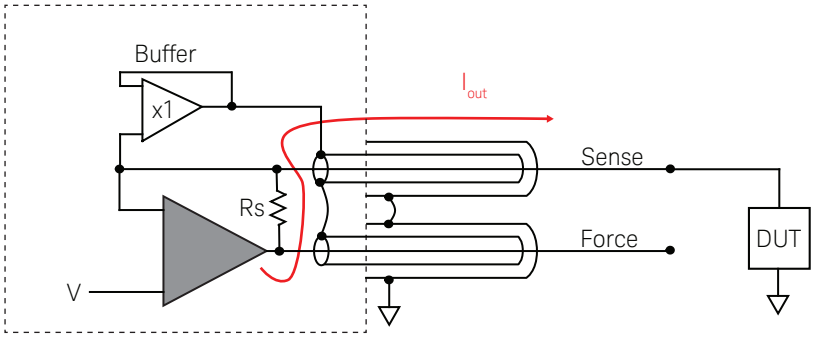


Figure 3.9. The problem with using only the sense line on a Kelvin SMU.

This will distort all of your measurement results and give completely incorrect measurement data.
Note: Never connect only the sense line of a Kelvin SMU to the DUT.

It is possible to use the SMU sense line as a buffered voltage monitor of the SMU force line in non-Kelvin measurement situations. The most common case is when you are driving the gate of a MOSFET with the force line of an SMU and you want to monitor the gate voltage. Using a triaxial to coaxial adapter (guard floating) you can connect the SMU sense line directly to the oscilloscope input as shown below.

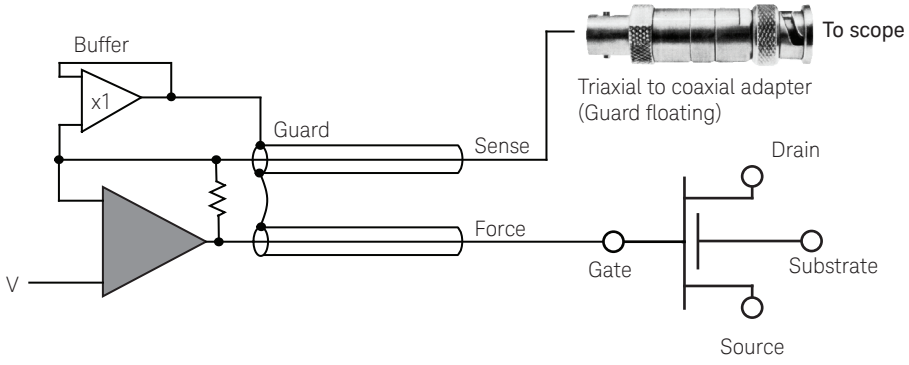


Figure 3.10. If the sense line of the SMU is not used, then you can connect it to an oscilloscope to monitor the SMU output.

Note: When placing any loading onto the output of an SMU you have to be careful, as there is a specified limit on the capacitance. Attaching large capacitive loads to SMU outputs can result in oscillation, so you need to make sure that your oscilloscope input capacitance does not exceed the specified limit of SMU load capacitance.

Understanding the ground unit

The ground unit (GNDU) is a special type of SMU that does not have any measurement capabilities. Its purpose is to provide an active ground to use with the other instrument resources. The ground unit will always maintain a voltage of zero volts as long as you do not exceed the maximum specified value that it can sink or source (for example ±4.2 A in the case of the Keysight B1500A, B1505A and B1506A). The benefit of using an active ground versus a passive ground should be obvious: you do not need to worry about the ohmic drop from large currents distorting your measurement results (at least if you maintain a Kelvin environment).

The configuration of the ground unit is a source of confusion for many users. This confusion is understandable given that, while the ground unit looks like a standard triaxial connection, it is not. The ground unit has the configuration that it does for historical reasons; in the past there simply was not enough room on some instruments for a separate force and sense line for the ground unit. Therefore, the force and sense lines were merged into a single connector. A comparison of the connections on a standard triaxial cable and the ground unit is shown below.

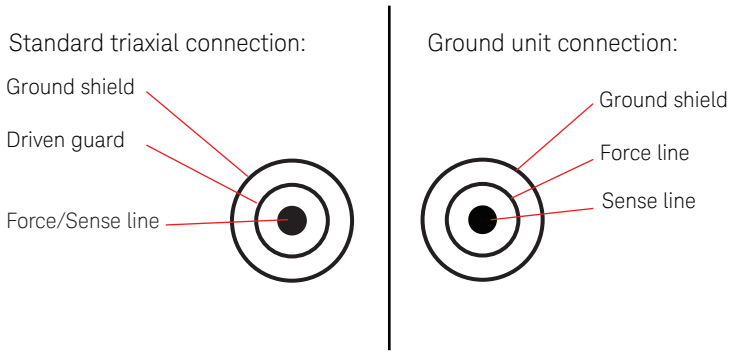


Figure 3.11. Comparison of standard triaxial connection and the ground unit.

The ground unit can have the configuration that it does because the potential of the force and sense lines is always at zero volts, so there is no need to isolate them from the outer ground shield to prevent leakage currents. It should be noted that the ground unit is the only case in which this scheme can work.

Understanding the ground unit *(continued)*

The illustration shown below highlights the problem with connecting up the ground unit like a standard triaxial output.

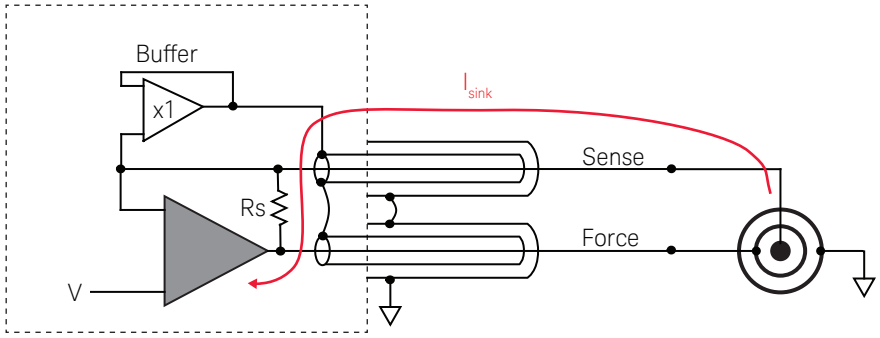


Figure 3.12. The problem with connecting up the ground unit like a standard triaxial connection.

As you can see, connecting up the ground unit like a standard triaxial connection is equivalent to connecting up only to the sense line on a Kelvin SMU. Obviously, this will lead to erroneous measurement results.

The following illustration shows the proper way to connect the ground unit to standard Kelvin SMU connections. Keysight can supply a ground unit to Kelvin adapter, the N1254A-100, which converts the ground unit output into the force and sense outputs as shown below.

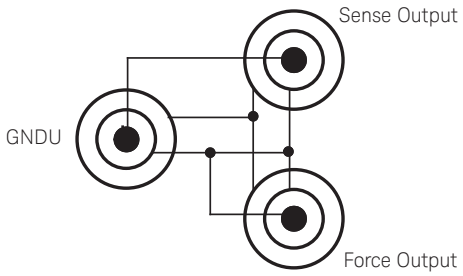
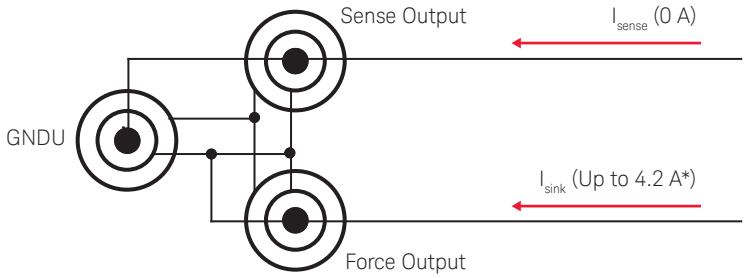


Figure 3.13. The proper way to connect a ground unit to standard triaxial connections.

Understanding the ground unit *(continued)*

Of course, the ground unit adapter can be used with only the force output connected just like a standard SMU. However, you need to be careful when doing this. Assumedly one reason that you are using the ground unit is to sink large currents, and large currents innately require a Kelvin connection (using both the force and sense outputs). Therefore, especially if you are using the ground unit to sink current from one or more high-power SMUs, it is highly recommended that you connect the ground unit in a Kelvin configuration as shown below.



*B1500A, B1505A & B1506A

Figure 3.14. The ground unit should be used in a Kelvin configuration when sinking large currents.

Note: Keysight makes a special triaxial cable designed to be used with the force output of the ground unit. This cable can handle the 4.2 amps maximum current. The part numbers are 16493L-001 (1.5 m), 16493L-002 (3 m) and 16493L-003 (5 m).

High current SMU connections (B1505A and B1506A)

As was mentioned earlier, the high current SMU (HCSMU) is a special module available only for the B1505A and the B1506A. Its structure is similar to that of an SMU except that it can source up to 20 A of current (pulsed). Because no other module can sink this much current, the HCSMU also has to have the ability to sink its own current. This gives the HCSMU a unique output configuration.

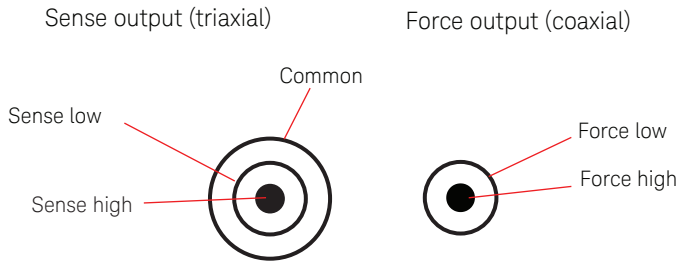


Figure 3.15. The output connections on the HCSMU.

The force lines of the HCSMU do not perform any measurement so they do not require any shielding and they can be coaxial. On the other hand, the sense lines of the HCSMU do perform measurement so they require shielding and they need to be triaxial. Fortunately, this arrangement makes it impossible to confuse the two outputs.

The HCSMU module is floating and is not tied internally to the instrument ground. This means that its low force and sense outputs must be tied to a reference level (normally the ground unit) when making a measurement. An example MOSFET measurement using the HCSMU is shown below.

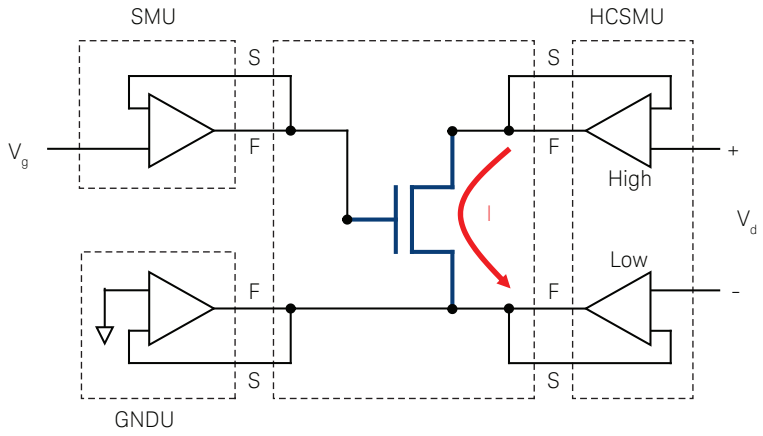


Figure 3.16. The HCSMU needs to have its low outputs (force low and sense low) tied to a known voltage reference when making a measurement.

When you are using the HCSMU with the N1259A or N1265A high-power test fixtures, both units take care of correctly separating out the HCSMU connections. However, you need to make sure that you use the correct cables and adapters when connecting the HCSMU up to a wafer prober. Keysight supplies a variety of adapters and connectors for this purpose. This will be covered in greater detail when we discuss making on-wafer measurements in chapter 4.

Medium Current SMU Connections (B1500A)

The MCSMU does not have standard internal triaxial connections, so an adapter is required for their use as standalone modules. As was mentioned earlier, the MCSMUs were designed to be used as control SMUs in the B1505A and B1506A. Their outputs are similar to the HCSMU (which can also be used as a control SMU) except that the MCSMU force output is triaxial. Also, the positioning of the force and sense outputs follows the convention of standard SMUs (force on the left and sense on the right).

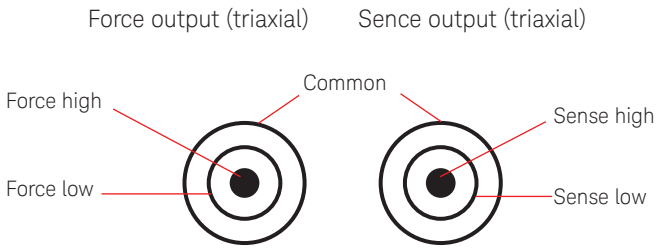


Figure 3.17 – The output configuration of the MCSMUs.

Just like the HCSMUs, the MCSMUs do not have an internal ground reference. To use the MCSMU as a stand-alone module you must use the N1255A connection box, which is included when a B1500A with these modules is specified. A picture of this box is shown below.

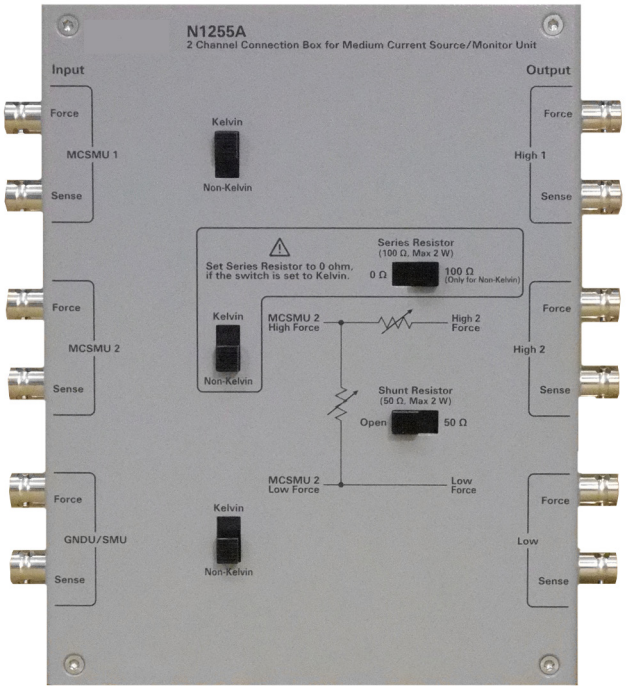


Figure 3.18 – N1255A connection box.

The connection box can be placed on top of a semiautomatic wafer prober, and it is heavy enough so that stability is not an issue. The following figure shows a schematic of the connection box.

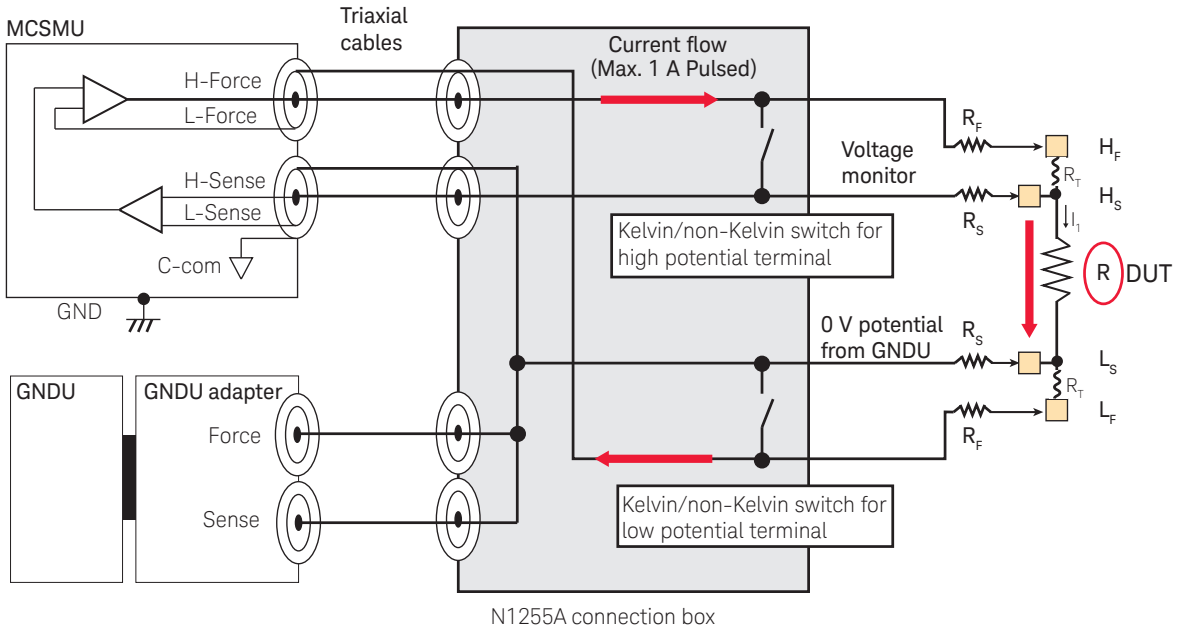


Figure 3.19 – Schematic of N1255A connection box.

Note that the MCSMUs require some sort of reference potential. Although this is typically supplied by the ground unit (as shown in the schematic), an SMU can also supply the reference potential. When using an SMU, the low potential supplied to the N1255A connection box can support floating measurements up to ±200 V.

Measurement ranging

Measurement ranging basics

Measurement ranging is intimately interrelated with measurement accuracy and resolution. However, before proceeding it is important to understand why SMUs have a range setting in the first place. The SMU circuitry has to switch in and out (using relays) different resistor values in order to handle the maximum expected current or voltage value based upon the initial compliance setting specified by the user. An example of this circuitry is shown below.

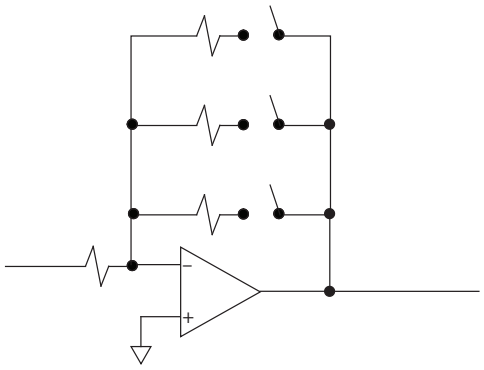


Figure 3.20. SMU ranging requires that resistors of different values be switched in and out of the circuit depending upon which particular range you are in.

Selecting one or more resistors via these relay switches places the SMU into a given measurement range. Obviously, it takes some time to switch these relays and move from one range to the next. While it would be possible to always have the SMU start at the highest possible measurement range and work its way down to the lowest measurement range that contains the quantity being measured, this would result in extremely slow measurements. By allowing flexibility as to how a measurement range is selected, you gain the ability to trade off measurement speed versus accuracy.

There are typically three types of ranging selectable on an SMU: fixed, limited, and auto. The illustration below shows how each of these choices impacts the measurement range used by the SMU:

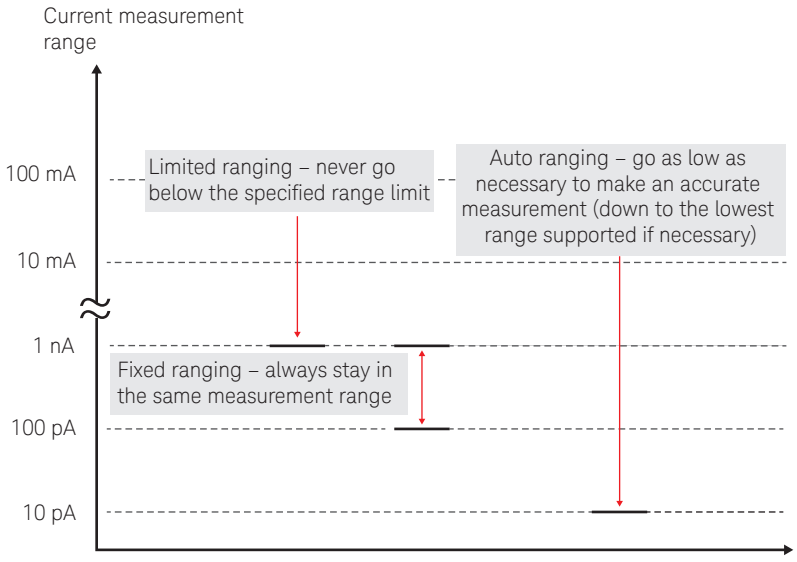


Figure 3.21. An explanation of the three types of measurement ranging: fixed, limited and auto.

Measurement ranging basics *(continued)*

Ranking the measurement ranges in terms of fastest to slowest, the order would be: fixed (fastest), limited (next fastest), and auto (slowest). Although fixed measurement ranging yields the fastest measurement results, it has the limitation that the SMU will not go into a lower measurement range to improve measurement accuracy. Also, if you attempt to measure a current or voltage in fixed measurement range that exceeds the maximum value of the fixed measurement range, you will get a measurement error. Limited ranging and auto ranging are similar in that they both start in the highest measurement range that contains the user-specified compliance value, and they both work their way down to find the optimal range in which to make the measurement. The difference between the two ranging choices is that limited ranging will never go below the user-specified range limit. Thus, limited ranging is useful when you are uncertain of the value of current or voltage that you will be measuring, and when you do not care about ultra-precise measurement. However, if you absolutely need to have the best measurement accuracy and measurement time is not a concern, then auto ranging is the correct choice.

There is no hard and fast rule to specify the measurement resolution achievable within a given measurement range. The measurement resolution discussed in chapter 2 pointed out that the primary determining factor is the number of bits in the SMU analog-to-digital converter (ADC). However, the actual measurement resolution achievable has to take into account other factors such as noise, drift, etc. This introduces a stochastic element to the measurement that requires averaging. The net result is that in most cases the minimum measurement resolution is 4-5 decades below the measurement range. Note that on some instruments, the SMUs may have more than one ADC available to them; in this case you need to check carefully to make sure that you understand the measurement resolution associated with the ADC that you are using.

There is one important point to understand regarding the use of fixed measurement ranging that was mentioned previously but is worth repeating again here. In fixed measurement ranging, you can measure currents or voltages smaller than the measurement range that you have chosen, since as mentioned above, the measurement resolution is 4-5 decades below the measurement range. Even if the actual measured value is more than 4-5 decades below the selected measurement range, the instrument will still return a result (albeit with reduced measurement resolution). However, if you try to measure a current or voltage that exceeds the specified fixed measurement range, then you will get a measurement error. Therefore, finding the optimal fixed measurement range is always a balancing act between selecting a range low enough to provide sufficient measurement resolution and high enough to always contain the quantity under measurement.

Measurement range management

In order to understand the range management feature and why it is sometimes necessary, we must first understand the issue of range compliance. We previously explained that the SMU switches various resistors in its internal circuitry in and out as it moves between different ranges. This process takes some time and it is part of the normal operation of the SMU. However, what was not mentioned was that as the SMU (for example) moves into lower current measurement ranges, its output sourcing capability is also temporarily reduced to the value of the measurement range plus a guardband factor (around 10%). This is called **range compliance** - to differentiate it from the SMU compliance that is specified by the user. Obviously, range compliance is only a transient issue in that if the SMU needs to source more current then it will move up into the next measurement range until it either can supply the required current or it reaches the user-specified SMU compliance.

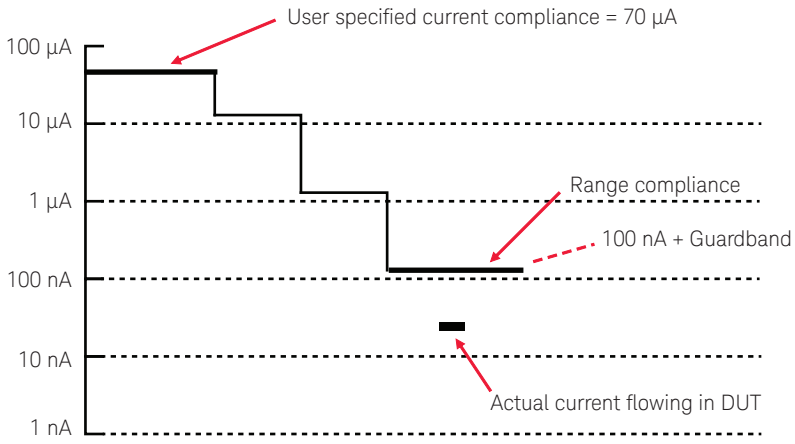


Figure 3.22. Range compliance limits the amount of current that an SMU can supply in a given measurement range.

In order to understand why range compliance can sometimes be an important issue, please refer to the figure shown below.

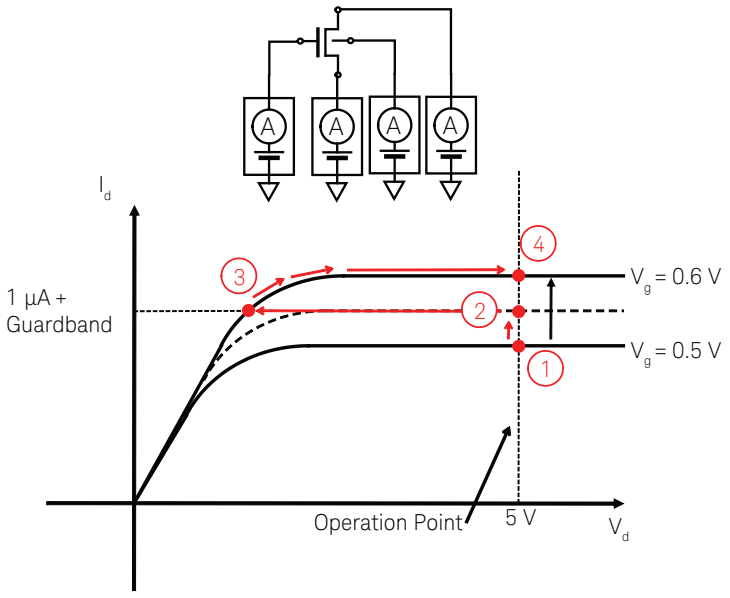


Figure 3.23. The effect of range compliance on SMU voltage output.

Measurement range management *(continued)*

Suppose that we are measuring a device with the Id-Vd characteristic shown in Figure 3.24 and that we are starting at the operating point denoted by "1" (V_g = 0.5 V). We want to move to the operating point denoted by "4" by changing the applied gate voltage to 0.6 V. However, this action requires us to move up into the next measurement range. Since this measurement range change is not instantaneous, as soon as the operating point moves to the position denoted by "2", the SMU cannot supply all of the current that the DUT wants. This means that the operating point must move to the position denoted by "3". When the SMU finally moves up into the next measurement range, it can supply the current requested by the DUT and the operating point moves to position "4". When viewed on an oscilloscope, this would give the appearance of an SMU voltage glitch, even though what really happened is that the DUT is causing the voltage droop because the SMU is starving it for current.

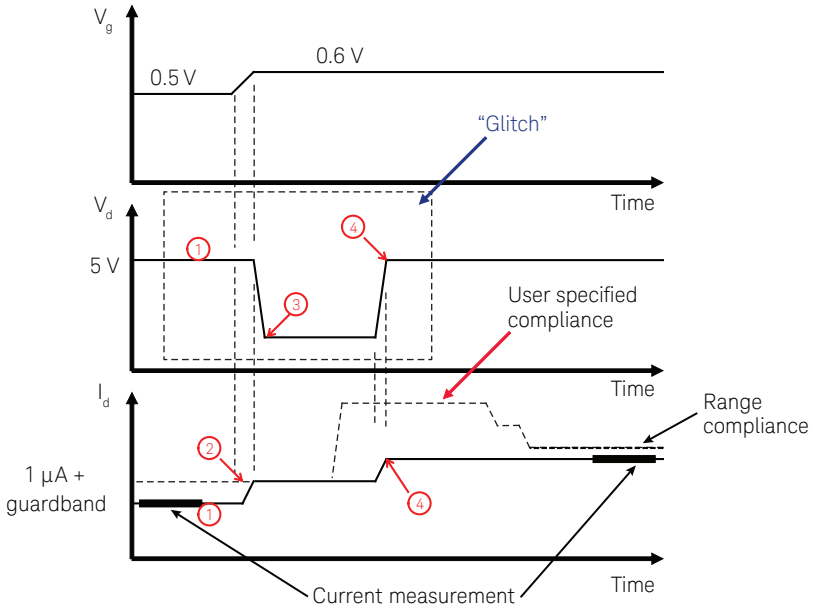


Figure 3.24. Voltage "glitch" on a MOSFET drain caused by temporary range compliance.

For the vast majority of parametric measurements, range compliance never causes any measurement issues. In fact, if we had not just discussed this issue you probably would never have suspected that it existed. However, in a certain small percentage of cases the voltage swings caused by the DUT due to current starvation can impact parametric measurements and even cause device damage (if the voltage swing is on the substrate and it causes the device to latch up). Until the development of the range management feature, the only solution to this issue was to replace the sweep measurement being performed with a series of spot measurements. Replacing the sweep measurement with a series of spot measurements always fixes this issue, since a spot measurement always starts at the measurement range containing the SMU compliance value and works its way down to the correct measurement range (thus avoiding any range compliance issues). However, performing a series of spot measurements in this fashion takes much more measurement time than an equivalent sweep measurement, and this was not acceptable to most users. To solve this issue, Keysight Technologies developed (and patented) the range management feature.

Measurement range management *(continued)*

Range management can be viewed as a sort of "range look-ahead" feature. The range management feature allows you to set trigger points within a measurement range (from 11% to 99% of the range) that force the SMU to go up (or optionally down) to a new measurement range **before** the measurement range limit is reached. The figure shown below helps to illustrate how the range management feature works.

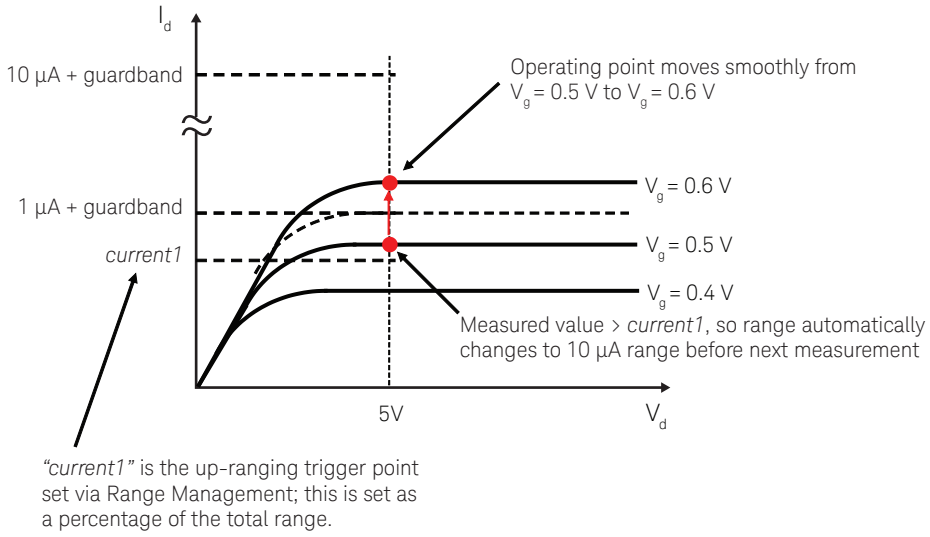


Figure 3.25. Illustration of how the range management feature causes the SMU to move up in range before reaching the range limit.

Applying the range management feature to the case previously discussed, we can see that the voltage droop seen at the MOSFET drain can be eliminated completely.

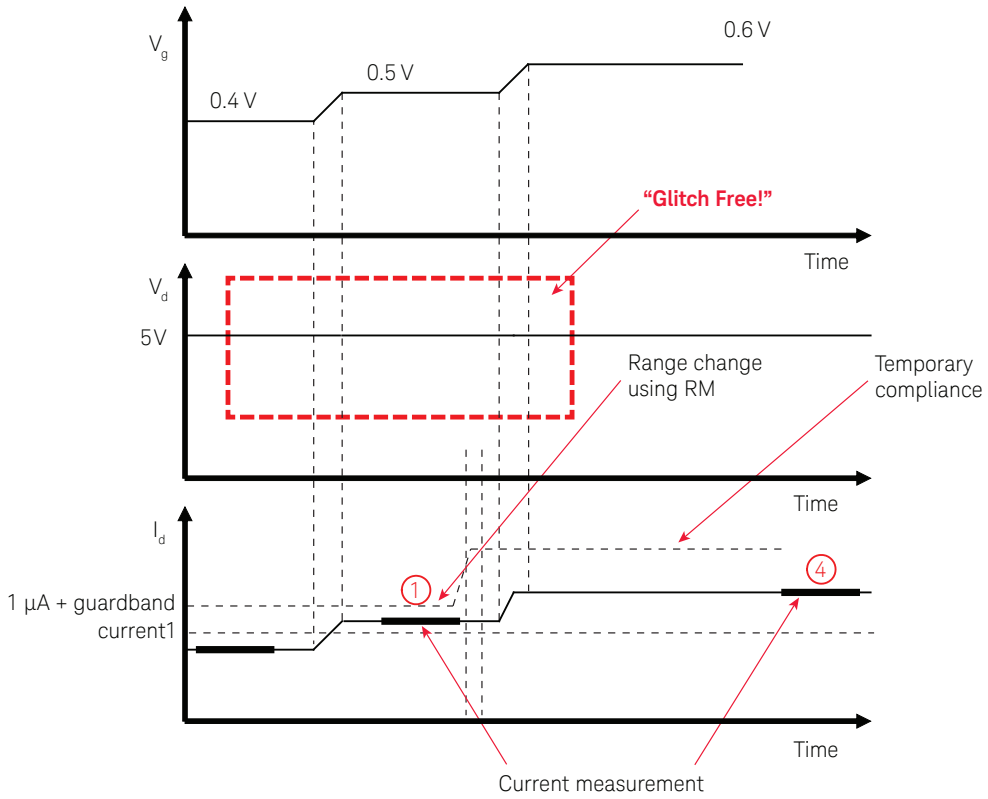


Figure 3.26. Eliminating the effects of range compliance through the use of the range management feature.

Measurement range management *(continued)*

It is a worthwhile question to ask if there are any disadvantages to using the range management feature. The answer is "yes", in the sense that if the range limits are set too low, then you might uprange unnecessarily and lose some measurement accuracy. However, this is a small potential price to pay if range compliance is causing repeated device damage. It should be obvious that the phenomena of SMU voltage spikes due to DUT current starvation can be affected by many factors, including the sweep range, the selected sweep step and variations in the DUT characteristics. Therefore, some trial and error is inevitable in order to find the optimal point at which to set the range trigger ("current1"). Finally, it also needs to be pointed out that not all device damage issues and SMU glitching are due to range compliance issues, so you need to be careful and look at all possible causes if you are experiencing these types of problems.

Keysight EasyEXPERT software supports the range management feature for the B1500A, B1505A and B1506A as part of its built-in GUI. By opening up the range setup window in Classic Test mode, you can modify the range change rule as shown below.

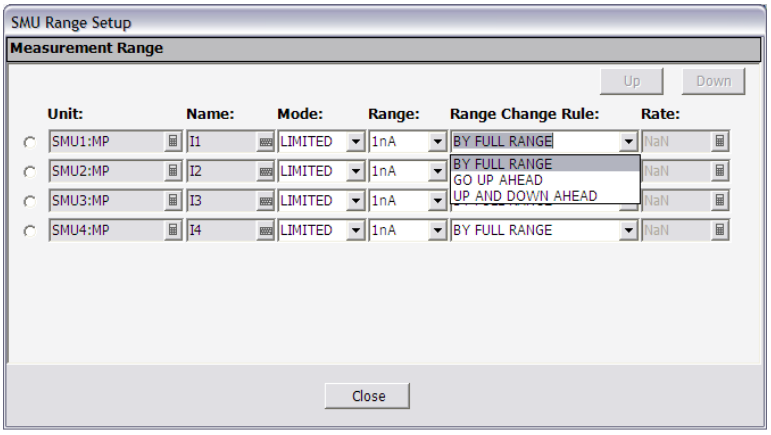


Figure 3.27. Keysight EasyEXPERT software allows you to activate the range management feature using its built-in GUI.

You can select an option to only uprange ("Go Up Ahead") or both uprange and downrange ("Up And Down Ahead"). Once you have specified a range change rule you can then select the rate or percentage of the total range at which you want to have the SMU uprange (or downrange too if "Up And Down Ahead" is selected).

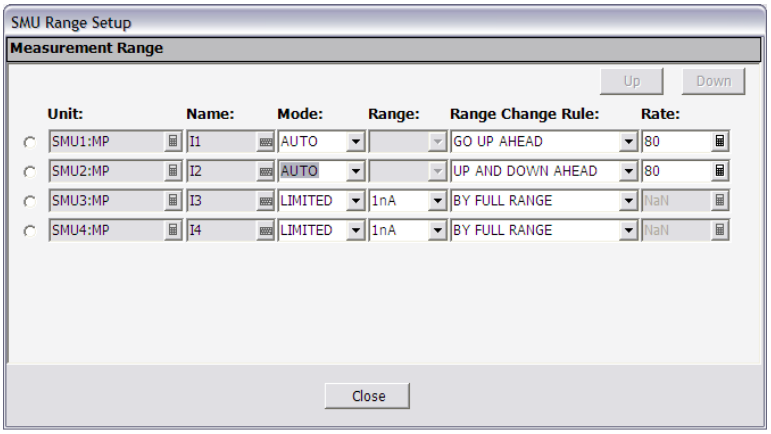


Figure 3.28. The "Rate" parameter sets the percentage of the total range at which the SMU will uprange (or optionally downrange).

Eliminating measurement noise and signal transients

Integration time

Inexperienced users sometimes confuse the purposes of measurement ranging with integration time. It is important to understand that the purpose of integrating a measurement over time is to eliminate noise. Increasing the integration time does not have the same effect as using a lower measurement range. Before increasing integration time to improve your measurement results, you should first determine if you have chosen the correct measurement range for the level of current or voltage that you are trying to measure. For example, it makes no sense to try and make a femtoamp current measurement using limited 1 nA ranging on an SMU, since the SMU needs to get down into the 10 pA range in order to make good femtoamp measurements. As a general rule, lower measurement ranges require longer integration times in order to obtain a satisfactory measurement because noise becomes more of an issue as you try to measure small currents and voltages.

The terminology used for integration time is not uniform across all products. In some cases, the term “Short” is used to refer to any integration time that occurs in a time period that is less than one power line cycle (PLC); “Medium” is then used to refer to integration over exactly one PLC, and “Long” is used to refer to integration over multiple PLCs. In other cases, the terms “Auto” and “Manual” are used for integration times of less than one PLC, and “PLC” is simply used to refer to integration over one or more PLCs. The documentation included with your instrument explains the exact terminology used for your equipment. The important point to understand is not the particular terminology used but rather when to use each type of integration time.

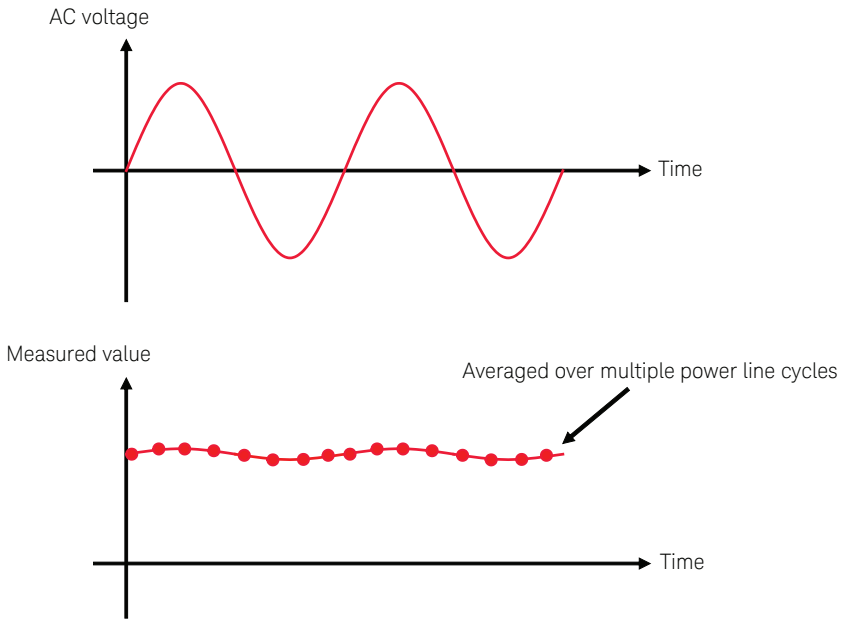


Figure 3.29. Power line cycle (PLC) integration eliminates measurement error caused by noise from the AC supply voltage by sampling over multiple power line cycles and averaging the samples.

Hold time and delay time

In addition to noise, capacitance on the SMU outputs can cause ringing and other transient phenomena each time the SMU applies a new voltage or current. To ensure that the applied voltage or current is stable before making a measurement, you can specify both a measurement hold time and a measurement delay time as shown below.

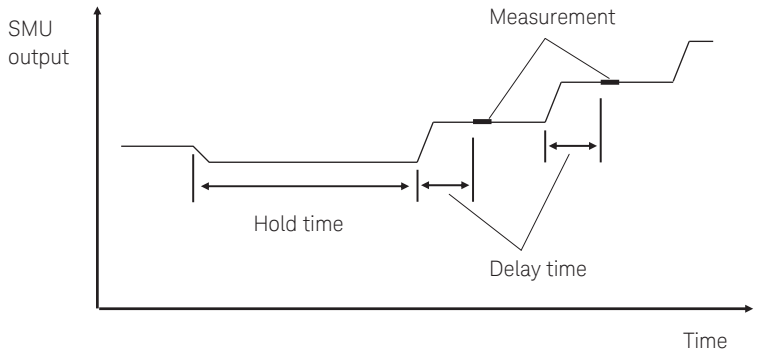


Figure 3.30. The hold time and delay time settings allow you to specify how long to wait before starting a measurement after the SMU applies voltage or current.

The hold time ensures that the SMU outputs are stable before the start of a measurement, and the delay time ensures that the SMU outputs are stable during a measurement if the value of the SMU output is being changed.

Low current measurement

Making low current measurements with SMUs (in the femtoamp or even sub-femtoamp range) presents a variety of measurement challenges. However, none of these challenges are insurmountable, and all of them can be met using well-established measurement techniques. A summary of these techniques is shown below.

- Create a Low-noise Probing Environment
  - Maintain a shielded measurement environment (Faraday cage)
  - Get rid of electrical noise sources
- Eliminate Stray Leakage Currents
  - Use fully guarded cabling and probes (down to the probe tips)
  - Use a guarded wafer chuck
- Choose Appropriate Measurement Setup Parameters
  - Select correct measurement range, integration time and hold time
- Perform Calibration and SMU Zero Offset

Techniques for guarding and shielding have already been discussed in the previous chapter, and techniques for making low-current measurements on-wafer will be discussed in the next chapter. However, there is one SMU feature that is critical for making successful low current measurement that has not yet been discussed: the zero cancel function.

SMU zero cancel function

Offset currents can create some significant measurement challenges when making current measurements in the femtoamp range. There are a variety of sources for offset currents, but the important point is that as long as they are consistent they can be eliminated. The 4155C, 4156C, B1500A, B1505A and B1506A SMUs all have a built-in feature known as “SMU Zero Offset Cancel”. With the outputs open the SMU zero cancel function measures the SMU offset current and then automatically subtract out this offset when the SMU performs subsequent measurements.

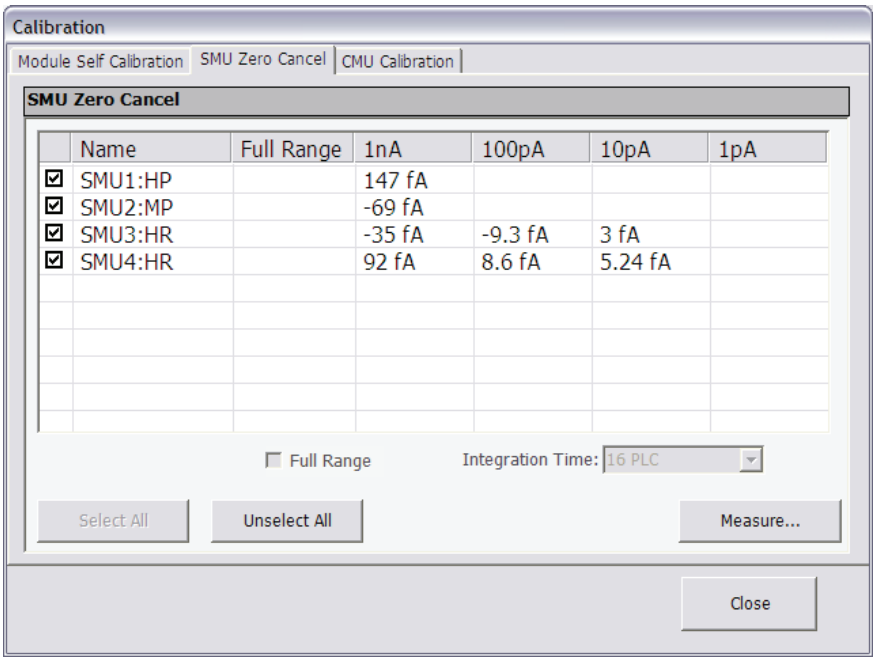


Figure 3.31. The results of performing an SMU zero cancel calibration.

SMU zero cancel function *(continued)*

On the B1500A, B1505A and B1506A (when using EasyEXPERT), the SMU zero offset cancel function status appears in the bottom menu bar as shown below.

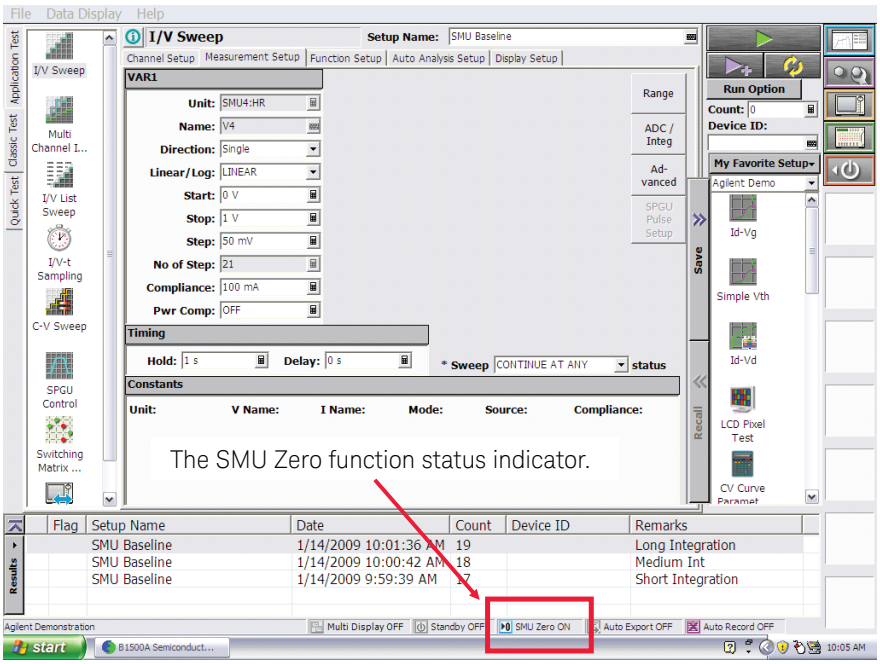


Figure 3.32. The location of the "SMU Zero" function status in the main EasyEXPERT window.

After performing a zero offset cancel calibration, you should be able to obtain a very good (±3 fA) low-current base line measurement on a HRSMU with the outputs open. The following plot shows the results of such a measurement using short integration on the B1500A.

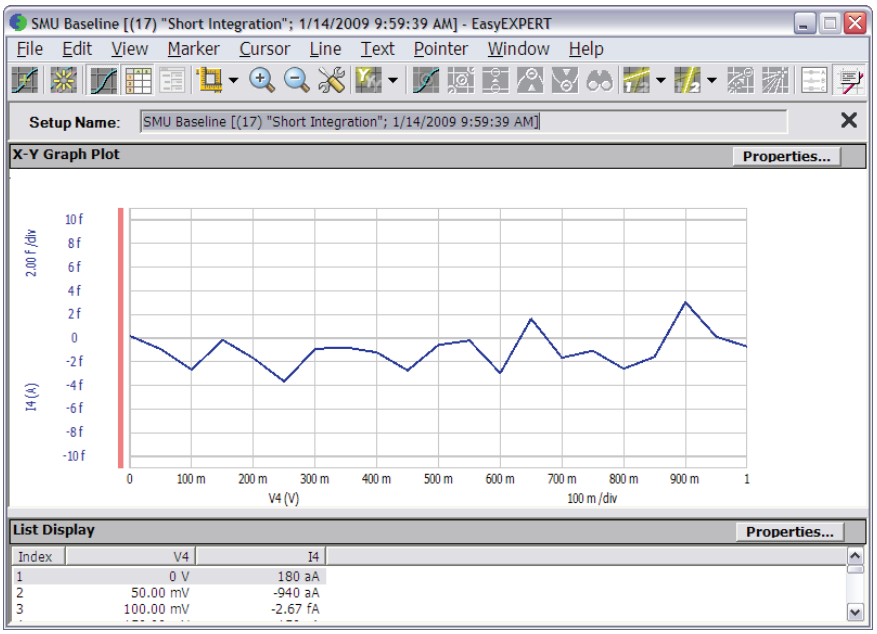


Figure 3.33. The low current base line for a high-resolution SMU (HRSMU) using short integration. As can be seen, this baseline only varies by about ±3 fA.

Debugging low current measurements

Debugging low current measurement issues is not a difficult process as long as you follow a logical procedure. The following debug flow assumes that high-resolution SMUs (HRSMUs) are being used. As mentioned earlier, it makes no sense to try making femtoamp measurements unless at least a 10 pA measurement range is available. If you are having trouble making low-current measurements, then the following debug flow should allow you to isolate the problem area.

1. Sweep 0 V to 1 V with no cables attached (after a 30 minute warm-up)
    a. Check to make sure that all SMU modules passed self-calibration
    b. Perform SMU zero offset cancel
    c. Verify that there is a ±3 fA base line
       Note: Success indicates that the SMUs are functioning correctly.
2. Connect prober cables but do not connect the cables to the wafer prober
    a. Wait several minutes to allow piezoelectric effects to dissipate
    b. Perform SMU zero offset cancel
    c. Check base line
       Note: Success indicates that the cables are not damaged (leaky).
3. Connect the prober cables to the prober connector plate but do not connect them to the wafer probes
    a. Wait several minutes to allow piezoelectric effects to dissipate
    b. Perform SMU zero offset cancel
    c. Check base line
       Note: Success indicates that the connector plates are not leaky.
4. Connect wafer probes to connector plate
    a. Wait several minutes to allow piezoelectric effects to dissipate
    b. With probes up (not contacting wafer), perform SMU zero offset cancel
    c. With probes up (not contacting wafer), check baseline
       Note: Success indicates that the wafer probes are functioning correctly.

Power line cycle noise tends to become more prominent at lower current measurement ranges. This means that you may need to integrate low-current measurements over multiple power line cycles in order to get noise-free results. For sub-femtoamp measurement made using the ASU, this can become even more of an issue. In fact, one experimenter once reported that the only way he could obtain decent sub-femtoamp measurements was to come in during the weekend when most of the electrical equipment in the building in which he was working was shut-off.

Spot and sweep measurements

Spot versus sweep measurements

Two basic types of measurements are made during parametric test: spot and sweep. A spot measurement is a single-point measurement that produces a scalar result. A sweep measurement is a series of measurements that produces a vector result. During a sweep measurement, one independent variable is changed and the dependent variable is then measured and plotted. At first, it might seem that a sweep measurement is simply a series of spot measurements; however, this is actually not the case. In order to make the sweep measurement complete quickly, the instrumentation makes certain assumptions as it moves from one measurement point to the next in the sweep. The basic assumption made is that the measurement range (if in limited or auto-ranging) does not need to change from one measurement point to the next. Of course, if in moving from one point in the sweep to the next, the measurement range needs to change, then the measurement circuitry will make this change. The important point is that, unlike a spot measurement (which always starts at compliance and works its way down to the correct measurement range), a sweep measurement tries to minimize unnecessary range changes. By making this assumption, a sweep measurement can complete in significantly less time than that of a similar collection of spot measurements.

Primary sweep measurement

For a primary sweep, you must specify three parameters: start, stop and step (the amount the sweep variable should be increased between points). The instrument will first measure with the sweep variable at the “start” value, and then it will increase the sweep variable by the amount specified by the “step” value and measure again. This process of incrementing by the step value and measuring continues until the sweep variable reaches the “stop” value.

Note: The 4145A/B referred to the primary sweep variable as VAR1 (for variable #1), and the 4155C, 4156C and B1500A/B1505A/B1506A (in classic mode) also use this same terminology for consistency.

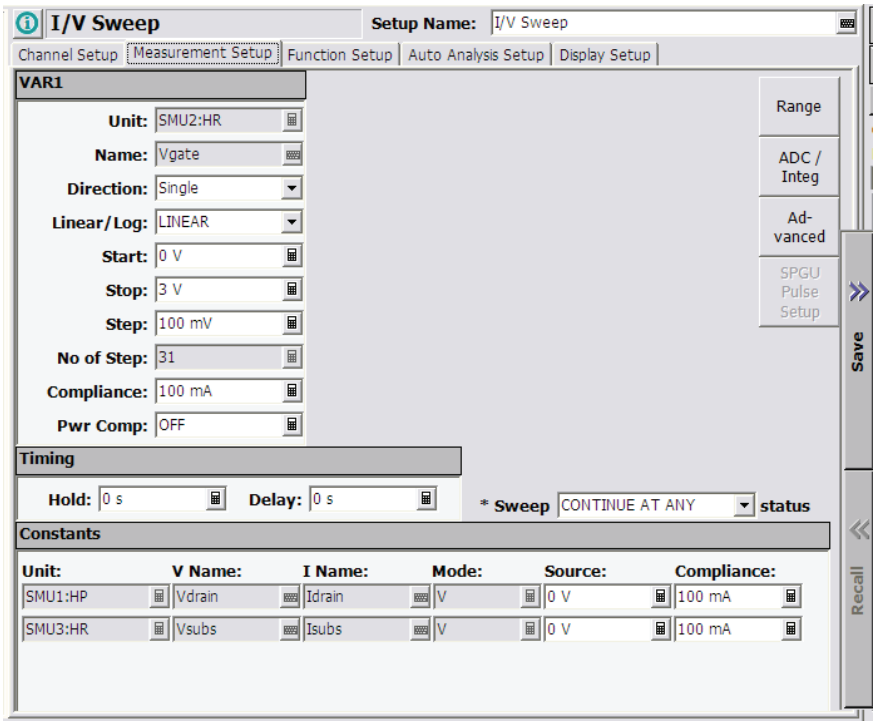


Figure 3.34. Setting up a sweep measurement in classic test mode.

In addition to the aforementioned sweep parameters, the other basic sweep setup parameters are direction and linear/log. The direction parameter allows you to have the sweep only perform a single sweep (from start to stop), or to perform a double sweep (from start to stop and then repeat backwards from the stop to the start value). The linear/log parameter allows you to specify whether or not the sweep steps are spaced linearly or logarithmically. In the case of a logarithmic sweep you can select 10, 25 or 50 points per decade.

Pulsed sweep measurement

As was mentioned in an earlier section, SMUs have the ability to do pulsed measurement.

For the B1500A's MPSMU, HPSMU and HRSMU modules, the minimum pulse width is 500 µs and the maximum pulse period is 5 s (for a minimum duty cycle of 0.01%). In general, this is more than adequate for the testing of most low to medium power devices. When used as a standalone module in the B1500A, the MCSMU module has a minimum pulse width of 50 µs and a maximum pulse period of 5 s (for a minimum duty cycle of 0.001%). The B1505/06A HCSMU also has a minimum pulse width of 5 µs (at 20 A) and a maximum pulse period of 5 s (for a minimum duty cycle of 0.001%). A comparison of a standard sweep measurement versus a pulsed sweep measurement is shown below.

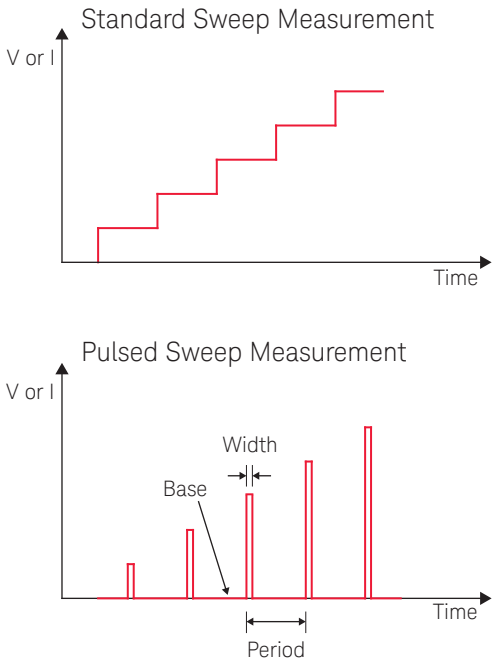


Figure 3.35. By making a pulsed sweep measurement with an SMU, the effects of device heating can be eliminated.

Note: SMUs do not have a small enough duty cycle to prevent self-heating in SOI transistors, as these can require pulse widths in the nanosecond range. Techniques for measuring these types of devices will be discussed in chapter 5.

It is important to remember that when making pulsed measurements you need to make sure that you are not specifying any other measurement conditions that might conflict with the pulse settings. Although this issue will be discussed in greater depth in chapter 5 when we talk about time dependent measurements, the basic rule for SMUs is that accuracy trumps all other settings. This means that you must specify short integration times and use fixed ranging when making very small pulsed measurements or you will not get the pulse width that you specify. In other words, the SMU will make the pulse width however long is necessary to accommodate the integration time and/or range changes that you have specified. If you are using the B1500A, B1505A or B1506A to make measurements, then this is taken care of for you automatically; however, if you are using an older instrument (4155 or 4156) then you need to take care of this manually.

Subordinate sweep measurement

The subordinate sweep function causes the primary sweep variable to repeat each time that the secondary sweep variable is incremented. For a subordinate sweep measurement, you must specify three additional parameters: start, step and the number of steps.

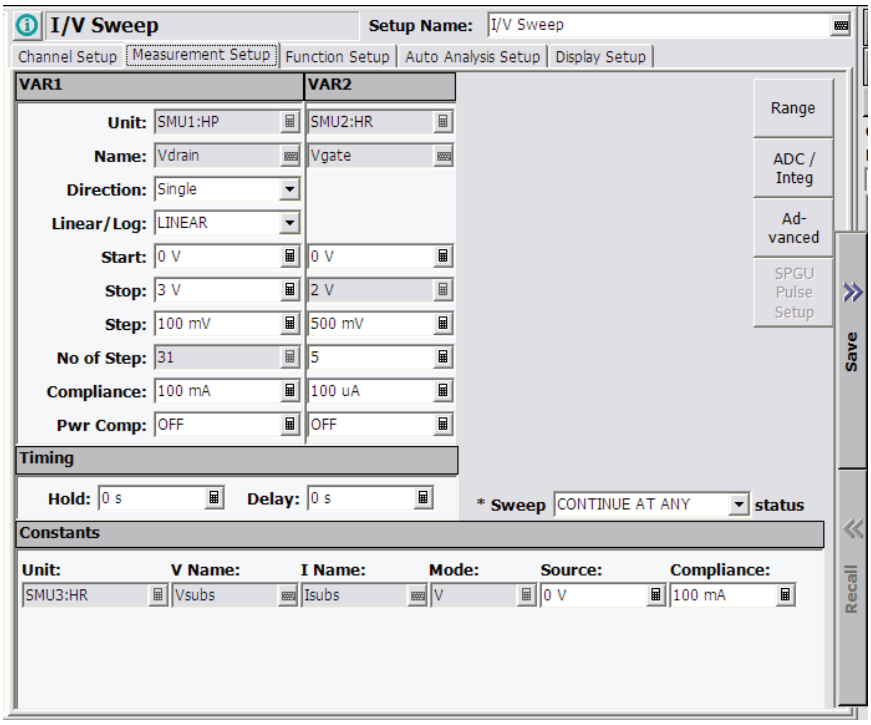


Figure 3.36. Setting up a subordinate sweep in classic test mode.

The instrument will perform a sweep measurement with the subordinate sweep variable at the “start” value, and then it will increment the subordinate sweep variable by the “step” value and repeat the sweep measurement. This process of incrementing by the step value and repeating the sweep measurement continues until the specified number of steps has been reached.

Note: The 4145A/B referred to this variable as VAR2 (for variable #2), and the 4155C, 4156C and B1500A/B1505A/B1506A (in classic mode) also use this same terminology for consistency.

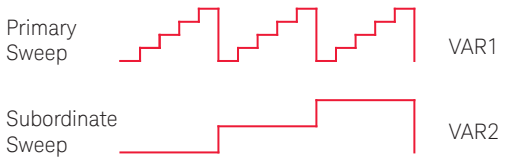


Figure 3.37. The subordinate sweep function causes the primary sweep SMU to repeat each time that the subordinate sweep variable is incremented.

Subordinate sweep measurement *(continued)*

The subordinate sweep function is useful for obtaining a family of curves, such as the I_d-V_d characteristics of a MOSFET as the gate voltage is varied over a range as shown below.

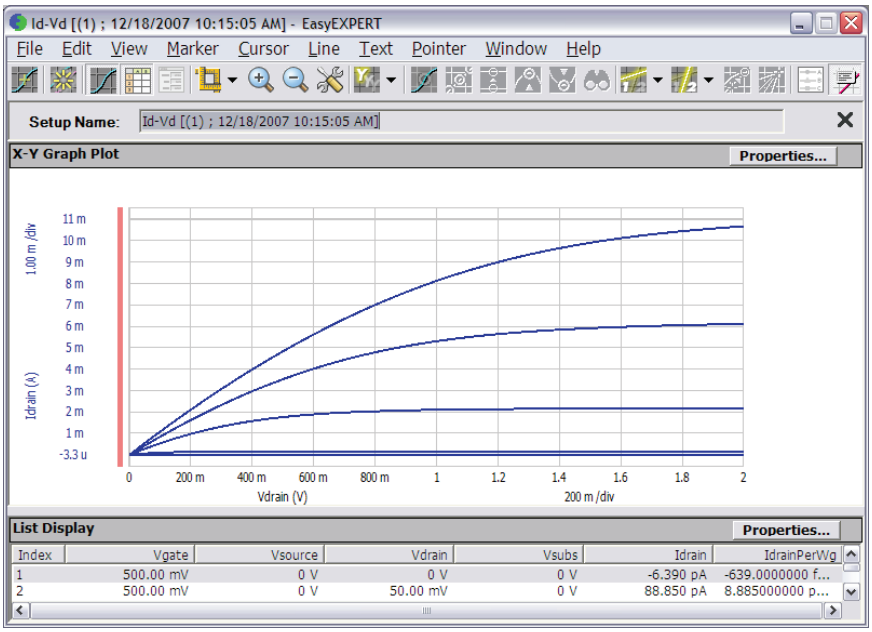


Figure 3.38. The subordinate sweep function can be used to create a family of I_d-V_d curves.

Synchronized sweep measurement

The synchronized sweep function allows you to set up an SMU to track the primary sweep source by a user-specified ratio. For a synchronized sweep measurement, you must specify two additional parameters: a ratio and an offset value as shown below.

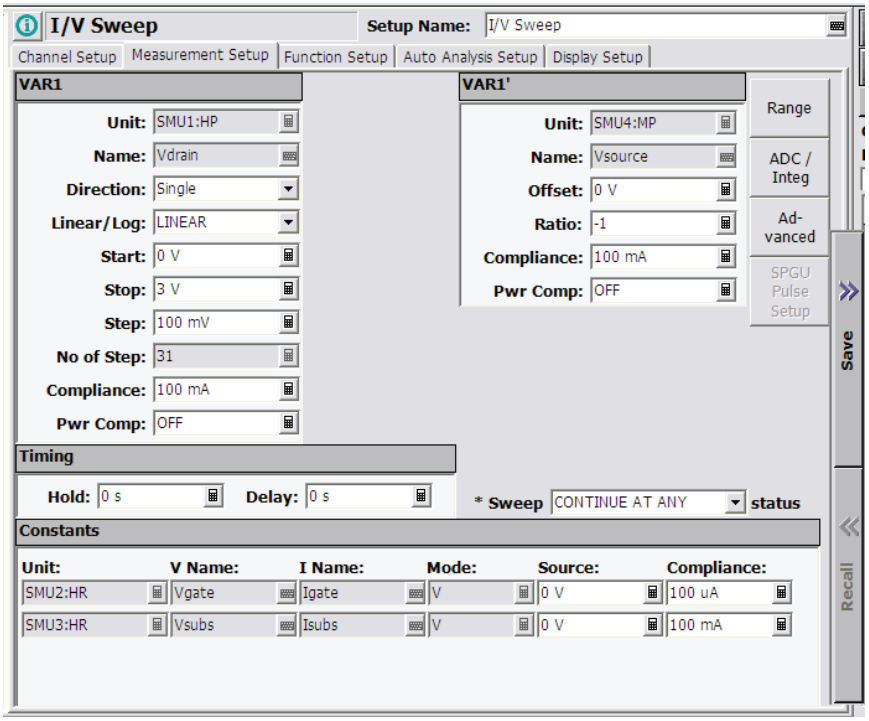


Figure 3.39. The synchronized sweep function allows one SMU to track the primary sweep source by a user-defined ratio.

Synchronized sweep measurement *(continued)*

The instrument will perform a sweep measurement with the synchronized sweep variable tracking the primary sweep variable by an amount determined by the "ratio". The "offset" parameter also allows for the synchronized sweep variable to have a DC offset from the primary sweep variable. The 4145A/B referred to this variable as VAR1**'** (the "prime" of the first variable), and the 4155C, 4156C and B1500A/B1505A/B1506A (in classic mode) also use this same terminology for consistency.

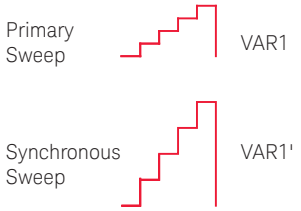


Figure 3.40. The synchronized sweep function allows one SMU to track the primary sweep source by a user-defined ratio.

Although the synchronized sweep function has many uses, the two important uses are placing SMUs in series and parallel (as will be discussed shortly).

Multi-channel sweep measurement

Some instruments (such as the B1500A, B1505A and B1506A) support a true multi-channel measurement sweep capability. An example of the B1500A multi-channel sweep measurement feature is shown below.

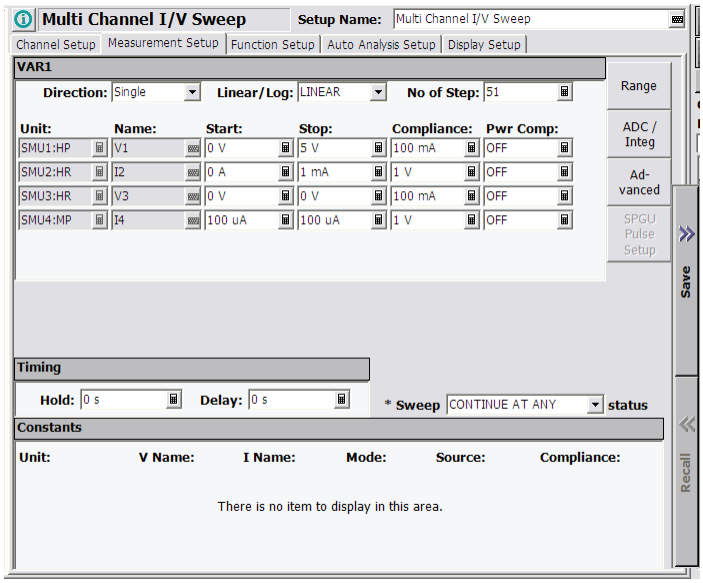


Figure 3.41. The multi-channel sweep function allows all of the available SMUs to be swept simultaneously.

The key points of the multi-channel sweep feature are:

1. You can independently select each swept source to be in either voltage or current force mode.
2. The start and stop values are independent for each swept source, but the number of points in the sweep has to be the same for each source.
3. You can specify the start and stop values to be the same, which effectively makes the swept source output a constant value (the use of this will be explained later).
4. As with the standard sweep function, you can specify power compliance for each swept source.
5. The subordinate sweep (VAR2) feature is supported, but the synchronous sweep (VAR1') feature is not supported.

List sweep measurement

The Keysight B1500A, B1505A and B1506A support a list sweep capability. This feature allows the measurement points in a sweep to be defined by an arbitrary vector list created in a spreadsheet type of format. An example of the B1500A/B1505A/B1506A list sweep function is shown below.

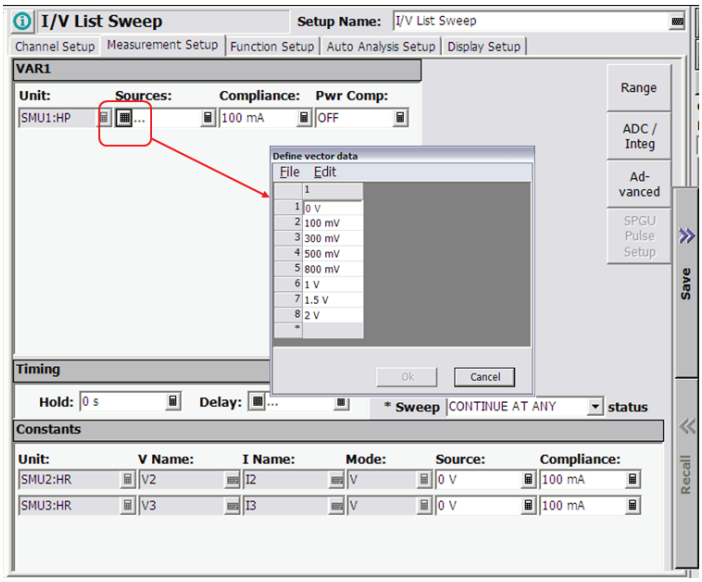


Figure 3.42. The list sweep feature allows the swept variable to be defined via a spreadsheet style format rather than in a linear fashion.

The list sweep function adds a great deal of flexibility to the basic sweep function. Instead of being forced to use constant sweep steps (either linear or logarithmic) as in the case of the standard sweep function, list sweep permits you to specify the points completely arbitrarily. One possible use of this feature would be a MOSFET subthreshold leakage measurement, where you could specify the initial values for the gate voltage sweep very close together and then spread them further apart once you have exited the subthreshold leakage region. You can create the vector list within applications such as MS Excel and cut-and-paste them into the vector data within EasyEXPERT. When used programmatically within an EasyEXPERT application test, you can also insert the name of a vector variable as the “source” for the list sweep.

Combining SMUs in series and parallel

It is possible to combine SMUs in both series and parallel to obtain more voltage and current than is possible using a single SMU. However, the methodology for doing this is not always simple or obvious.

Combining SMUs in series

Although SMUs are single-ended devices (one end is always tied to ground), by having one SMU track another SMU via the synchronized sweep function (VAR1'), it is possible to double the voltage applied to a DUT (just as if the SMUs were actually in series). The following figure illustrates how this can be achieved.

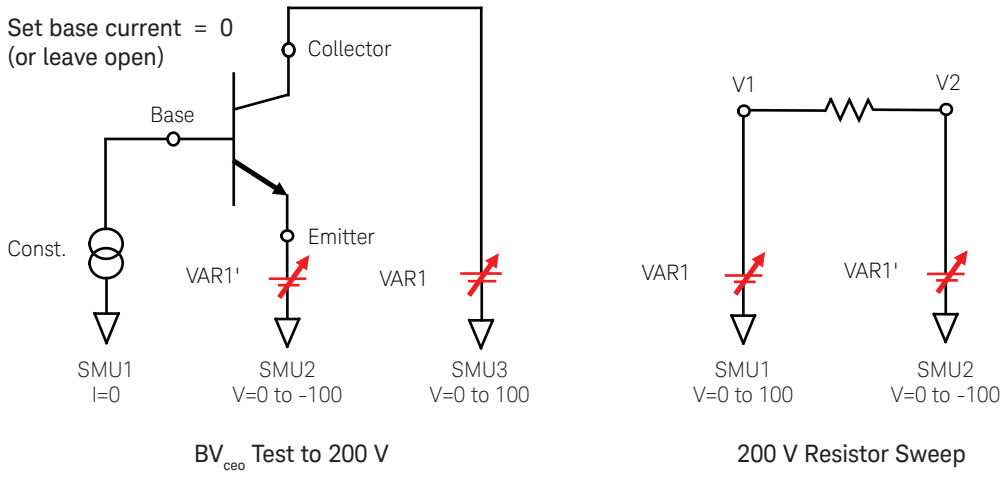


Figure 3.43. By using the VAR1' function SMUs can effectively be placed in series.

Although the above example is for medium power SMUs and high resolution SMUs (which both have a maximum voltage output of ±100 V), high-power SMUs (which have a maximum voltage output of ±200 V) can achieve differential voltages of up to 400 V.

Combining SMUs in parallel

Connecting SMUs in parallel in current force mode is relatively trivial. However, the usefulness of this procedure is limited. In most cases we want to force a voltage, and we want to place SMUs in parallel in order to improve the current sourcing capability of our voltage source as shown below.

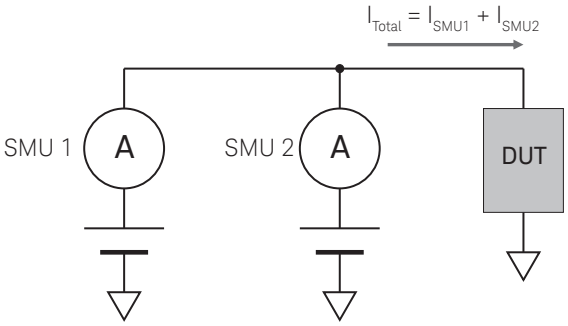


Figure 3.44. By placing two SMUs in voltage force mode in parallel, we can increase the total amount of current available.

Let us examine the general case of placing two SMUs in parallel in voltage force mode with a Kelvin configuration.

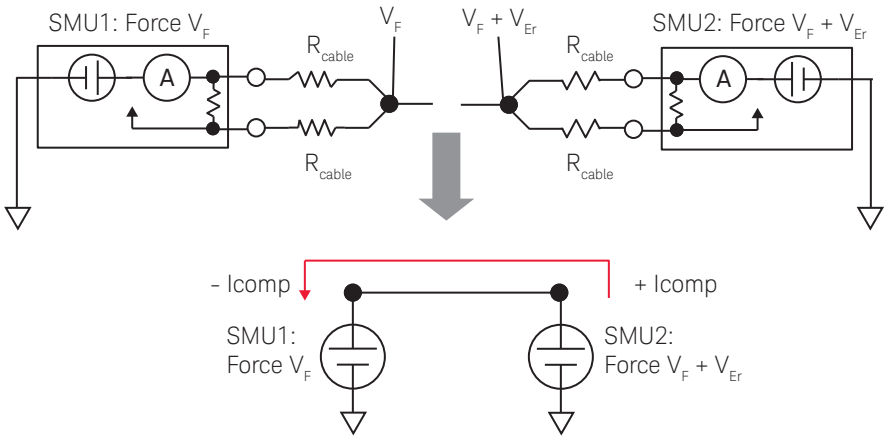


Figure 3.45. When placing SMUs in parallel, even a slight forcing voltage difference (error) can cause one or both SMUs to hit their current compliance limit.

The problem with this approach is that, even if you specify the exact same voltage for both SMUs, in practice there will be some voltage force error between the SMUs. This will cause one SMU to source current into the other SMU and very quickly one or both SMUs will hit their current compliance limit.

Combining SMUs in parallel *(continued)*

To prevent this situation, we can create a "quasi-Kelvin" configuration using two small resistors. These resistors limit the current flow to keep the SMUs from hitting compliance.

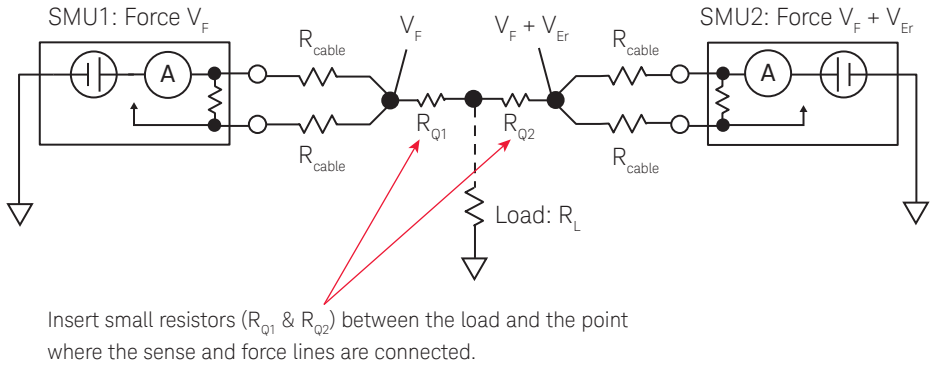


Figure 3.46. By inserting small resistors between the load and the point where the force and sense lines are connected, we can prevent the SMUs from hitting current compliance.

A worst-case scenario for this situation is two high-power SMUs placed in parallel each outputting 1 A of current. We know that the 20 V range is the maximum measurement range in which we can output 1 A of current, and that in this measurement range the maximum voltage force error is 10 mV.

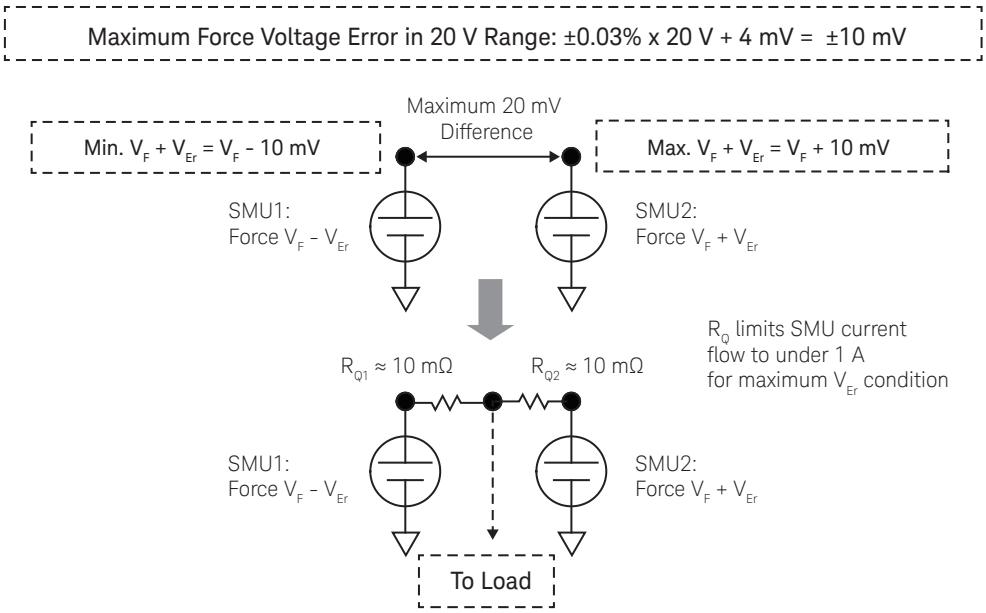


Figure 3.47. Calculating the required value of the resistors for this quasi-Kelvin connection from SMU specifications.

From these calculations, we can see that a 10 mΩ value for R_Q is sufficient to compensate for the 10 mV error for each SMU. In many cases a small piece of wire can be used to create the 10 mΩ resistor.

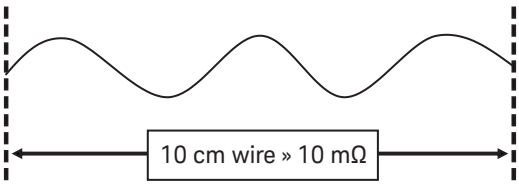
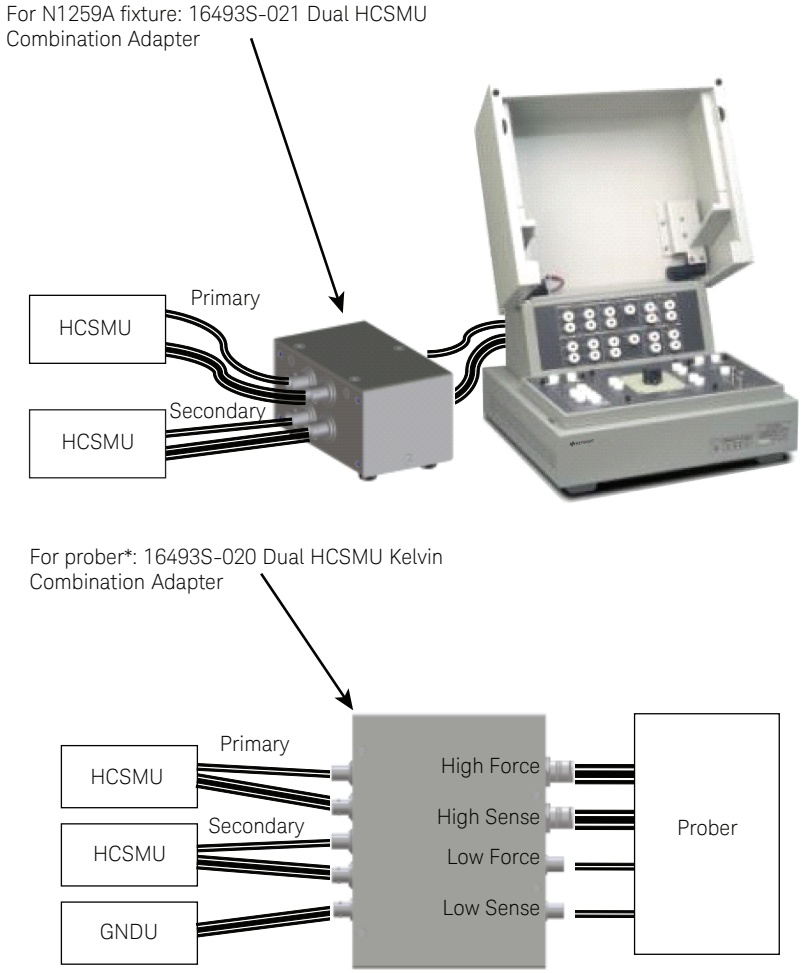


Figure 3.48. One simple way to create a 10 mΩ resistor is to use a small piece of wire.

Combining two HCSMUs in parallel

The B1505A supports accessories that allow you to combine two HCSMU modules in parallel to enable you to output 40 A of current from the mainframe. To use two HCSMUs with the N1259A test fixture, you need the 16493S-021 Dual HCSMU Combination Adapter. To use two HCSMUs with a wafer prober if you are not using the module selector (N1258A), you need the 16493S-020 Dual HCSMU Kelvin Combination adapter. To use two HCSMUs with a wafer prober if you are using the module selector, then you should use the 16493S-021. In addition to having one or both of these adapters, you also need to configure the dual combination adapter in EasyEXPERT. However, once this is done the software takes care of managing all of the issues involved with using the two HCSMU modules in parallel.



*If using the module selector (N1258A), use the 16493S-021.

Figure 3.49. Solutions for combining two HCSMUs in parallel using the N1259A test fixture or on a wafer prober.

Safety issues

Even medium power SMUs can output voltages and currents that can be lethal. Therefore, it is important to understand the precautions that must be observed to ensure a safe parametric measurement environment.

Interlock

To prevent accidental electrocution, all Keysight parametric instruments have an interlock connection located in the rear of the instrument. The interlock prevents the instrument from sourcing more than 42 V unless a safety interlock is in-place. The location of this interlock on the B1500A (and B1505A/B1506A) is shown below.

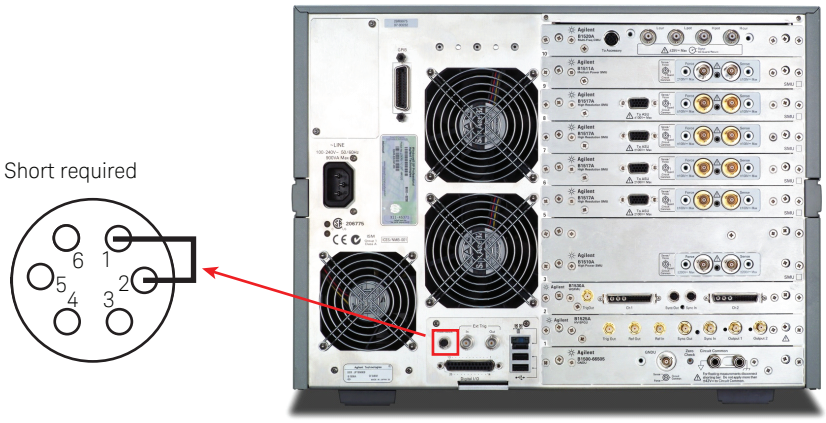


Figure 3.50. Rear view of the B1500A showing the location of the safety interlock.

In order for any of the SMUs to supply more than 42 V, a short between pins 1 and 2 must be made. **Do not attempt to defeat this scheme by shorting these pins together as an extreme safety hazard will be created.** The rear interlock should be used only with an approved interlock cable and equipment designed to work with an interlock cable.

Interlock *(continued)*

If you plan on using your parametric instrument with a semiautomatic wafer prober, then you should consult with your wafer prober supplier for the proper accessories to support the safety interlock. All major semi-automatic wafer prober companies can supply solutions that are compatible with this scheme. In addition, if you want to test packaged devices at up to 1 A and 200 V, Keysight Technologies can supply the 16442B test fixture for this purpose. The 16442B supports the Keysight 16493J interlock cable, which requires that the lid of the 16442B be closed in order for voltages greater than 42 V to be sourced. A picture of this test fixture is shown below.

Figure 3.51. The Keysight 16442B Test Fixture supports a variety of socket types for testing packaged devices and also has a built-in interlock.

Keysight has two test fixtures available for use with the B1505A Power Device Analyzer/Curve Tracer. The N1259A test fixture supports component testing at up to 30 A and 3000 V.

Figure 3.52. The Keysight N1259A High-Power Test Fixture provides a safe and supported solution for testing power devices at up to 30 A and 3000 V.

For users requiring currents up to 1500 A and voltages up to 10 kV, the N1265A ultra-high current expander/ test fixture is available.

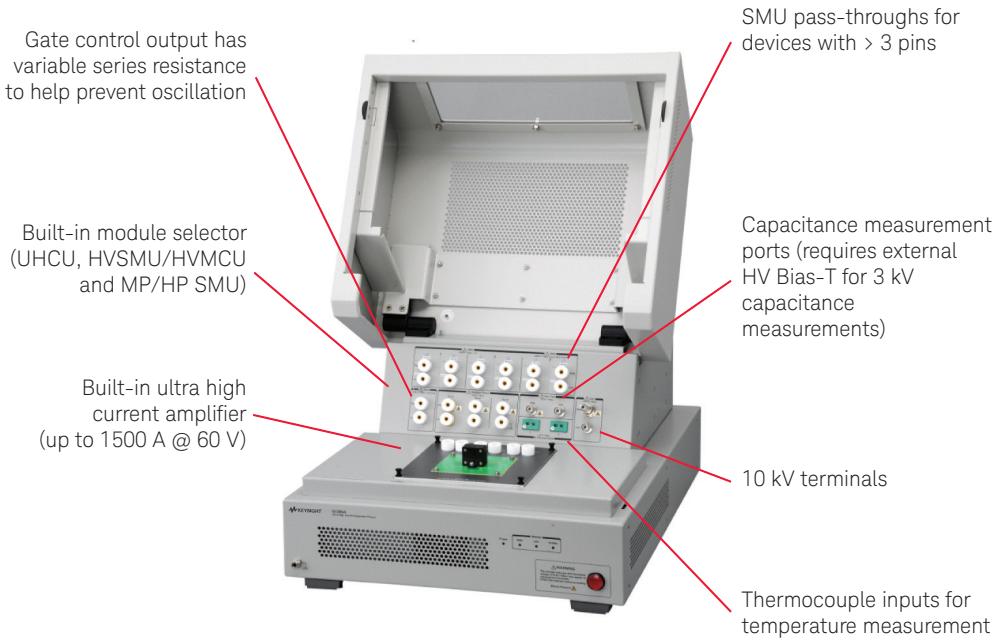


Figure 3.53. The Keysight N1265A Ultra-High Current Expander/Test Fixture can support the safe testing of power devices up to 1500 A and 10 kV.

Since the N1265A includes either a 500 A or a 1500 A ultra-high current unit (UHCU), it requires two control SMUs in order to use it. These SMUs can be either MCSMUs or HCSMUs. Also, another MCSMU or HCSMU is required for the gate control input. If an HPSMU or MPSMU is available in the B1505A for low-level precision measurement, or if an HVSMU is available in the B1505A, then they can be connected to the integrated module selector unit (explained in the next section). An illustration of a typical connection scheme using the N1265A is shown in the following figure.

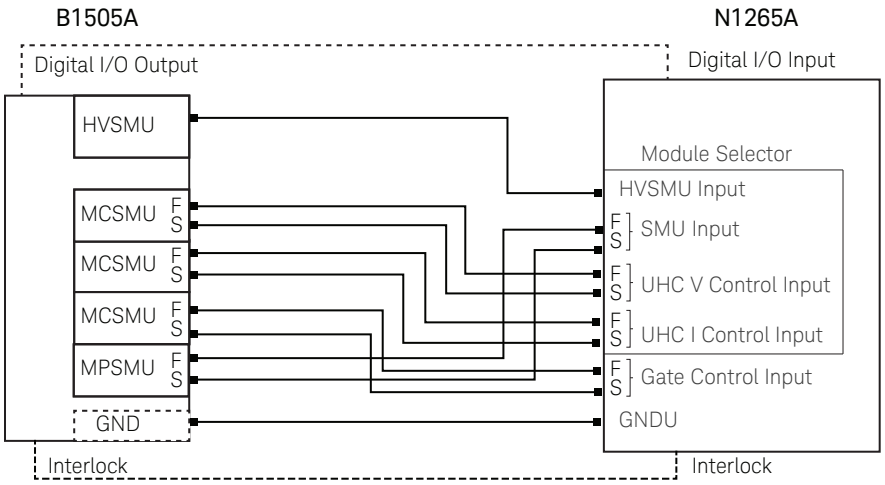


Figure 3.54. An example connection scheme between the B1505A and the N1265A.

Both the N1259A and N1265A test fixtures have a modular configuration that supports a variety of different package types. This includes universal and Teflon board socket modules that can be used to create solutions for custom package types, and adapters that support legacy Tektronix curve tracer test sockets. Both test fixtures also have a safety interlock that prevents measurements at dangerous voltage levels unless the lid is closed. In addition, optional resistor boxes (R-boxes) are available to work with the test fixture to ensure that lower-power modules (such as an MPSMU or HPSMU) are not inadvertently damaged.

Additional Modules and Accessories for Power Device Test

Module selector unit

The HVSMU module has a high-voltage triaxial output, and the HCSMU module has a unique configuration consisting of triaxial and BNC outputs (which was explained earlier in this chapter). Standard SMUs (MPSMU and HPSMU) of course have two standard triaxial outputs (force and sense). The incompatibility between these three module types makes it tedious to switch between high-voltage, high-current and low-level measurements due to all of the recabling involved. To solve this issue and make it easy to switch between these various modules, Keysight created the module selector unit. A schematic of the module selector unit is shown below.

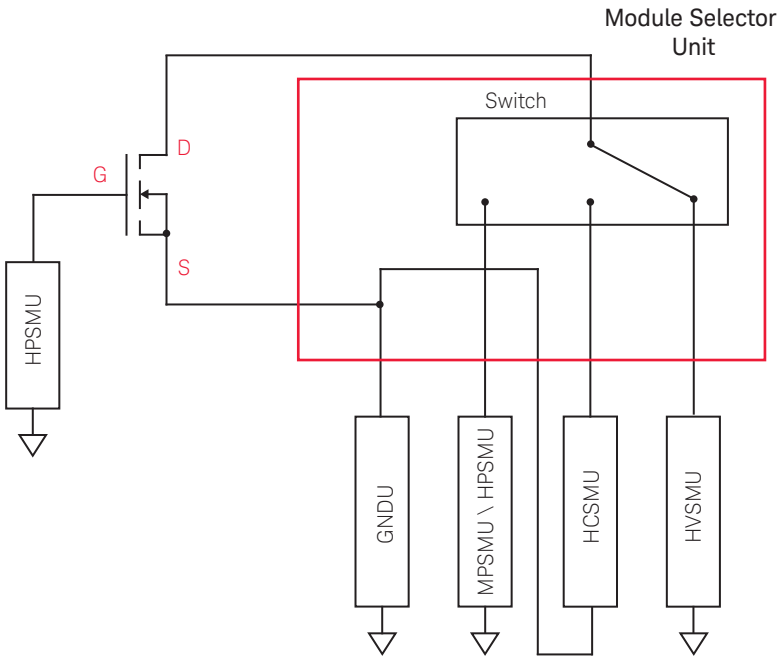


Figure 3.55. The module selector unit allows you to switch between standard, high-voltage and high-current measurements without having to change any cabling.

The module selector unit is connected to the B1505A via its digital I/O interface, and the B1505A's EasyEXPERT software automatically takes care of connecting the correct measurement resources.

The module selector unit for the N1259A test fixture is not a standard feature, but it is available as an option (N1259A-300). However, for the N1259A test fixture, there is no way to use the built-in test fixture for wafer probing, so a separate module selector unit (N1258A) is required if you want to perform both packaged device and on-wafer measurements. The wafer prober version of the module selector unit has two high-voltage triaxial outputs (force and sense high) and two BNC outputs (force and sense low). The correct cabling and adapters necessary to support the N1258A on a high-power wafer prober will be discussed when we talk about on-wafer parametric measurement in chapter 4.

In the case of the N1265A test fixture/high current expander, the module selector unit is included automatically. The B1505A's ultra-high current unit (UHCU) is integrated into the N1265A Ultra-High Current Expander / Test Fixture. The N1265A module selector unit can support current measurements up to 1500 A. The N1265A also has series resistors built in to its gate control path, and these can be configured via software.

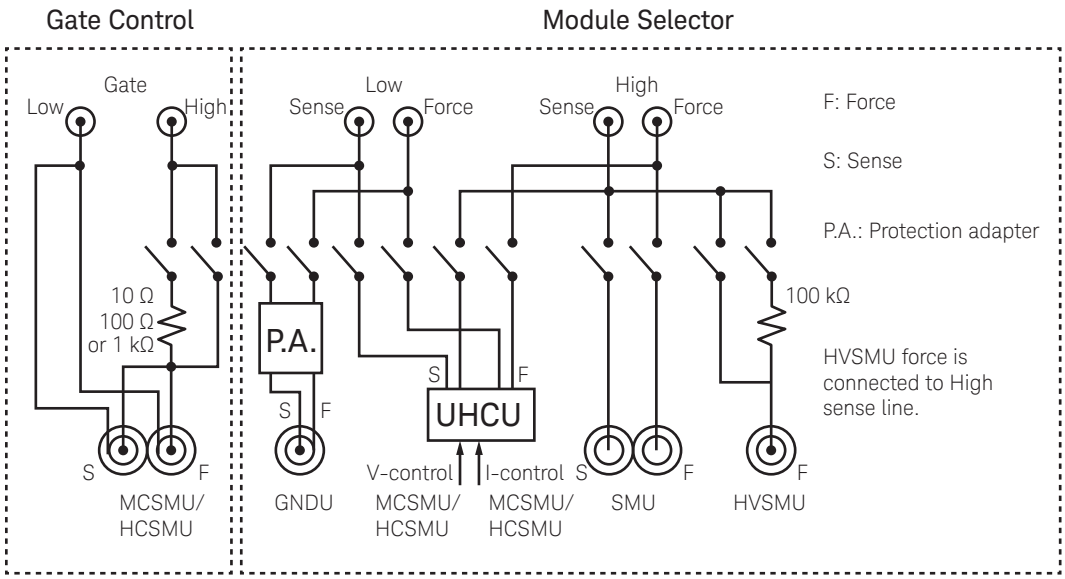


Figure 3.56. The N1265A module selector unit and gate control path support seamless switching between ultra-high current (up to 1500 A), high voltage (up to 3 kV) and a precision measurement resource (MPSMU or HPSMU).

As this figure indicates, the N1265A's ultra-high current (UHC) expander requires two SMUs to control it (either MCSMUs or HCSMUs). Note: Unlike the N1259A, the N1265A's module selector unit can be used for both packaged device and on-wafer measurements (to be discussed in chapter 4).

High-voltage Bias-T

In chapter 9, when we talk about power device testing, the theory and usage of the high-voltage Bias-T will be discussed in detail. This section will explain how the HV Bias-T works for the N1259A and N1265A. For the N1259A, the HV Bias-T is an available option (N1259A-020) that is integrated into the test fixture.

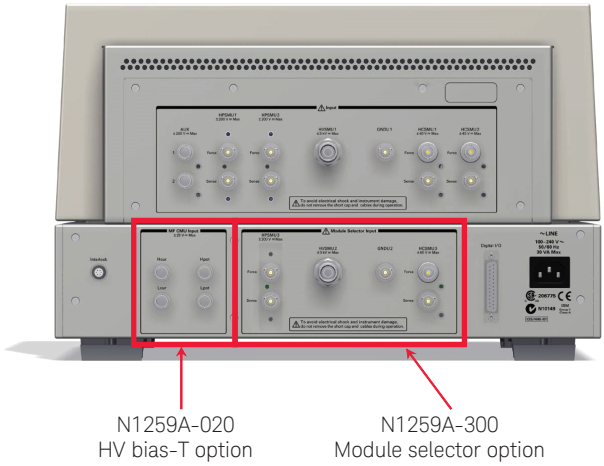


Figure 3.57 – Rear illustration of the N1259A high power test fixture with module selector unit and HV Bias-T options.

The N1259A's HV Bias-T cannot be used for wafer probing, so a separate part (N1260A) is required to perform high-voltage capacitance measurements on wafer. For the N1265A test fixture/high current expander, there is no integrated HV Bias-T option. Instead, the N1265A uses the same external Bias-T (N1260A) that is used for on-wafer measurements.

High-voltage SMU current expander

The N1266A high-voltage SMU current expander enables the HVSMU to output up to 2.5 A. The N1266A is a separate mainframe, and when combined with a HVSMU and two control MCSMUs, it forms the high-voltage medium current unit (HVMCU). The HVMCU can provide simultaneous high-voltage and high-current output for applications such as safe operating area (SOA) test and breakdown test where the breakdown currents are large. The HVMCU can only operate in pulsed mode, and the supported pulse widths depend upon the output range as shown in the following table.

Output Range	Supported Pulse Widths	Resolution
1500 V/2.5 A	10 µs - 100 µs	2 µs
2200 V/1.1 A	10 µs - 100 µs	2 µs
2200 V/110 mA	10 µs - 1 ms	2 µs

Figure 3.58 – The allowable HVMCU pulse widths and resolution for each of its measurement ranges.

Ultra-high voltage unit (10 kV)

The N1268A ultra-high voltage unit (UHVU) is available to test devices at up to 10 kV. The UHVU is a separate mainframe, and it requires two control SMUs (either MCSMUs or HCSMUs) in order to use it. The N1268A can operate in both continuous and pulsed mode, and the maximum output current for these two cases is shown in the following table.

Voltage	Max. Current (DC and Pulsed Mode)	Max. Current (Pulsed Mode Only)
+10 kV	+10 mA	+20 mA
–10 kV	–10 mA	–20 mA

Figure 3.59 – The UHVU maximum current output for DC and pulsed modes.

The N1268A can be used with either the N1265A test fixture or a wafer prober, but special ultra-high voltage triaxial cables are required in both cases.

Summary of B1505A IV options

Using the options just described, it is possible to create a very comprehensive power device testing solution using the B1505A. The following figure shows a solution capable of testing power devices at up to 10 kV and 1500 A, and with the ability to test SOA at high-voltage and medium current.

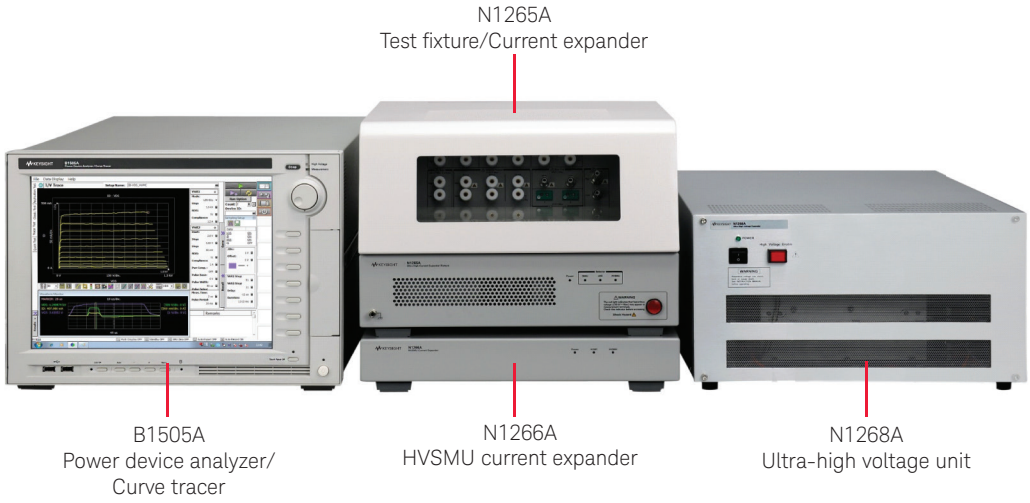


Figure 3.60. The B1505A power device analyzer/curve tracer, N1265A test fixture/current expander, N12566A high-voltage current expander and N1268A ultra-high voltage unit.

Note: This solution assumes that the N1266A and N1268A will share two control MCSMUs. Since the applications using these solutions are quite different, the time needed to recable and reconfigure the modules in the software is not a big issue.

Chapter 4 On-Wafer Parametric Measurement

"One must learn by doing the thing; for though you think you know it, you have no certainty until you try."
— Sophocles

Introduction

The vast majority of parametric measurements (at least low-power ones) are performed on-wafer. This implies that you need some sort of a wafer prober to make these measurements.

Many engineers performing on-wafer parametric measurements often miss taking into account factors such as the wafer chuck, cables, wafer probes and switching matrix (if used), all of which can have profound effects on their measurement results. Note: The first part of this chapter is focused on traditional low-power (V<=200 V, I<=1 A) parametric measurements, and these power limits are expressed implicitly in the following discussions. The last section covers the issues involved with the probing of power semiconductor devices on-wafer.

Fully automatic versus analytical wafer probers

Fully automatic wafer probers are designed to be used in production environments in conjunction with either parametric or functional test equipment. Fully automatic wafer probers can test an entire cassette or FOUP (front-opening unified pod) of wafers at a time, and they also have the ability to automatically load and align a wafer for testing. Fully automatic wafer probers are virtually always used with some sort of a probe card and switching matrix. Due to their sizes, costs and complexities, fully automatic wafer probers are used almost exclusively in production test environments with high-volume production testers. Since this handbook is primarily focused on parametric test instruments, fully automatic wafer probers will not be included in the discussions.

Analytical wafer probers are designed to be used in laboratory environments, and they come in both manual and semi-automatic versions. Manual wafer probers do not possess any electrical motors to automate any part of the wafer alignment and stepping process, so they must instead (as the name implies) be manually adjusted each time you want to move to a new location on the wafer. In contrast, most semi-automatic wafer probers have the capability to automatically align an individual wafer. However, they typically do not support any sort of automated wafer loading and they require manual loading of each wafer. Semi-automatic wafer probers can be used with probe cards and switching matrices, but they are more commonly seen used with multiple individual positioners (probes) for maximum measurement flexibility. Most semi-automatic wafer probers also support automated testing across an entire wafer in conjunction with some type of parametric instrumentation.

Figure 4.1. Semi-automatic wafer probers.

Wafer prober measurement concerns

It should be clear upon reflection that the wafer prober measurement environment has a significant impact on the quality of parametric measurements that you can obtain. Unfortunately, these effects are often not adequately accounted for by users trying to make parametric measurements. Some of the key issues are as follows.

1. Current noise floor – An inadequately shielded wafer prober will greatly increase the noise floor within a measurement system.
2. Slow measurements – Capacitance and piezoelectric effects can significantly alter the speed at which parametric measurements can be made.
3. Low and high temperature measurements – Low and high temperature measurements both present issues in terms of moisture and noise.

Chuck isolation

The basic rules for proper guarding and shielding discussed in chapter 2 take on great significance when applied to the wafer probing environment. It is important to understand that a wafer chuck is an extremely large capacitor and that it can act as a large antenna (collecting random noise). Therefore, optimal measurement performance is obtained using a guarded Kelvin wafer chuck as shown below.

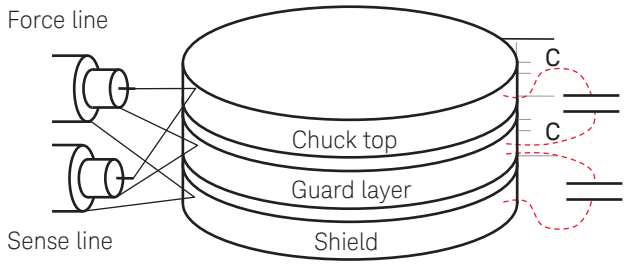


Figure 4.2. A guarded Kelvin wafer chuck can eliminate the effects of the parasitic chuck capacitance and greatly reduce the leakage from the chuck to the outside environment.

By using a guard layer in the wafer chuck, the effects of the parasitic chuck capacitance and the leakage currents through the wafer chuck can be virtually eliminated. The following plot compares the settling time for both guarded and unguarded wafer chucks, and it is clear that for the case of the guarded chuck, the waiting time before measurement can begin is greatly reduced.

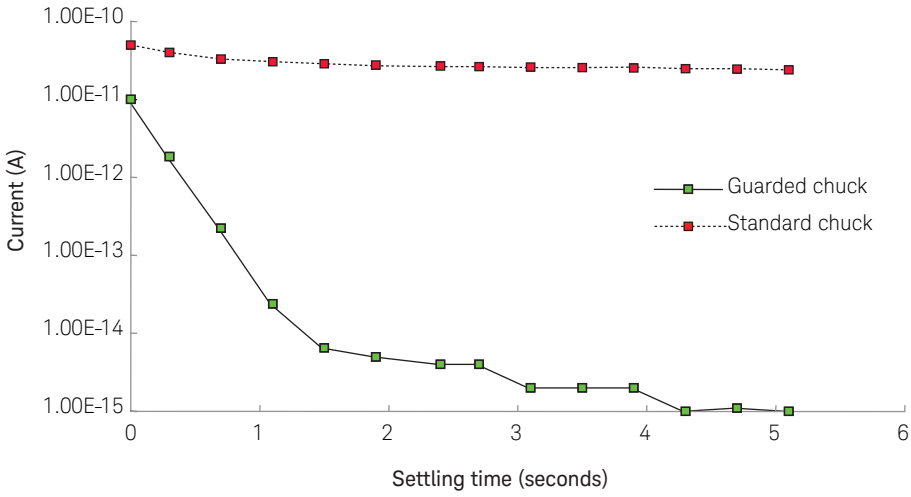


Figure 4.3. Plot comparing the settling time for both guarded and unguarded (standard) wafer chucks.

Chuck isolation *(continued)*

In addition to guarding the wafer chuck, if the guard can also be placed above the wafer being measured then noise can be reduced further and an optimal low-noise measurement environment is guaranteed. Such a measurement scheme is shown below.

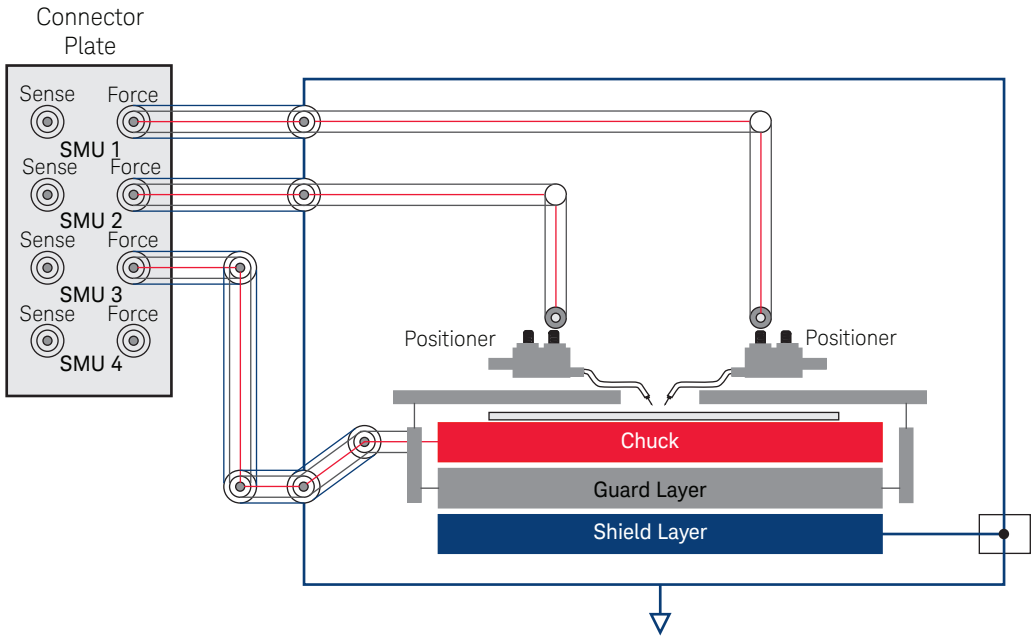


Figure 4.4. A completely guarded and shielded wafer prober environment.
Note: Scheme shown is FormFactor's patented AttoGuard technology

Low and high temperature measurement issues

Performing parametric measurements at low and high temperatures on-wafer presents the following measurement challenges.

1. High electrical noise caused by the thermal control circuitry.
2. Slow measurement times due to parasitic capacitance.
3. Large transient noise after wafer chuck moves.
4. Frost induced moisture leakage.

We will proceed to examine each one of these in-turn.

The electrical noise generated by the thermal control circuitry is probably the single greatest source of error when making low-level measurements on-wafer at low and high temperatures. While it is not possible to completely eliminate all of the effects of this circuitry, many analytical wafer prober companies have mitigated them to a greater extent. The graph on the following page shows the noise current associated with a standard triaxial wafer chuck as the chuck temperature varies over time.

Low and high temperature measurement issues *(continued)*

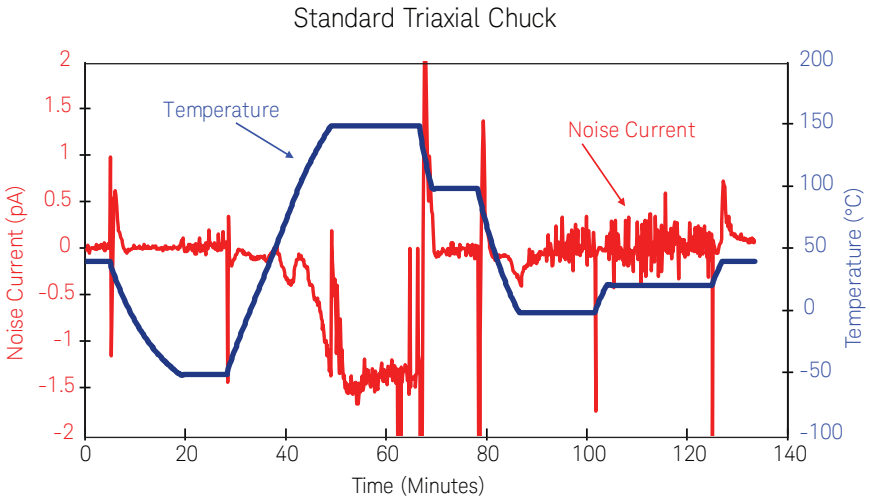


Figure 4.5. Noise current for a standard triaxial wafer chuck as the temperature varies over time.

In contrast, the following graph shows the noise current for a well-designed thermal wafer chuck.

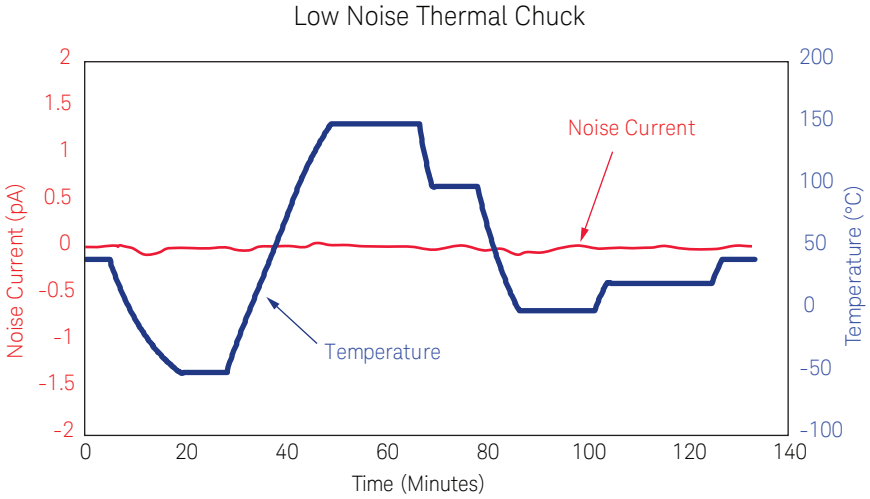


Figure 4.6. Noise current for a well-designed low-noise thermal wafer chuck.

In the end, you get what you pay for, so it does not make sense to try and make low-level measurements at high or low temperatures on anything other than a wafer chuck specifically designed for that purpose. Before purchasing any analytical wafer prober with a thermal chuck option, you should very carefully evaluate its low-current noise performance to make sure that it meets the requirements of your parametric measurements.

Low and high temperature measurement issues *(continued)*

In addition to reducing noise on the wafer chuck, it is important to reduce noise and capacitance effects caused by the probes. Ceramic bladed probe tips are the best solutions for making on-wafer measurements at high temperatures. Bladed probe tips can easily handle temperatures of up to 300°C, and they have minimal residual capacitance. A picture of a ceramic bladed probe tip for positioner-based wafer probing is shown below.

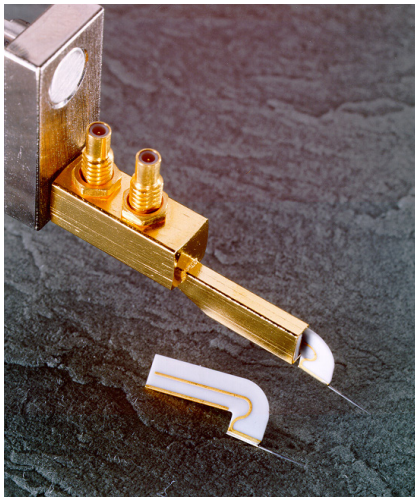


Figure 4.7. Ceramic bladed probe tips are the best solutions for on-wafer high-temperature measurement.
Note: Photo courtesy of FormFactor

Note that this type of probe tips can be easily replaced if they are damaged.

The same type of methodology should be implemented when using probe card schemes (which will be discussed later in this chapter). However, in the case of probe cards, the juxtaposition of many signal lines close together adds to the design complexity. The key is to shield the entire probe card within a metal enclosure to isolate the signal lines as much as possible from thermal noise. An example of a fully guarded ceramic probe card for high temperature measurement is shown below.

Figure 4.8. A ceramic bladed probe card with full shielding that works from –65°C to +300°C.
Note: This is a picture of FormFactor's Attofast probe card.

Low and high temperature measurement issues *(continued)*

On a semiautomatic wafer prober, the electronics and cable connections used to move the wafer chuck and to control its temperature have a strong impact on low-noise measurements, and these are exacerbated at higher temperatures. There is no single "magic bullet" that can completely eliminate these noise sources, but sound design principles can mitigate them. Some key points to look for in a wafer prober are listed below.

1. Low triboelectric effect materials to minimize the electronic charge generated via friction.
2. Low dielectric absorption materials to minimize the residual capacitive charge.
3. Low noise design through proper shielding.

The final issue to consider when making on-wafer low temperature parametric measurements is frost, which is related to the dew point. Moisture degrades parametric measurement significantly, and can cause high residual capacitance, elevated leakage currents, and increased noise. For this reason, it is often recommended to "bake" a wafer chuck at +200°C for twenty four hours prior to making any parametric measurements at low temperatures. In any case, low temperature measurements require an atmospheric flow (using either clean dry air or nitrogen) to prevent frost from forming. However, as the graph shown below indicates, increase in air flow also causes noise to increase.

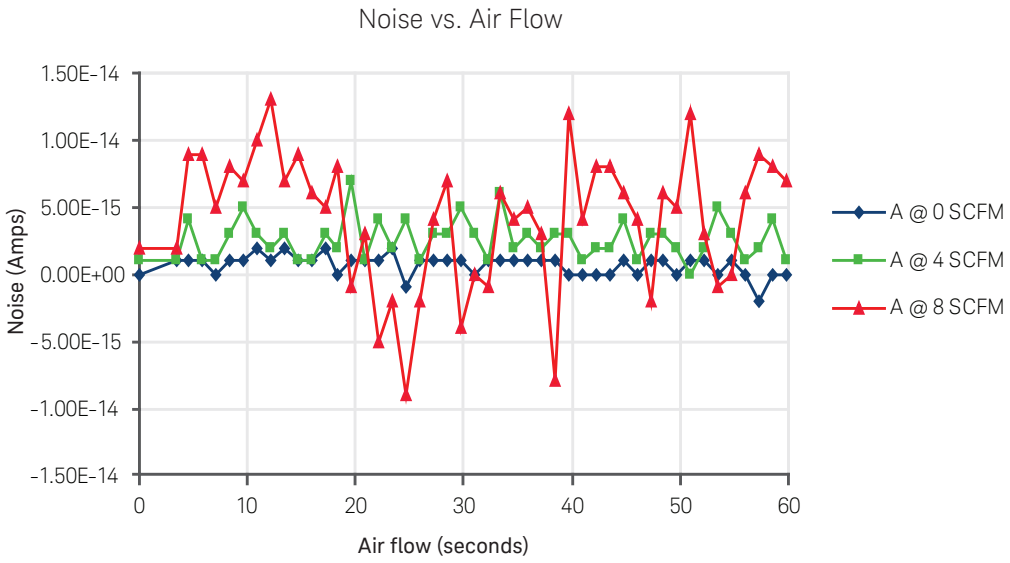


Figure 4.9. The effect of air flow on noise.

Therefore, the best probing systems are those that can achieve (for example) frost-free operation at –55°C with minimal air flow.

Note: An SCFM flow of less than 1 is probably sufficient for most measurements.

DC and RF wafer probes

For analytical wafer probers, there are two distinct types of positioners: DC and RF. Although many people might assume that RF probes are only required for true RF measurements made in the GHz range (such as S-parameter extraction), this belief is not true. Capacitance measurements made above 5 MHz may require RF probes in order to achieve satisfactory measurement results, and many high-speed pulsed measurements require RF probes to avoid ringing on the pulse edges. Therefore, it is important to understand the differences and limitations of both DC and RF probes.

There are several key points to keep in mind when designing test structures for RF wafer probes.

1. RF wafer probes are typically mounted in fixed positions on the wafer prober (180 degrees apart, opposite to each other), so any structure to be placed on the wafer must adhere to this limitation.
2. RF wafer probes primarily come in ground-signal (GS) or ground-signal-ground (GSG) configurations, and they are available in only certain pad pitches (typically ranging from 50 µm to 250 µm).
3. The probes themselves should be separated by at least 200 µm to avoid crosstalk.
4. All grounds should be connected together.
5. Any biasing or control pads should be placed completely out of the main signal path.
6. RF probes require BNC cabling and usually use an SMA style connector.

A generic example of a test structure adhering to these guidelines is shown below.

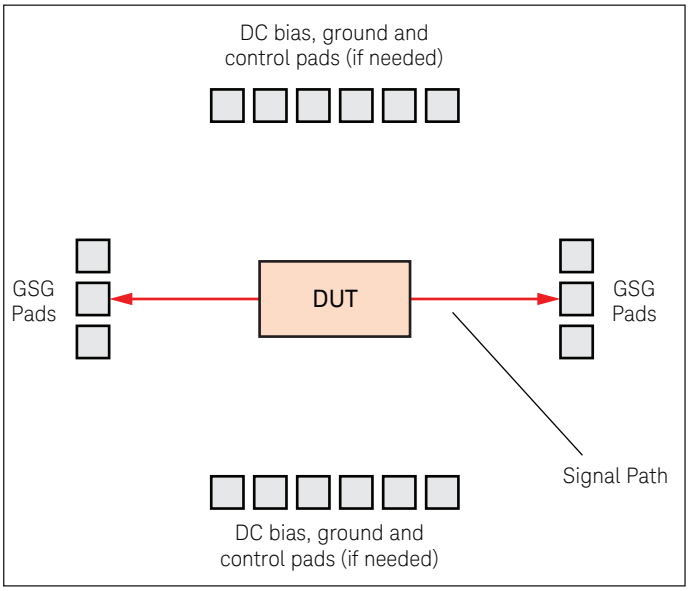


Figure 4.10. The proper way to layout a structure for use with RF positioners.

DC and RF wafer probes *(continued)*

A picture of RF probes used for on-wafer measurement is shown below.

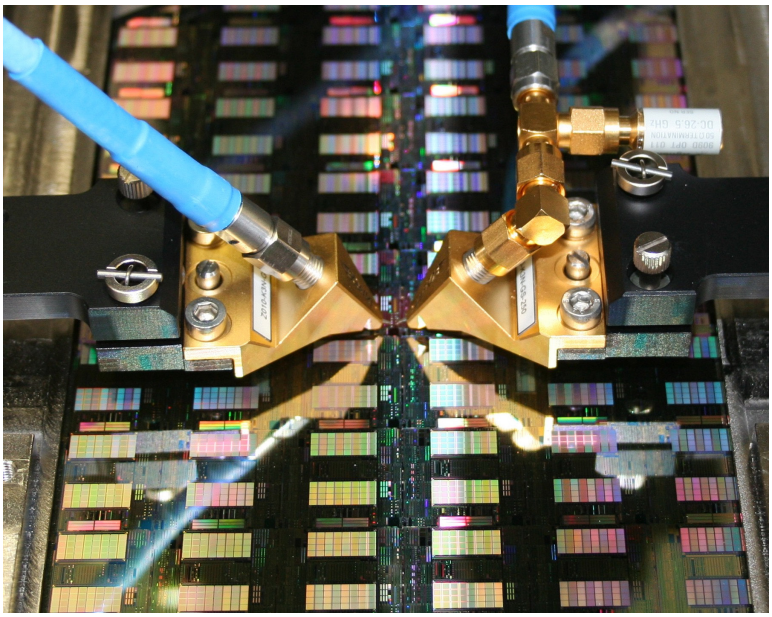


Figure 4.11. An example of on-wafer measurement using RF wafer probes.
Note: Photo courtesy of Form Factor

RF positioner device layout and design will be discussed further when we cover high-speed pulsed measurement in chapter 5 and high-frequency capacitance measurement in chapter 8. Note that the RF probes use SMA connectors, so BNC to SMA connector adapters will be required if you are trying to connect up instruments such as capacitance meters.

Switching matrices

Introduction

Although switching matrices do not necessarily have to be used in conjunction with on-wafer semiconductor measurements, it is very rare to see them used outside of this type of measurement environment. When making on-wafer measurements, a switching matrix can both facilitate faster measurement and simplify the connection environment (for example switching between IV and CV measurements). A photo of a switching matrix for use in parametric testing is shown below.

Figure 4.12. The Keysight B2200A femtoamp low leakage switch.

The basic concept of a switching matrix is not hard to understand. A switching matrix consists of a series of inputs and outputs; connections between the inputs and outputs can be made by closing one or more relays. Since virtually all connections in parametric tests are made with triaxial cables, almost all switching matrices convert some number of coaxial inputs into triaxial outputs. While the number of inputs on a switching matrix is fixed, the number of outputs is usually determined by the number of cards installed in that matrix mainframe. Each card will add some number of outputs to the configuration. An example of a switching matrix card is shown Figure 4.13.

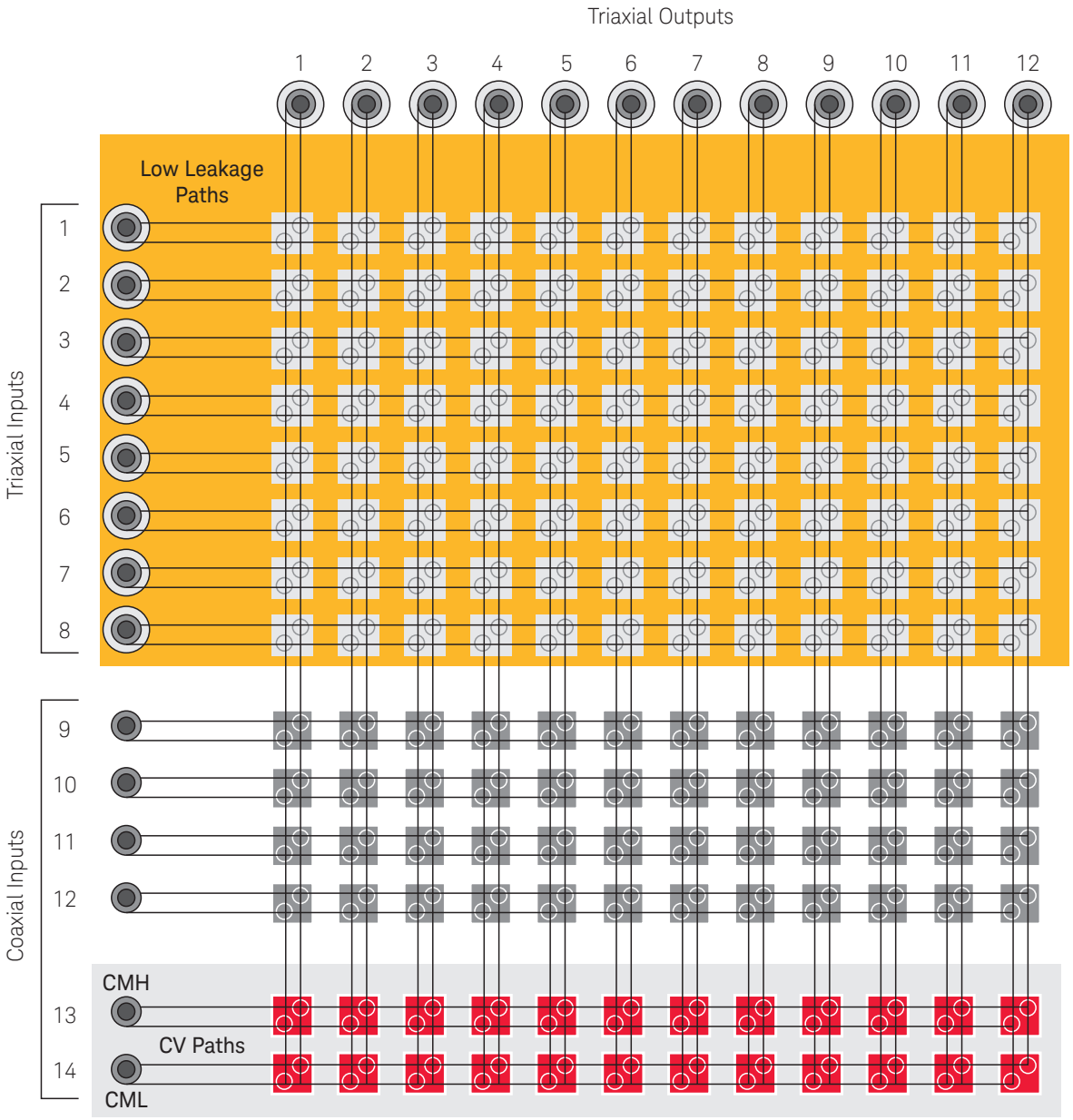


Figure 4.13. An example of a switching matrix card with both triaxial and coaxial inputs (as well as a dedicated path for CV measurement).

Introduction *(continued)*

Here are some key takeaways from the schematic just shown. The first of these is that not all triaxial switching matrix paths are necessarily “low-current”. Although this is true in the above example, it is not always the case for other switching matrix cards. The second observation is that a switching matrix converts all inputs (triaxial and BNC) into triaxial outputs. This is extremely convenient since it eliminates the need to worry about triaxial to BNC adapters. The third (and final) observation is that some matrix cards have a dedicated path for capacitance measurement. However, we will defer the discussion of making CV measurements through a switching matrix until we reach chapter 8.

Probe cards and module testing

In order to understand the benefits of a switching matrix when making on-wafer measurements, one must first know how parametric test structures are organized on the wafer. Parametric test structures are typically organized into modules with some sort of regular pad organization (such as two rows of twelve pads). When used in conjunction with a probe card designed to fit the module pad configuration, a switching matrix permits the testing of all of the devices in that module without the need to physically move the probes. By using the switching matrix to connect and disconnect the DUTs from the parametric measurement resources, the time associated with physically moving the probes from one device to the next is eliminated. The following figure shows a photo of probe card tips in contact with a test module on a wafer.

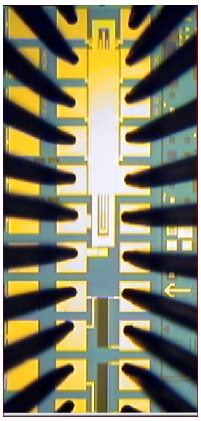


Figure 4.14. A photo showing a probe card in contact with a module on-wafer.

Even in a semi-automatic wafer prober environment, the use of a probe card can significantly improve test throughput.

Introduction *(continued)*

It should be obvious that in order to take full advantage of a switching matrix and probe card solution, some sort of software test shell is necessary to control everything. The key requirements of this software are shown below.

1. The ability to coordinate testing with the wafer map stored on the semi-automatic wafer prober.
2. The ability to specify sub-die moves so that different modules within a die can be tested.
3. The ability to store and recall switching matrix settings to connect to a particular device to be tested.
4. The ability to store measurement results for each device tested along with information as to the exact location (both die and sub-die) of that device on the wafer.

Keysight Technologies' EasyEXPERT software possesses all of these capabilities and can be used to automate the testing across a semi-automatic wafer prober.

In order to perform automated wafer testing using Keysight EasyEXPERT, you must first set up a wafer map on the semi-automatic analytical wafer prober that you plan on using. An example of this is shown below.

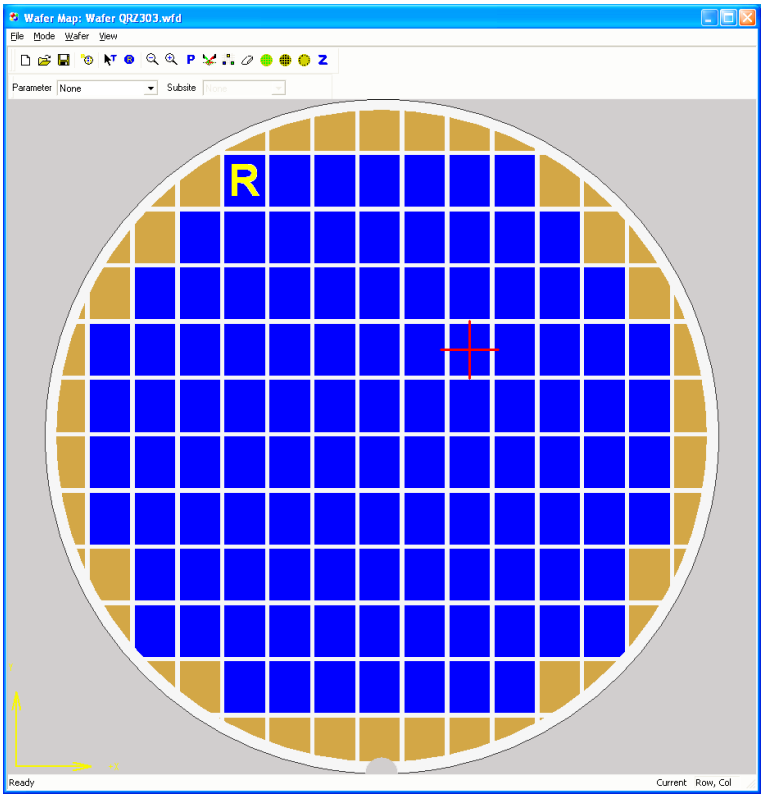


Figure 4.15. Setting up a wafer map on a semi-automatic wafer prober in preparation for automated wafer probing.

Introduction *(continued)*

Once the wafer map is set up, you can start the automated testing by calling up the “Repeat Measurement Setup” window by clicking on the repeat button in the upper right corner of the main EasyEXPERT screen.

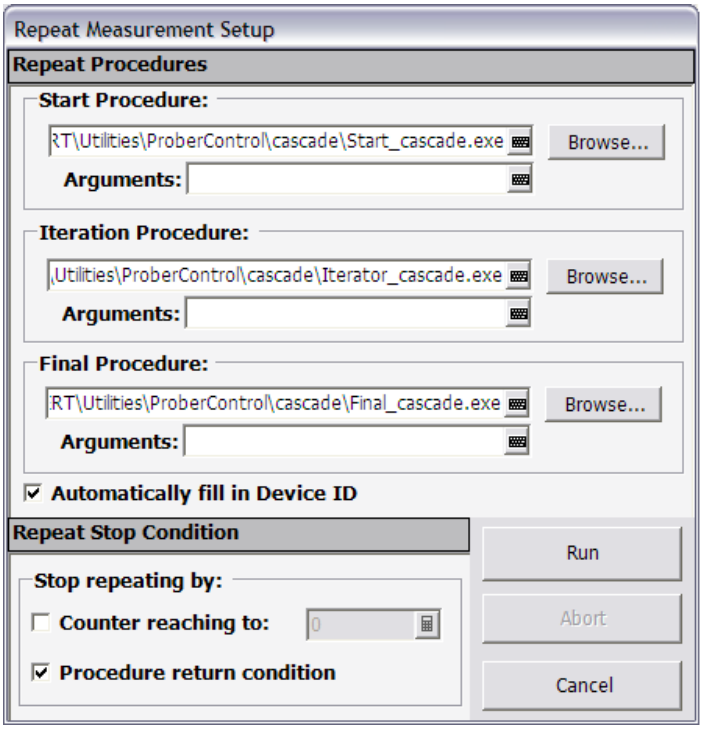


Figure 4.16. The Repeat Measurement Setup screen allows you to set up the conditions to automatically step and repeat one or more measurements across a wafer.

Note: You must have the data auto-record feature turned on in order to use this feature.

EasyEXPERT has built-in wafer prober drivers for FormFactor (also known as Cascade Microtech), SUSS, Signatone and Vector semi-automatic wafer probers. In addition, sample source code is available so that you can create your own wafer prober drivers if necessary. The prober driver files can be found in the following location: *Program Files\Agilent\B1500A\EasyEXPERT\Utilities\Prober Control*.

Each supported prober has a file folder, and there are several executable files located in these file folders. In the “Repeat Procedures” section of the “Repeat Measurement Setup” window, the correct executable file can be selected using the “Browse” button. For each of the three lines, Start, Iteration and Final, there is a corresponding executable file (as can be seen in figure 4.16). By default, the “Automatically fill in Device ID” box is checked, which will automatically create a device ID label for each EasyEXPERT test record to keep track of which test or tests are associated with which die on the wafer. In addition, you can select a “Repeat Stop Condition” if the “Procedure return condition” is checked then the testing will continue until the wafer map reaches the last die; if the “Counter reaching to:” condition is checked, then the testing will continue until the specified number of die has been tested. If both conditions are checked, then testing will stop at whichever condition is reached first. After filling in the above information, simply click on the “Run” button to begin coordinated testing of your Keysight parametric instruments with your semi-automatic wafer prober.

Reliability Test Switching Solutions

Certain types of reliability test (such as HCI and TDDB) require the stressing of tens or even hundreds of devices or structures in parallel. In addition, these types of tests require constant voltage biases that are best supplied by a power supply (rather than an SMU). Conventional crosspoint switching matrix cards do not work well for these types of reliability testing. However, Keysight has a solution tailored to address this need.

The E5250A Low-Leakage Switch Mainframe supports the E5255A 24 (8x3) Channel Multiplexer cards, which can support the testing of hundreds of devices in parallel for very minimal cost. The E5250A mainframe can contain up to four of the E5255A cards, and each card has 24 outputs (for a total of up to 96 channels per mainframe). In addition, you can gang up to four mainframes together for a total of up to 384 channels. The E5255A cards also allow you to use standard power supplies to provide the stress voltage biases during test, and you can connect up to six SMUs (using triaxial cabling) to measure the DUTs during different phases of the stress testing. A picture of the E5250A with the E5255A cards installed is shown below.

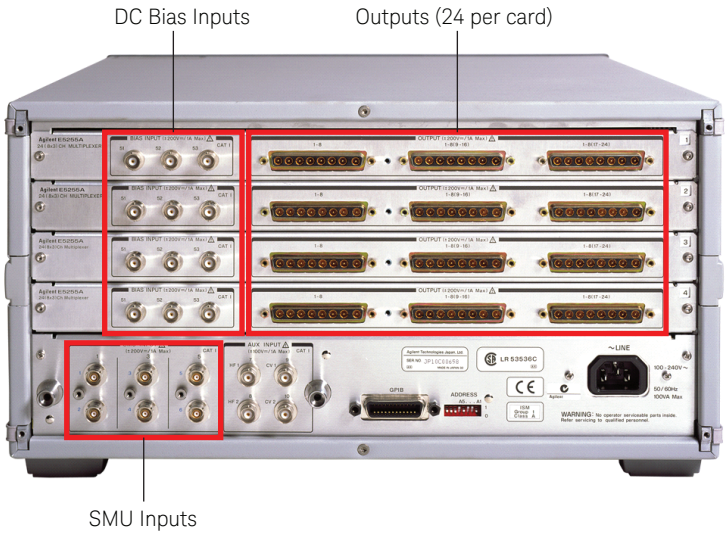


Figure 4.17. The E5250A low leakage switch mainframe shown with four E5255A multiplexer cards installed.

To understand how this solution works, please refer to the block diagram below that shows the connections on one of the three blocks that are present on each E5255A card.

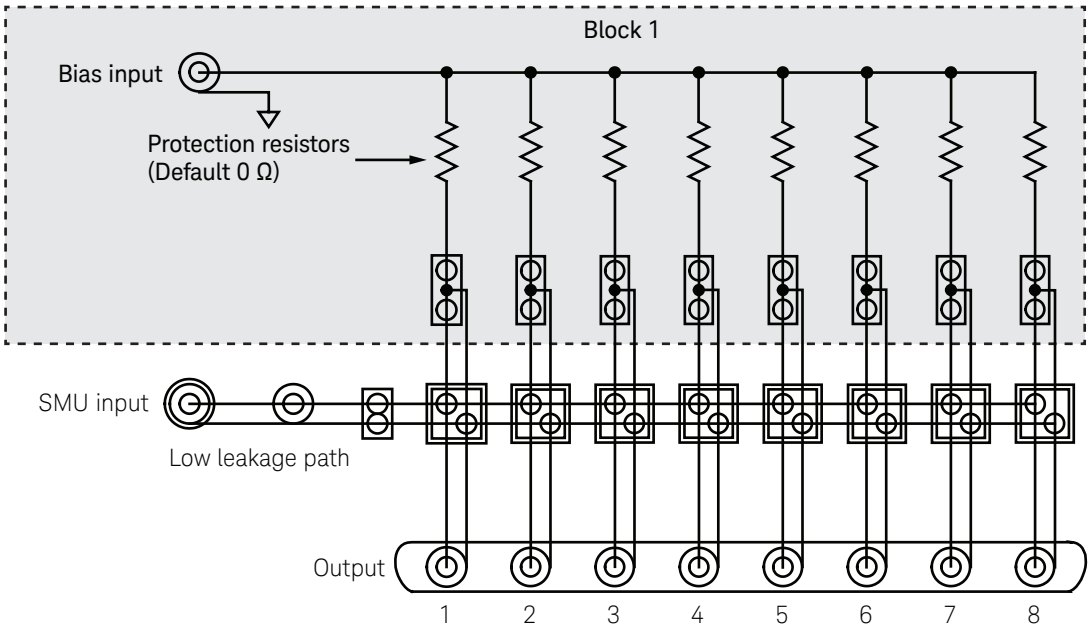
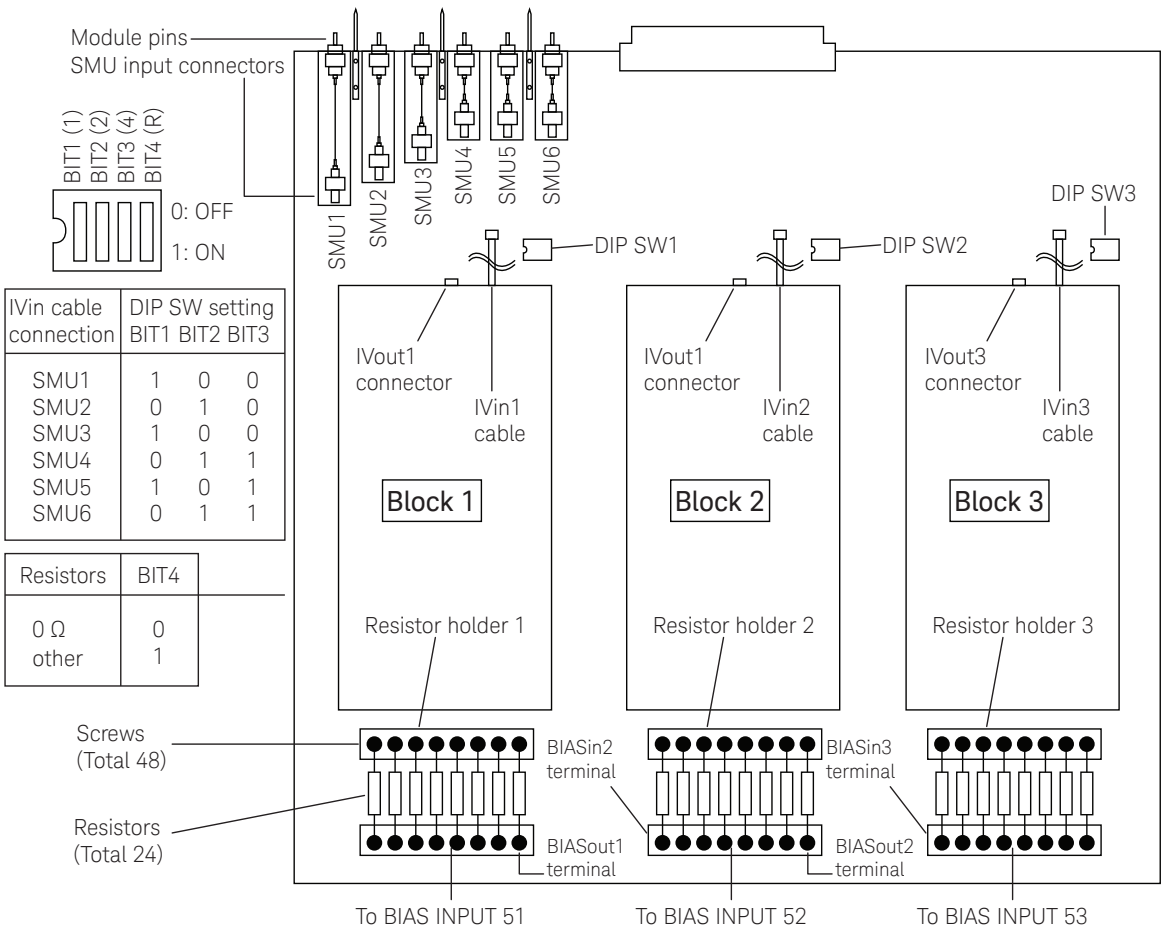


Figure 4.18. One of the three independent blocks on each E5255A card.

As this figure shows, each block of eight outputs is tied to the same DC bias source (BNC connection) through protection resistors and relays. The protection resistors have a default value of zero ohms, but they can be replaced with any value desired. For example, if you are testing oxide breakdown and you are concerned about current surges when the oxide ruptures, then you can substitute a high-value resistor to limit the current spike. Notice that this card allows you to easily switch from applying stress to a device to performing measurements on that device using an SMU. Moreover, while you are measuring one device you can still continue to apply stress to the other devices.

Each E5255A card can be individually programmed via internal jumper cables and DIP switches. Each block of eight outputs on a card can be tied to separate SMU inputs and DC bias sources, or they can all be tied to the same SMU and DC bias source (and every combination in-between). The following illustration displays the component locations on the E5255A multiplexer card.

Keysight E5255A Component Locations



IVin cable connection	DIP SW setting BIT1	BIT2	BIT3
SMU1	1	0	0
SMU2	0	1	0
SMU3	1	0	0
SMU4	0	1	1
SMU5	1	0	1
SMU6	0	1	1

Resistors	BIT4
0 Ω	0
other	1

Figure 4.19. Overview of the component locations on the E5255A multiplexer card.

Configuring an E5255A card requires two processes: setting the card's DIP switches, and arranging the jumper cables appropriately. For example, if you want all three blocks of multiplexer card outputs to be connected to the same SMU and the same bias source, you need to set the DIP switches and arrange the connections as shown below.

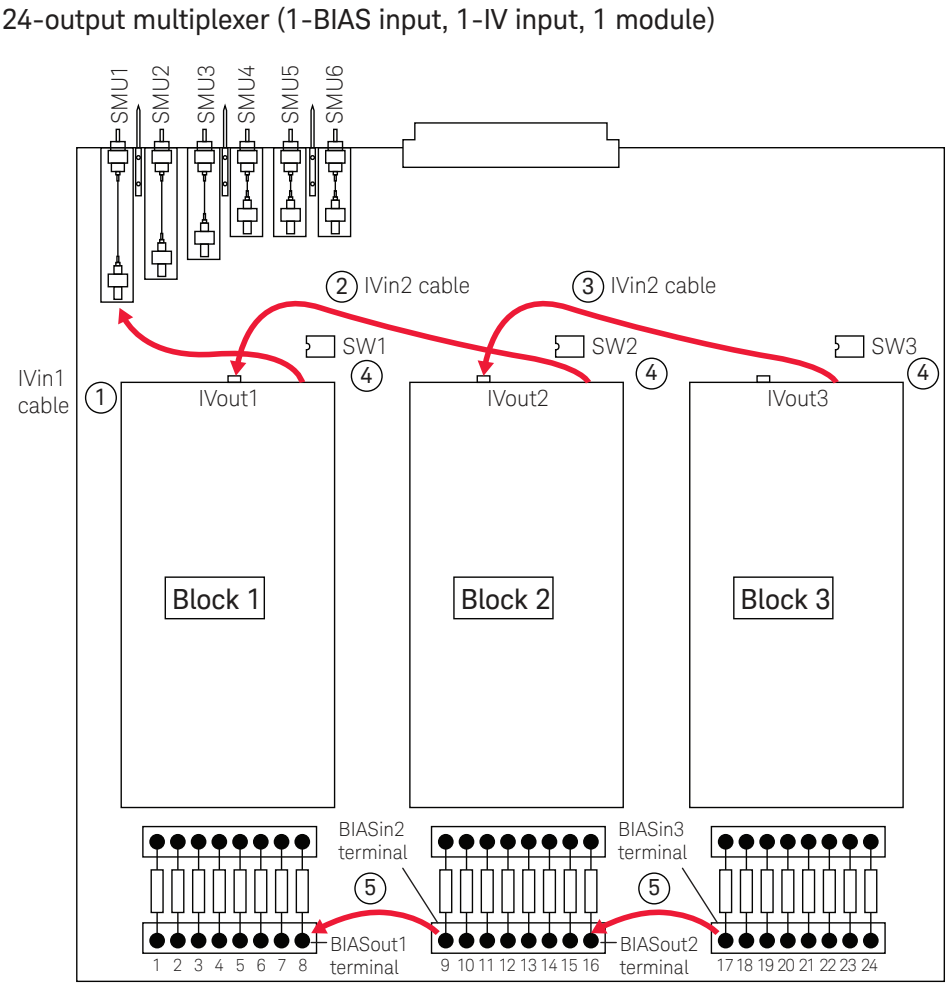


Figure 4.20. Example E5255A card connections where all three blocks are using the same bias source and are connected to the same SMU.

By using external BNC cables, it is easier to tie the bias inputs from multiple cards together (for example if you want to bias 96 channels with the same power supply).

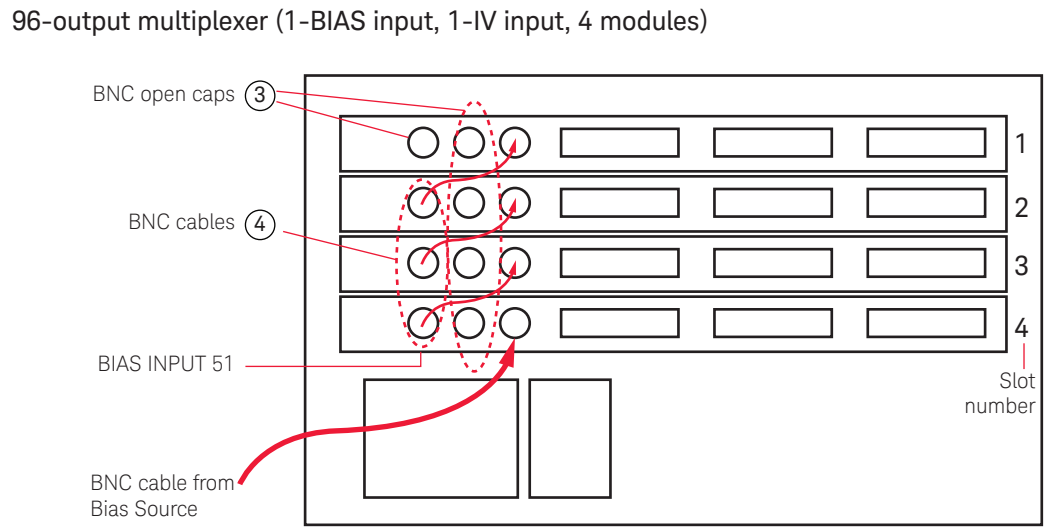
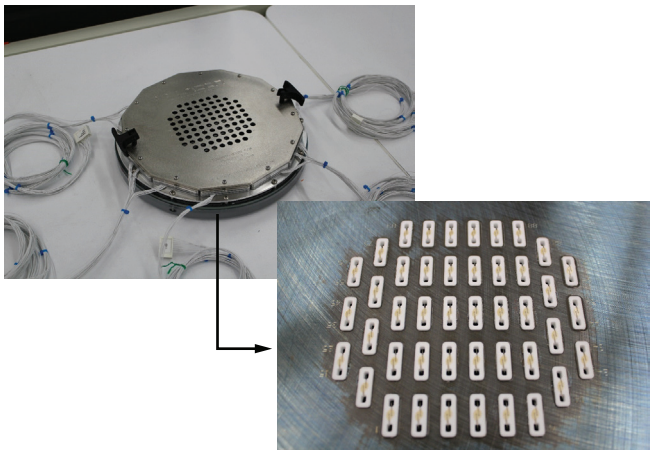


Figure 4.21. Illustration showing how multiple cards can use the same power supply bias source.

This discussion has shown how useful the E5250A switching matrix and the E5255A card can be for reliability applications requiring long-term stress and periodic measurements. Keysight can supply cables (16494D) of various lengths that mate with the E5255A card outputs and that can pass through connector plates (16495C/D) into either a thermal test chamber or a wafer prober. If you want to perform multi-site on-wafer reliability testing, then companies such as Celadon® Systems can supply probe cards specifically designed for this purpose. In this case, Celadon will supply a customized cable to connect their probe card to the outputs of the E5255A multiplexer cards.



Underside of multi-site probe card

Figure 4.22. A probe card designed for multi-site reliability testing.
Note: Photo courtesy of Celadon Systems

Positioner-based switching solutions

Introduction

Switching matrices can be a very cost-effective solution when probe cards are used. However, the vast majority of laboratory-based analytical wafer probing involves the use of individual positioners rather than probe cards. Although it is certainly possible to use a switching matrix with positioners, the cost and complexity of this type of solution may not be optimal. Fortunately, Keysight Technologies has several options for positioner-based wafer probing switching that do not require a switching matrix. The following sections will explain these available options.

Note: Since RF positioners should never be used with any sort of a switching solution, all further discussions in this section refer to DC positioners.

Atto-sense and switch unit

In the earlier discussion of SMUs in chapter 3, it was mentioned that an atto-sense and switch unit (ASU) is available, and that this unit can be used with either the MPSMU or HRSMU to achieve current measurement resolution down to 0.1 fA. In addition to possessing this impressive low-current measurement capability, the ASU (as the name implies) also has a built-in switching capability. Each ASU has two BNC inputs that are nominally designed to be used with two of the outputs from a capacitance meter, although they can be used with any type of instrumentation that has BNC outputs. A schematic of a single ASU is shown below.

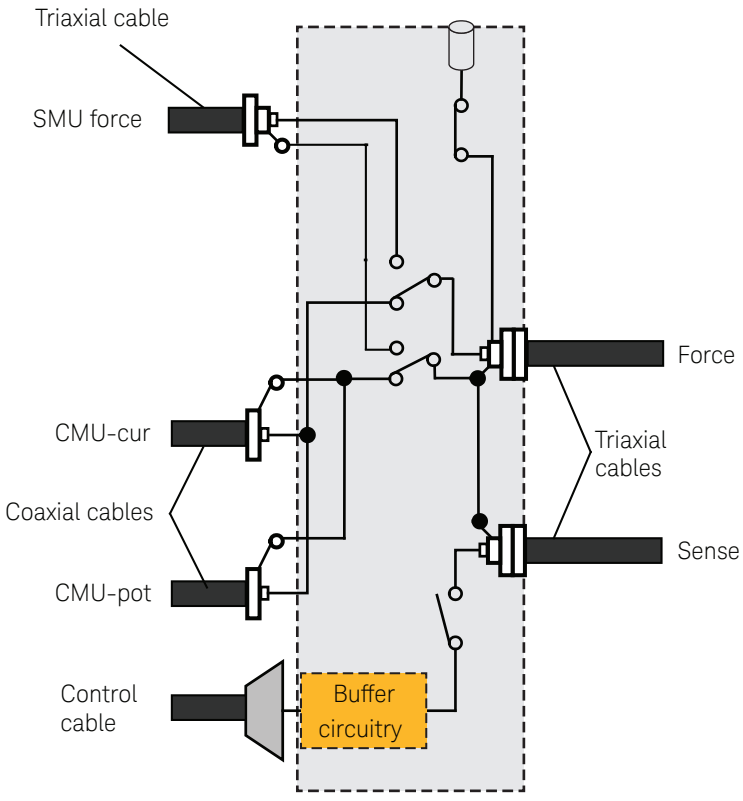


Figure 4.23. A simplified ASU schematic showing its switching capabilities.

Atto-sense and switch unit *(continued)*

A picture of the ASU's inputs is shown below.

Figure 4.24. The inputs to the ASU, showing the two coaxial inputs (top), the triaxial input (from the force line of the MPSMU or HRSMU, lower right) and the control cable input (also coming from the MPSMU or HRSMU, lower left).

The advantage of combining switching capability with 0.1 fA measurement capability should be obvious: you can switch between IV and CV measurements on positioners without having to change any cables. Although the issue of capacitance compensation will not be discussed until chapter 8, it is sufficient to state here that the B1500A will automatically detect and compensate for the effects of the two supported cabling connections (1.5 m and 3.0 m) between the MPSMU/HRSMU and the ASU. This means that the only further cabling compensation required is for the additional items (triaxial cables, positioner and probe tip) that are attached to the triaxial outputs of the ASU.

Note: Of course, the length of these "cable extensions" should be kept as short as possible.

SMU CMU unify unit

The ASU works well for CV-IV positioner-based wafer probing, but its 0.1 fA current measurement resolution is not always necessary. For this reason, Keysight supports another positioner-based switching solutions for the HRSMU and MPSMU. This solution is the SMU CMU unify unit (SCUU). The SCUU is a switching module designed to work with the B1500A's MFCMU and two of either the high-resolution or medium power SMUs. The SCUU form-factor is designed to accept the four BNC outputs of the MFCMU and the two pairs of Kelvin triaxial outputs from the two HRSMUs and/or MPSMUs. The SCUU provides the capability to switch between the MFCMU and the HRSMU/MPSMU pair automatically, and it outputs into two pairs of triaxial outputs. The SCUU also allows you to use the SMUs as DC bias souces during capacitance measurement, supporting up to ±100 V. A diagram of the SCUU showing its connections is shown below.

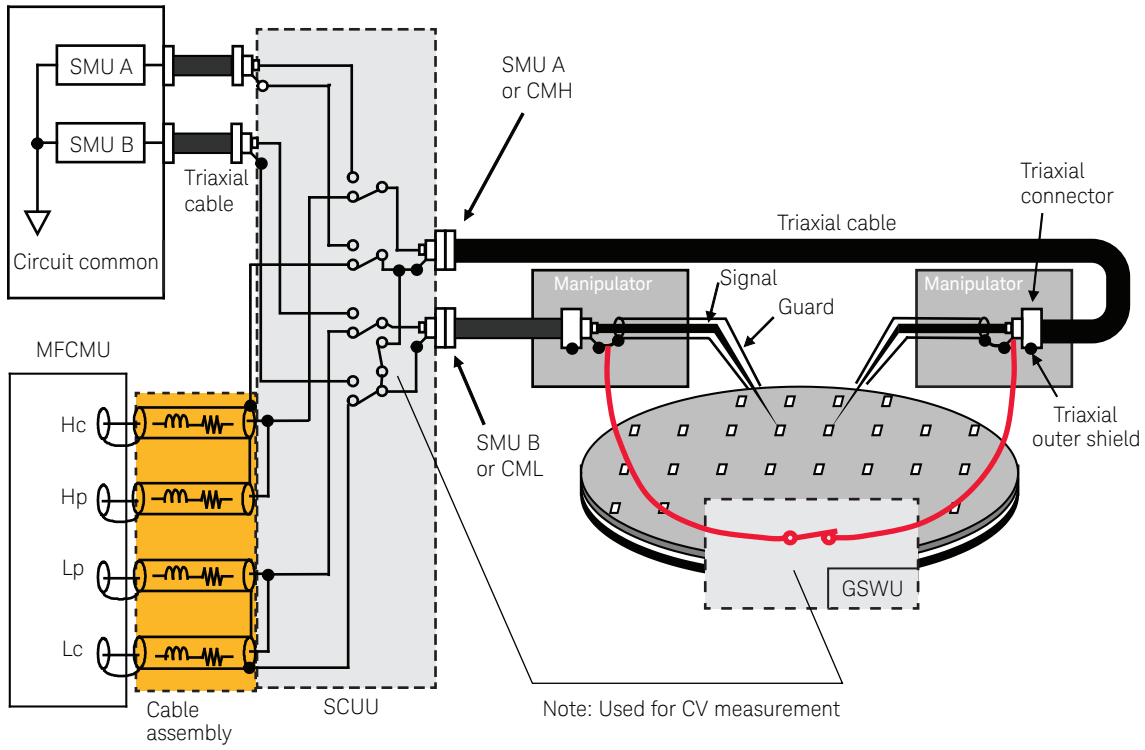


Figure 4.25. The SCUU supports switching between the MFCMU and two SMUs (HRSMU or MPSMU).

SMU CMU unify unit *(continued)*

A picture of the SCUU (mounted on the rear of the B1500A) is shown below.

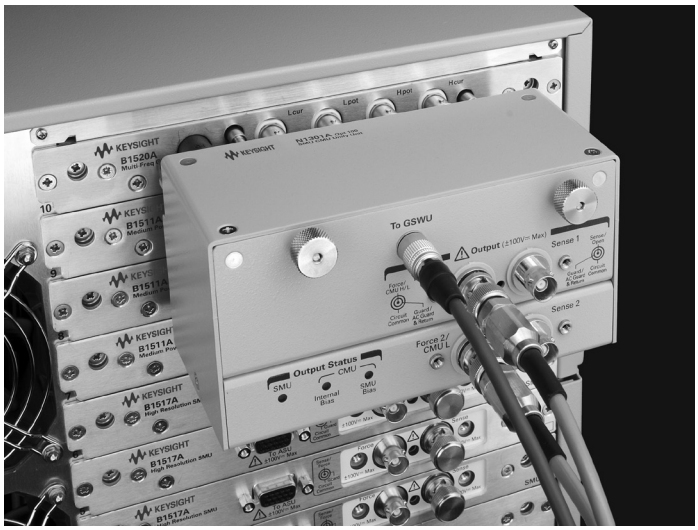


Figure 4.26. A picture of the SCUU mounted on the rear of the B1500A to illustrate how the MFCMU and SMU outputs are converted into two sets of triaxial outputs (force and sense).

Note: It is important to understand that figure 4.26 is for illustrative purposes only and that when making on-wafer measurements, the SCUU should never be mounted directly on the rear of the B1500A.

The advantages of this solution for a positioner-based wafer probing environment are the same as those of the ASU; namely, you can switch between IV and CV measurements without having to change any cables. Also, as in the case of the ASUs, Keysight can supply a cable (3.0 m) that allows the SCUU to be mounted on an analytical wafer prober in close proximity to the DUT and the B1500A will automatically detect and compensate for the effects of the supported cabling connection. Like for the ASU, the only further cabling compensation required is for additional items such as the triaxial cables, positioner and probe tip that are attached to the triaxial outputs of the SCUU. Keysight can also supply a magnetic mount for the SCUU, although most analytical wafer prober companies can also supply mounting hardware for the SCUU.

There is one other optional component of the SCUU solution that has not yet been mentioned even though it was shown in the diagram (Figure 4.25) and the cable connected to it was shown in the SCUU picture (Figure 4.26). It is the guard switch unit (GSWU). The GSWU can be connected to the SCUU, and its purpose is to short the guard lines of the two positioners used together (using an internal relay) whenever a capacitance measurement is being made. In normal current-voltage (IV) operation, the GSWU does not short the guards together, which is necessary since the guards have to track the independent signal lines in order to permit low-current measurement. The reasons for shorting the guards together during CV measurement will be discussed in the next section.

Shorting the guards together during measurement

Although it is not possible to go into all of the details why shorting the guards together is important when making capacitance measurements until after we cover capacitance meter operation in chapter 8, we can at least give a basic explanation here. We will cover the issues mentioned here in greater depth in chapter 8.

In order to measure capacitance, we have to measure impedance, which is a complex quantity consisting of a magnitude and a phase (or a resistance and reactance, if you prefer). In addition to having inherent resistance, all cabling also has inherent capacitance and inductance. When we perform capacitance compensation, we are essentially measuring the impedance characteristics of the cabling so that we can subtract them out from our measurement in order to arrive at the true (actual) value of the capacitor that we are trying to measure. It is implicitly assumed in this process that the impedance characteristics of the cables being used are stable. However, in the case of flexible cabling, it turns out that this is not necessarily a valid assumption. If the guards are not shorted, then any movement of the cables can cause their series inductance to vary, thereby invalidating any compensation performed and resulting in inaccurate capacitance measurements. The series inductance of coaxial cables can vary by several hundred nano-Henries (nH) per meter unless the cable guards are shorted together. The graph shown below illustrates to what extent a measurement error can occur when the guards are not shorted during a capacitance measurement.

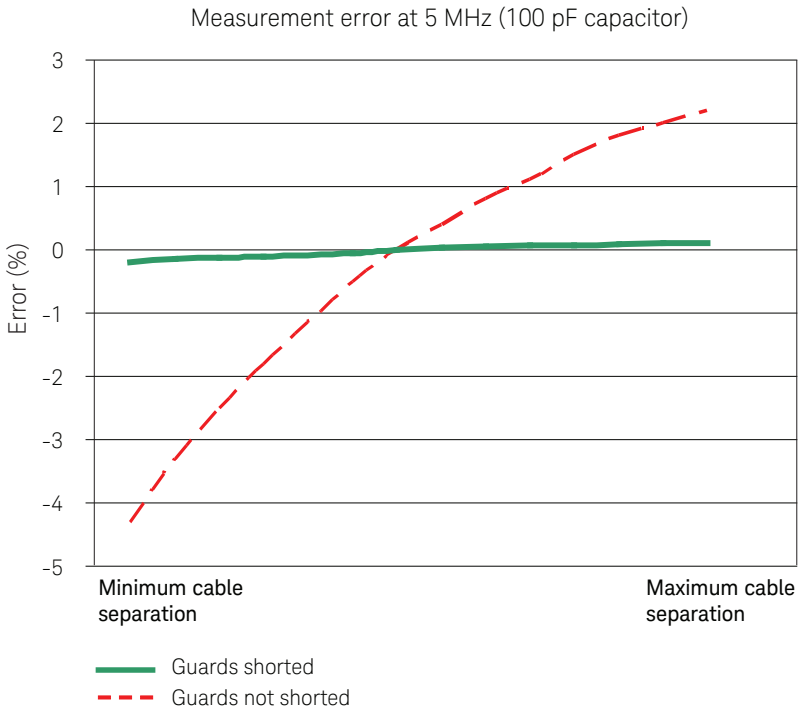


Figure 4.27. A comparison of the measurement error (%) for a 5 pF capacitor measured at 5 MHz when the guards are shorted (solid line) versus that when not shorted (dashed line).

Shorting the guards together during measurement *(continued)*

In addition to the concerns with variable series inductance, there is another reason for shorting the guards together during capacitance measurement. A wafer resting on a chuck creates an LCR circuit, which has a resonant frequency. In order to short-out this circuit and prevent oscillations, you need to connect the probes together using a short cable. This is illustrated in the figure shown below.

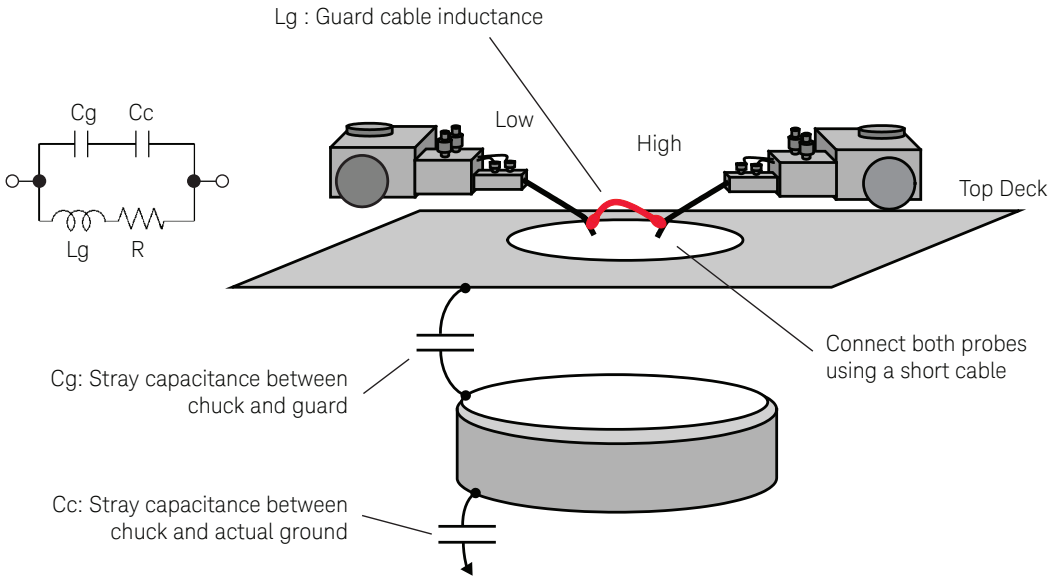


Figure 4.28. When parallel resonance occurs during a measurement, shorting the guards together will eliminate the problem.

Based on the above, it should be clear that we need to short the guards together during capacitance measurements for both the ASU and SCUU cases. In the case of the ASUs, Keysight supplies a shorting cable with the units for this purpose. If you are using the ASUs with the B1500A and EasyEXPERT software, then the software will automatically take care of shorting the guards together whenever you make a capacitance measurement. A simplified schematic showing the ASU connections for making CV-IV measurements is shown below.

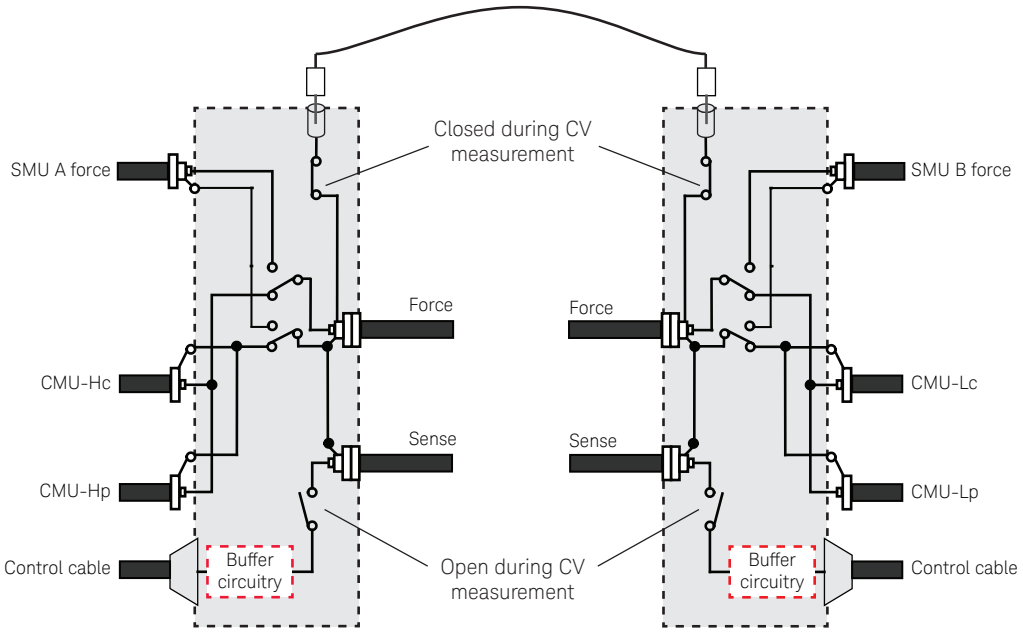


Figure 4.29. Simplified schematic showing the ASU connections when performing both CV and IV measurements.

Shorting the guards together during measurement *(continued)*

A photo of two ASUs mounted on an analytical wafer prober with their guards tied together is shown below.

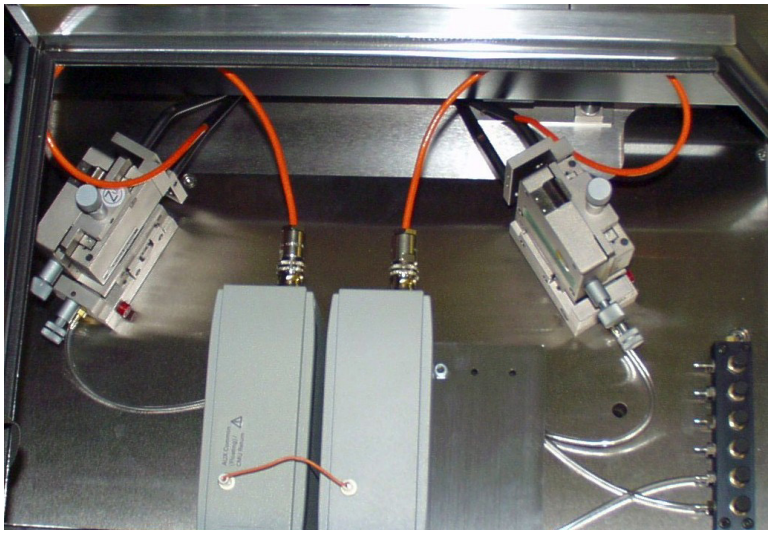


Figure 4.30. A picture showing the ASU guard connection as implemented on an analytical wafer prober.
Note: Photo courtesy of FormFactor

In the case of the SCUU, you need the optional GSWU in order to short the guards together during CV measurements. As in the case of the ASU, the EasyEXPERT software automatically takes care of shorting the guards together during a CV measurement (and an LED on the GSWU will light up when this occurs). A photo showing the GSWU mounted on a wafer prober and connected to the guards of two positioners is shown below.

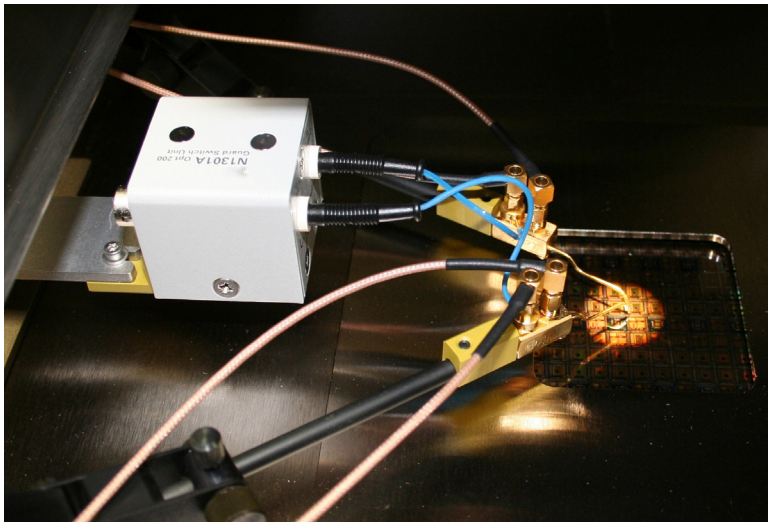


Figure 4.31. Picture showing the GSWU mounted on an analytical wafer prober and connected to the guards of two positioners.
Note: Photo courtesy of FormFactor

The following figure displays all the recommended accessories to perform accurate IV and CV measurements on-wafer using the B1500A's SCUU.

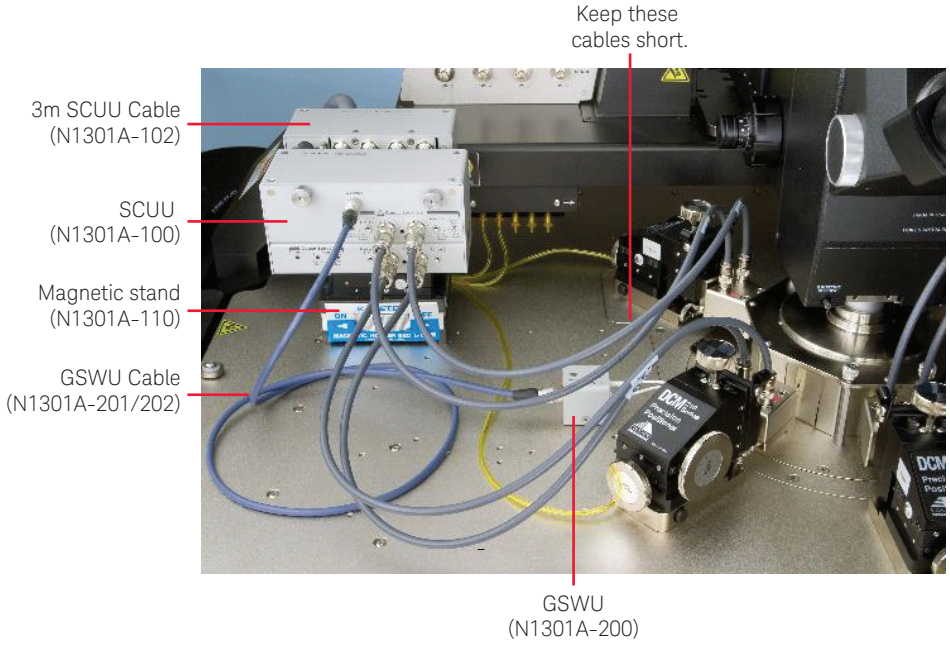


Figure 4.32. An optimal wafer probing configuration for performing both IV and CV measurements using the SCUU.

Summary of Keysight positioner-based switching solutions

The following table summarized the three Keysight positioner-based switching solutions. You can combine the B1500A's MFCMU and MPSMUs or HRSMUs with the SCUU or ASUs to create positioner-based CV-IV switching solutions to meet a variety of needs. All solutions are full Kelvin, support CV-IV switching and have integrated capacitance compensation.

10 fA/0.5 µV solution	**1 fA/0.5 µV solution**	**0.1 fA/0.5 µV solution**
1 × MFCMU 2 × MPSMU 1 × SCUU	1 × MFCMU 2 × HRSMU 1 × SCUU	1 × MFCMU 2 x MPSMU/HRSMU 2 × ASU
Kelvin measurement Multi-frequency capacitance measurement CV-IV switching Integrated capacitance compensation		

Figure 4.33. Table showing the three positioner-based switching solutions supported by the B1500A.

Note: The techniques involved in making capacitance measurements through a switching matrix will be deferred until chapter 8 after we have completed a more thorough discussion on capacitance measurement.

High voltage and high current wafer probing

Overview

To perform high-voltage and high-current wafer probing, you need to use a wafer prober designed for this purpose. While this statement might seem apparent, it is sometimes forgotten or ignored by users who are eager to begin making high power on-wafer measurements. High power analytical wafer probers differ from standard analytical wafer probers in several key ways:

1. Carefully designed safety schemes to prevent any chance of accidental exposure to high voltage.
2. Support for the unique connectors used to make high-voltage and high-current measurements.
3. Special positioners for high-current wafer probing.

We will see as we progress through this section, the cabling and connectors for high-voltage and high-current wafer probing are much more complicated than that for standard wafer probing. In addition, different analytical wafer prober companies use different connector schemes. This makes it extremely important to specify exactly what wafer prober company you plan on using so that Keysight can correctly configure a solution for you.

Connection considerations

The first point you need to consider is exactly what type or types of high-voltage or high-current wafer probing you want to perform. If you only plan on performing one specific type of high-power wafer probing (only high-voltage or only high-current), then you probably do not need the module selector unit. However, if you want to perform both high-voltage and high-current measurements, (or you think you might want to do both in the future) then it is highly advisable to purchase the module selector unit. As we discussed in chapter 3, the module selector unit allows you to switch between the HVSMU, HCSMU or HPSMU without having to change any cables. Of course, you can still use the module selector unit even if only one of these modules is connected to it, making it easy to expand your B1505A's measurement capabilities in the future if you add additional modules to it. If you are using the N1265A current expander/test fixture, then the module selector unit is integrated into it as a standard feature.

Connection considerations *(continued)*

When you order a B1505A, with either the N1259A or N1265A test fixtures, all of the necessary cables to connect to the test fixture are included. However, if you want to connect to a wafer prober, you will need additional cables and connectors. Some of these can be supplied by Keysight, and some may need to be supplied by your wafer prober vendor. Another factor that must be taken into consideration is that some of the lower power B1505A modules (the HPSMUs, the MPSMUs, the MCSMU and the GNDU) can be damaged if exposed to the high voltages generated by the HVSMU. To avert this, Keysight can supply a variety of protection adapters that prevent accidental damage to these modules. There are several different versions of the protection adapters that can support the different types of connectors used by the various analytical wafer prober companies. Keysight can also supply resistor boxes (R-boxes) in different resistor values (1 MΩ, 100 kΩ and 1 kΩ) to place a resistor in series with the cables that you are using. The 1 MΩ and 100 kΩ R-boxes are used to prevent device damage due to voltage spiking, while the 1 kΩ R-box can be used to suppress device oscillation. An illustration showing the most common connectors and adapters is shown below.

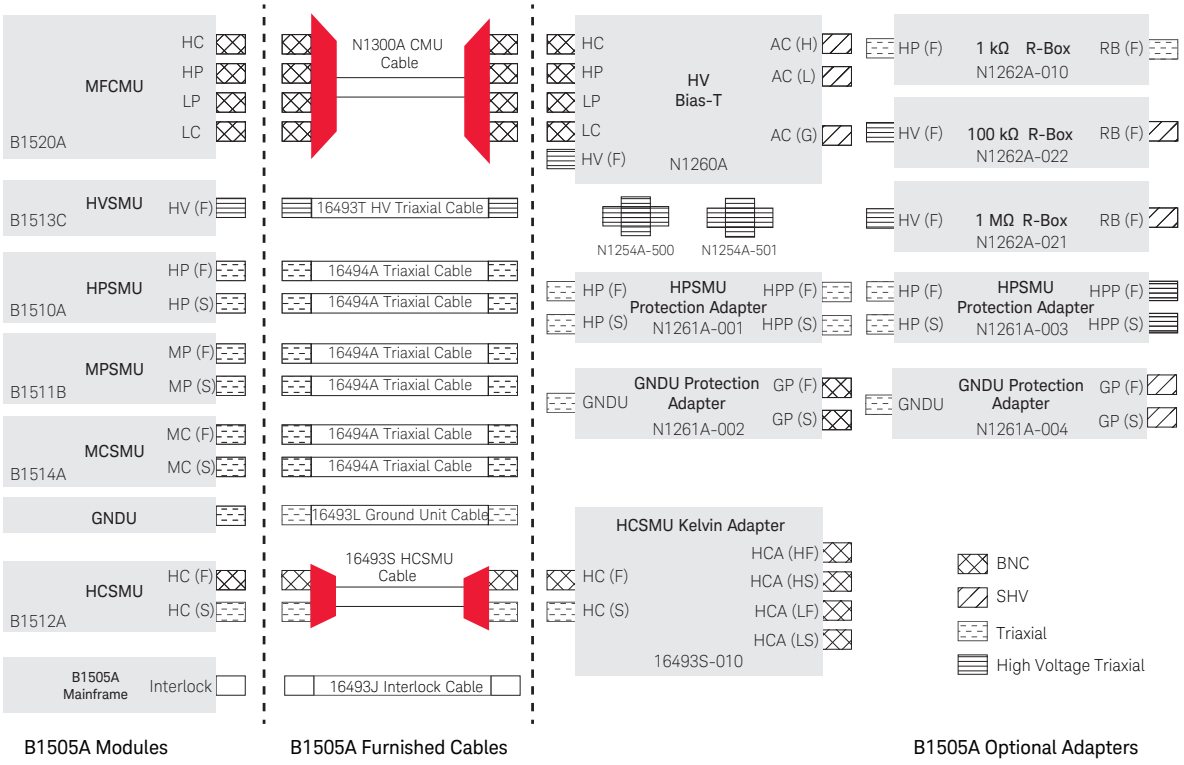


Figure 4.34. Cables and adapters available to connect the B1505A to a wafer prober.

Connection considerations *(continued)*

If you are using the N1258A module selector unit with your wafer prober, then the cables and accessories that you need will be different than when you are not using the module selector unit. The main difference is that the module selector unit contains device protection circuitry so that you do not need to use the protection adapters for the modules connected to the module selector. Of course, any modules not connected to the module selector would still require the use of protection adapters if there is any possibility of them being exposed to high voltages and currents. Again, Keysight can help you with the correct configuration for your use-model and the wafer prober that you plan on using. An illustration showing the most common connectors and adapters that would be used with the module selector unit is shown below.

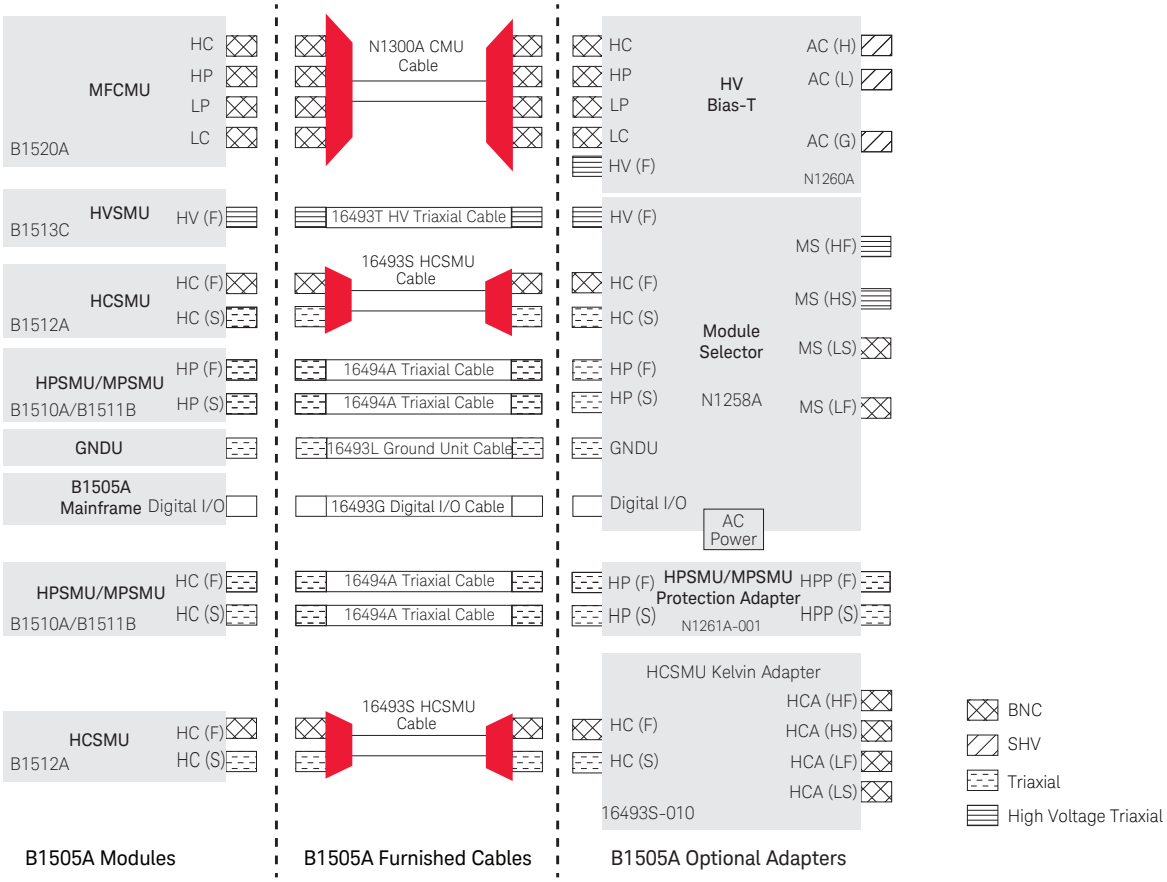


Figure 4.35. The module selector unit also performs protection functions, so the protection adapters are not needed for the modules connected to the module selector unit.

This collection of cabling and adapter accessories is rather daunting, and it is easier to understand their use by showing different sample wafer probing configurations.

Note: The explanation on how to use the high-voltage bias-T will be deferred until chapter 9 (in which we will discuss high-power capacitance measurement).

Connection considerations *(continued)*

The first case we will examine is high-voltage wafer probing to measure lateral device breakdown and leakage currents (such as MOSFET drain to source breakdown). For high-voltage wafer probing, it is essential to connect up the interlock from the B1505A to an approved interlock scheme supported by your wafer prober vendor. It is also important to use protection adapters for the HPSMU, MPSMU, MCSMU and GNDU modules to prevent high-voltages from accidentally damaging these modules. An illustration of this scheme is shown below.

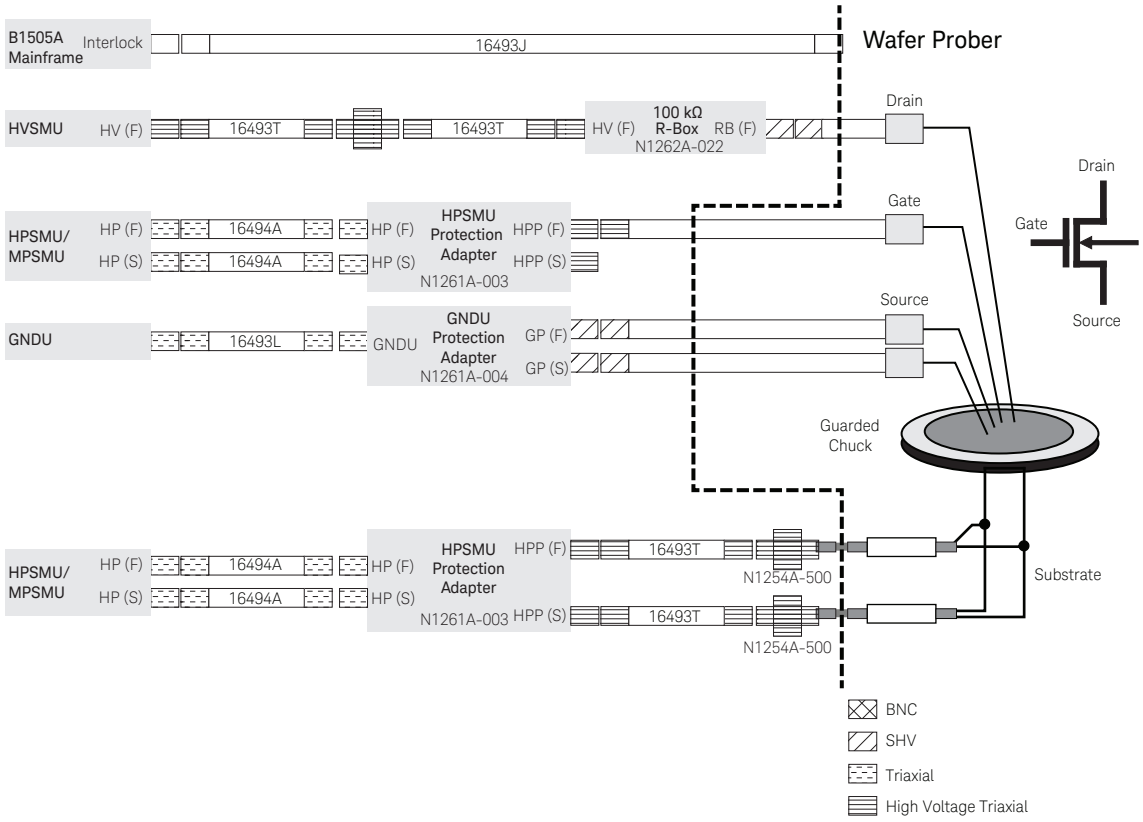


Figure 4.36. A B1505A wafer probing configuration to perform high-voltage (3 kV) breakdown measurements on lateral devices.

Note: If your wafer prober does not support SHV or high voltage triaxial connections, then you can use the HPSMU/MPSMU protection adapter N1261A-001 which has triaxial outputs and the GNDU protection adapter N1261A-002 which has BNC outputs.

Connection considerations *(continued)*

The next case we will examine is high-current wafer probing to measure high lateral currents using a Kelvin configuration (such as high-current MOSFET Ids measurement). Note that the BNC and triaxial outputs of the HCSMU must be separated into the high force/high sense and low force/low sense BNC outputs using the HCSMU Kelvin adapter. In this situation, Kelvin connections are virtually mandatory to prevent measurement error due to the large resistive voltage drops in the cabling that occurs when many amps of current are pumped through them. Also, remember that in chapter 3, we explained that the HCSMU outputs are floating and that the low force (LF) and low sense (LS) outputs need to be tied to ground. Therefore, in this case, they are tied to the ground unit using BNC T-connectors. An illustration of this scheme is shown below.

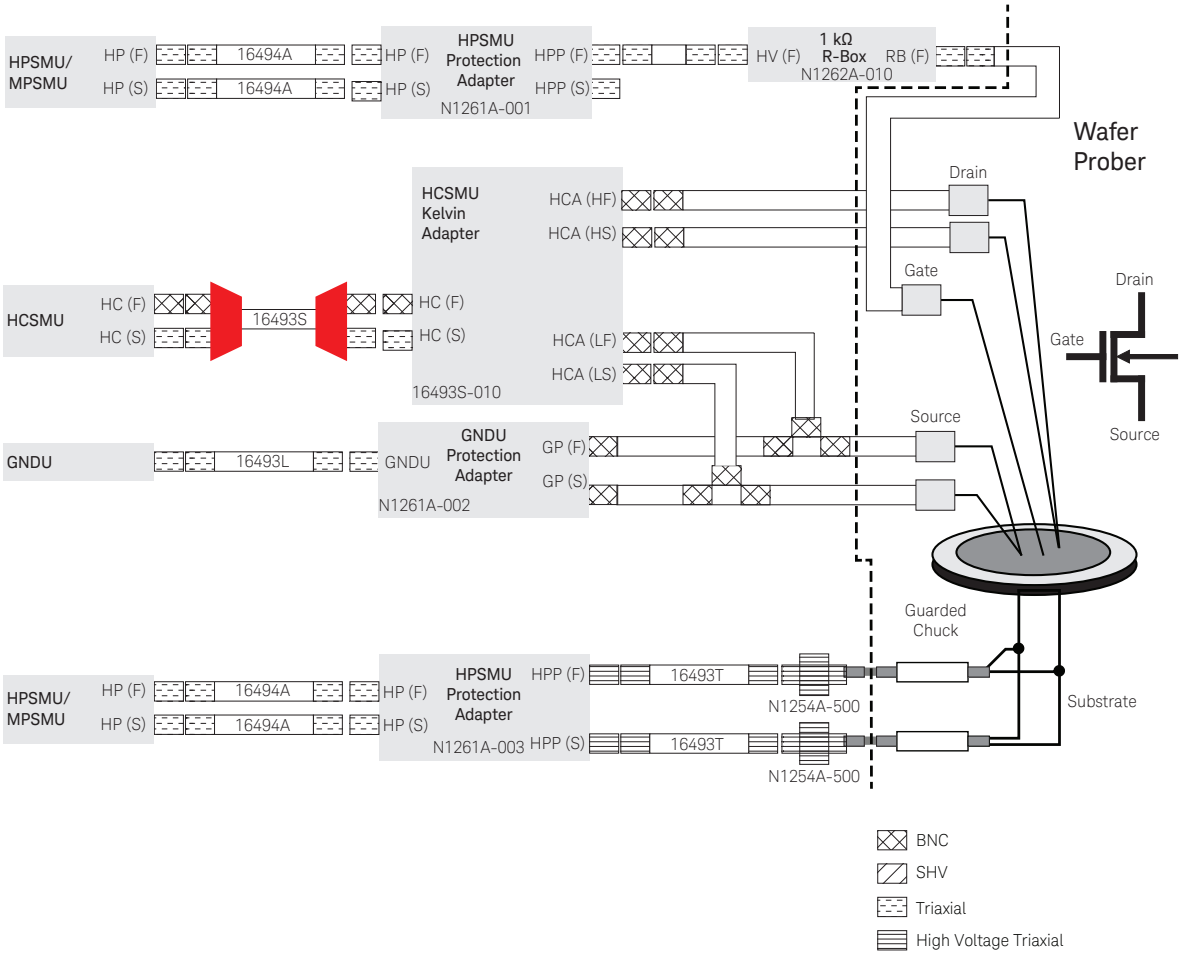


Figure 4.37. A B1505A wafer probing configuration to perform high lateral current (20 A) measurements.

Connection considerations *(continued)*

As can be seen from the previous two examples, it is not a trivial task to manually switch between high-voltage and high-current measurements on a wafer prober. Therefore, if you want to perform both types of measurements, you should invest in a module selector unit. In addition to allowing you to easily switch between high-voltage and high-current measurements, the module selector unit also allows you to use the HPSMU to perform high-accuracy, low level measurements not possible with the HVSMU and HCSMU. As the illustration below shows, the wafer prober connections when using the module selector unit are greatly simplified.

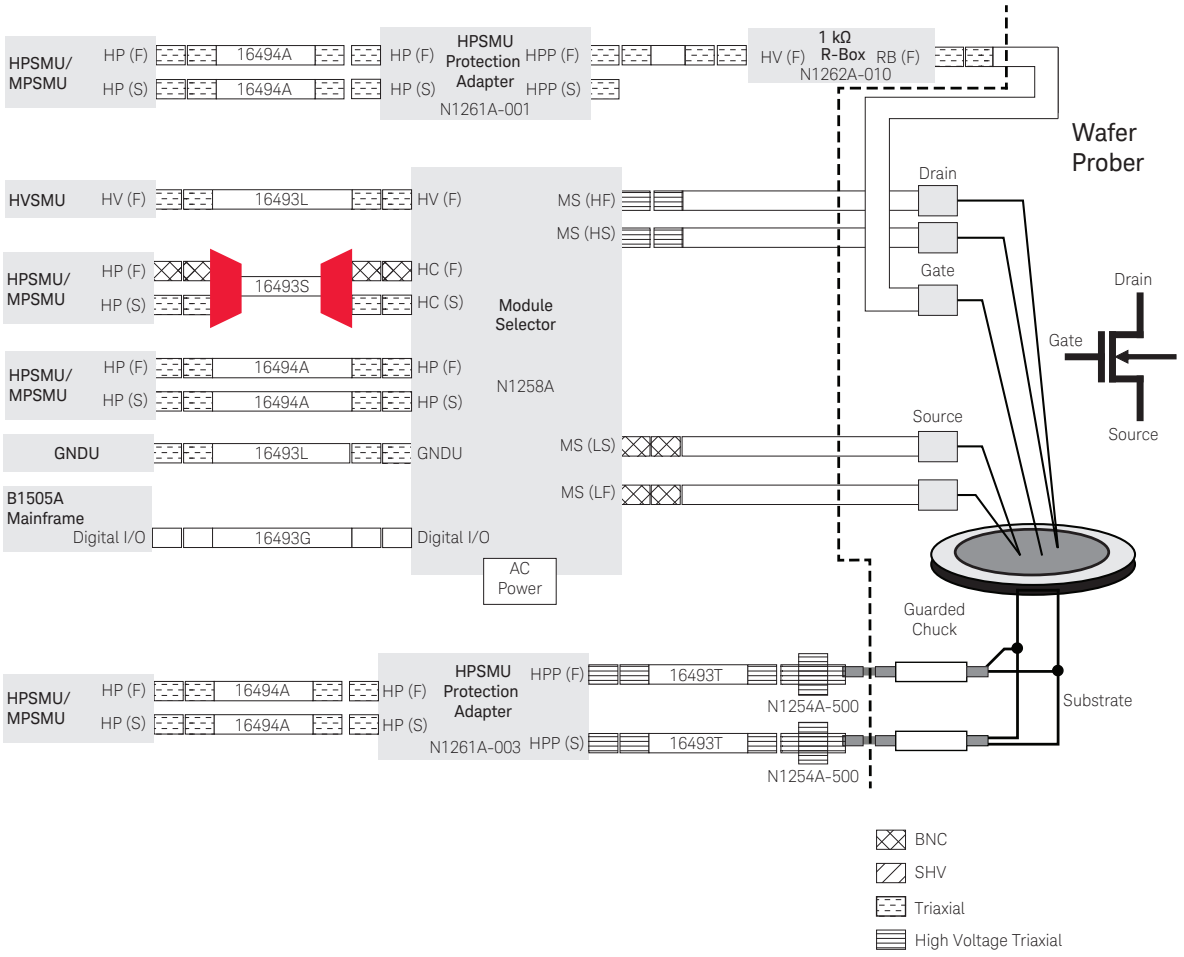


Figure 4.38. A B1505A wafer probing configuration with the module selector unit that can perform both high-voltage (3 kV) and high-current (20 A) measurements without having to change any cabling.

The additional cost of the module selector unit is more than compensated for by the improvements in convenience and productivity.

The previous schemes are limited in current by the capabilities of the HCSMU. Although it is possible to combine two HCSMU modules together to get up to 40 A from the B1505A mainframe, for device testing above 20 A it is much better to purchase the N1265A current expander/test fixture. The N1265A has options to support both 500 A and 1500 A, and it has a built-in module selector unit. In addition, as we will see in the next section, it is much easier to connect to a wafer prober if you have the N1265A.

Prober connections using the N1265A and N1254A-524

The B1505A's N1265A test fixture/current expander has the advantage that it can be used for high power wafer probing (up to 3 kV and several hundred amps) with minimal additional cabling and adapters. Furthermore, the same module selector unit (integrated into the N1265A) used for packaged device and module testing can be used for on-wafer measurement. This economy of performance is achieved using the N1254A-524 cable adapter. The N1254A-524 cable is approximately 1.8 meters long, and one end of the cable has blocks with banana plugs that mate with the outputs of the N1265A test fixture/current expander. The standard front window of the N1265A fixture is then replaced with one that has a notch (supplied with the cable) that allows the lid to close and the cable to be routed out to a wafer prober. The details of this connection scheme are shown below.

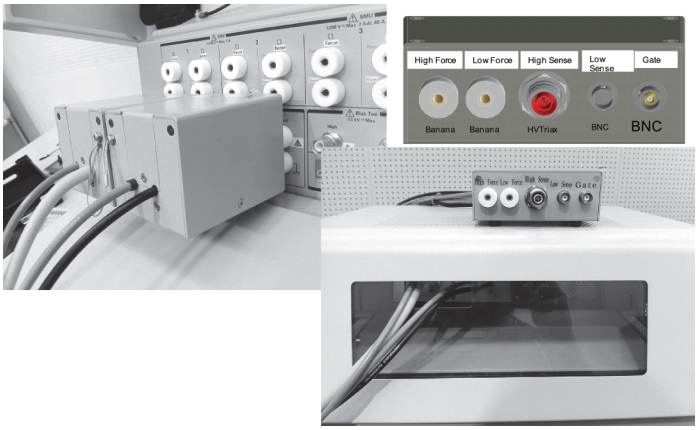


Figure 4.39. N1254A-524 cable connection details.

For simplicity of connection, the high force and low force outputs of the N1254A-524 cable are banana jacks (since these connections do not perform any measurement). The high sense line output of the N1254A-524 is high-voltage triaxial to support low-current measurement at high voltage. Finally, the low sense line and gate outputs of the N1254A-524 are BNC. The following figure shows how the N1254A-524 cable can be used to connect to a wafer prober.

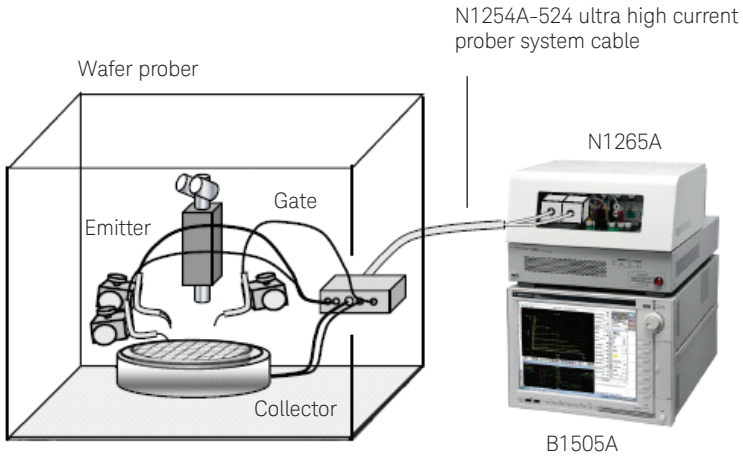


Figure 4.40. Connecting to a wafer prober using the N1265A test fixture and N1254A-524 cable.

As shown in Figure 4.40, the wafer prober connections are greatly simplified by using this cable versus directly connecting the B1505A mainframe outputs to a wafer prober.

Chapter 5 Time Dependent and High-Speed Measurements

"Experience is something you don't get until just after you need it." – Anonymous

Introduction

One of the most prominent trends in parametric testing is the transition from simple sweep and spot measurements (where measurement time is not an important part of the measurement) to measurements whereby the speed at which the measurement must be made is critical to the accuracy of the measurement. In some cases, this requires new types of measurement modules other than SMUs, or even the integration of external instruments (pulse generators, oscilloscopes, etc.) with parametric measurement equipment. However, some of the newer parametric instruments (such as the Keysight B1500A, B1505A/B1506A and B2900A series SMUs) can make impressive time sampling measurements just with their SMUs, so it is worthwhile to discuss all of the available measurement options.

Note: In the following discussions, I will often refer to the B1505A and B1506A together as "the B1505A/06A" for the sake of convenience. However, it is important to understand that in this context I am assuming that the B1506A is being controlled via EasyEXPERT rather than Easy Test Navigator.

Parallel measurement with SMUs

Parallel measurement is an important capability for all time dependent testing done with SMUs. True parallel measurement with SMUs is only possible when each SMU has its own analog-to-digital converter (ADC) available to it. The Keysight parametric instruments with this capability are the E5260 Series, the E5270B, the B1500A and the B1505A/06A. With all of these instruments (except for the E5260A Series), you can select either a shared high-resolution (HR) ADC or a per-SMU high-speed (HS) ADC for all of the various types of IV measurements as shown below.

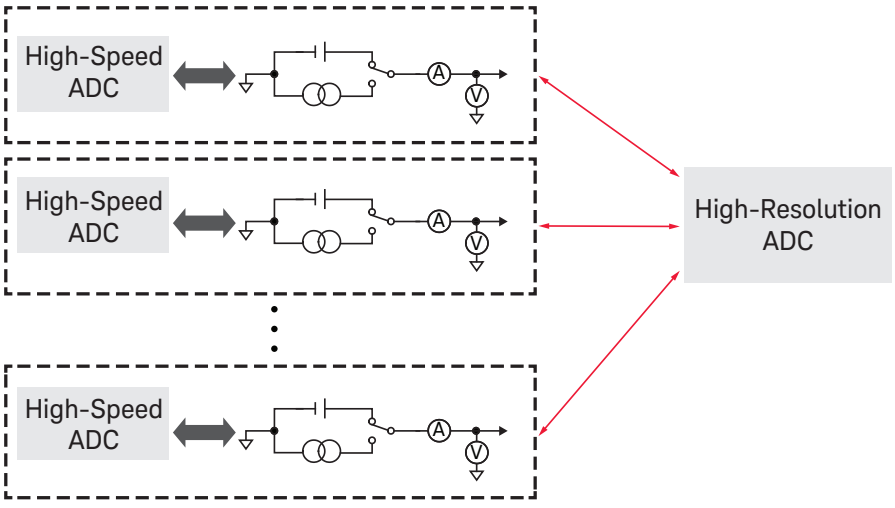


Figure 5.1. On instruments such as the E5260A, E5270B, B1500A and B1505A/06A you can select either the HS ADC or the HR ADC on a per SMU basis.

Parallel measurement with SMUs *(continued)*

In chapter 3, we discussed the multi-channel sweep mode of the B1500A and B1505A/B1506A. This feature enables the B1500A/B1505A/B1506A to make parallel measurements on multiple channels as long as the following conditions are met:

1. All parallel measurement SMUs must be in multi-channel sweep mode.
2. All parallel measurement SMUs must be using the HS ADC.
3. All parallel measurement SMUs must be in fixed measurement ranging.

If all three of these conditions are met, then parallel measurement occurs automatically. As was mentioned when we discussed multi-channel sweep mode, if the start and stop values for a VAR1 (sweeping) SMU are set to be the same, then the SMU functions as a constant voltage or current source during the sweep (and it can also make parallel measurements).

One question which is often asked regarding this procedure is how do we verify whether the SMUs are indeed performing measurement in parallel (since there is no actual parallel measurement command). The suggested method to verify parallel test is to set both the HS ADC and HR ADC to very long integration times (four or more PLCs). Repeat the measurement twice, the first time with the SMUs set to HS ADC and the second time with the SMUs set to HR ADC. It should be easy to detect a difference in the measurement times for these two cases (parallel measurement versus non-parallel measurement). Once you are convinced that the parallel measurement is working correctly, you can return the SMUs to their normal HS ADC setting and adjust the integration time for the HS ADC back to the value you want it at for the actual measurement.

Time sampling with SMUs

So far, we have discussed making various types of sweep and spot measurements with SMUs, but we have barely mentioned the element of time. The 4155C, 4156C, B1500A and B1505A/06A all possess a built-in time sampling capability that allows their SMUs to measure voltages or currents at specified intervals of time. This measurement capability is useful for certain types of reliability measurements such as time dependent dielectric breakdown (TDDB) where a device is placed under stress and continually monitored until the insulating layer ruptures. The B1500A/B1505A/B1506A setup screen for a time sampling measurement is shown below.

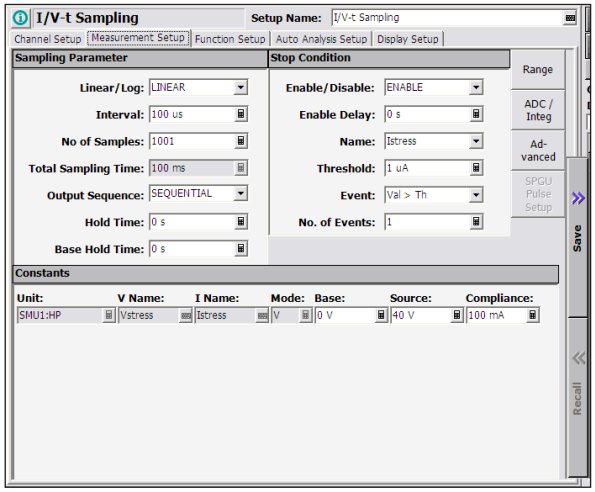


Figure 5.2. Screen capture showing how to set up a time sampling measurement on the Keysight B1500A/B1505A/B1506A (Classic Test mode).

In addition to specifying how frequently samples should be taken in time (the sampling "Interval"), there are many other parameters that can be specified. The "No. of Samples" parameter specifies the maximum number of sample points to be taken, and in the case of the B1500A/B1505A/B1506A, this can be as much as 100,001 points. Moreover, it is possible to take samples in either linearly or logarithmically spaced intervals and a variety of different logarithmic sampling intervals (points per decade) are supported. It is also possible to specify a "Stop Condition", which will stop the sampling measurement before the maximum number of sample points is reached if the specified stop conditions are met. This is a very important capability, since it can reduce test times by stopping a measurement once an event of interest (such as the rupture of a gate oxide) has occurs. An example of this is shown below.

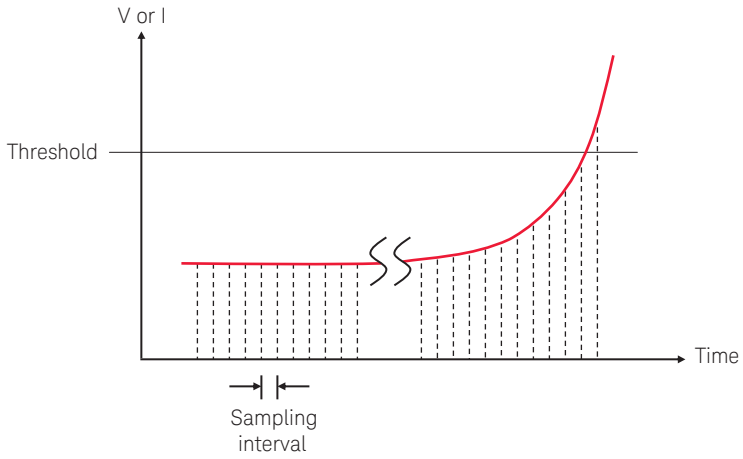


Figure 5.3. By setting up a stop condition, you can terminate a time sampling measurement once an event of interest has occurs (such as the rupture of a gate oxide).

Sequential versus synchronous sequencing

For time sampling measurements, it is sometimes important to consider how the SMU resources will be energized. For certain types of measurements, you need to be very careful with respect to the order in which the SMUs are turned on. For other types of measurements, you may actually want all of the SMUs to be energized simultaneously. For this reason, Keysight parametric instruments that support time sampling (the 4155C, 4156C, B1500A, B1505A and B1506A) all support both sequential and synchronous SMU output sequencing. In the case of the B1500A/B1505A/06A, if the "sequential" setting is selected, then the SMUs will turn on in the order shown on the "Channel Setup" page (from top to bottom). However, if the "simultaneous" setting is selected then the SMUs will all turn on at the same time.

Note: In both of these cases (sequential or simultaneous) the turn off order of the SMUs will follow the inverse order shown on the "Channel Setup" page (from bottom to top).

It is important to understand that the time sampling capabilities of the newer instruments (as well as their default settings) are different from those of the older 4145A/B. In many cases the failure to duplicate 4145A/B time sampling measurement results on the 4155C, 4156C, B1500A or B1505A/B1506A can be traced to differences in the turn-on sequence of the SMUs. If you are having trouble correlating measurements results between these instruments, you should check the settings on both instruments very carefully.

Setting up the time sampling interval

The most important point when making time sampling measurements with SMUs is that accuracy trumps every other setting. This point is often neglected by many users, resulting in much frustration and confusion. The point of this statement is that you cannot set up a time dependent measurement on an SMU that conflicts with the measurement accuracy you have specified. The SMU will always take as much time as it needs to meet the specified measurement accuracy, regardless of how fast you are telling it to make a measurement.

It is very easy to miss this point when setting up measurement conditions. For example, both limited ranging and auto ranging require the SMU to work its way down until the optimal measurement range (or range limit) is reached. Obviously, this takes time to complete. If you try to specify a time sampling interval that is smaller than the time required for the SMU to complete its measurement ranging, then the SMU will simply ignore the time settings and take as long as necessary to complete the measurement ranging. Therefore, it is always recommended to use fixed measurement ranging when performing time sampling measurements. Similarly, you cannot use power line cycle integration when making time sampling measurements. You must use an integration time of less than one PLC; in fact, even when using integration times of less than one PLC, you will need to be careful not to specify an integration time that exceeds the measurement interval you are trying to achieve.

Setting up the time sampling interval *(continued)*

The exact procedure and limitations on setting up a time sampling measurement are idiosyncratic to the instrument that you are using. The following are important considerations to keep in mind if you set the sampling interval to less than 2 ms on the B1500A/B1505A/B1506A.

1. All measurement channels must use the high-speed ADC; sampling intervals of less than 2 ms are not supported when using the high-resolution ADC.
2. If fixed measurement ranging is not selected, then the measurement channels automatically select the lowest fixed measurement range that covers the compliance value selected for that channel.
3. If the expected measurement time as determined by the number of averaging samples specified in the integration time is longer than the sampling interval, then the measurement channels will attempt to adjust the number of averaging samples so as to keep the specified sampling interval.
4. If multiple measurement channels are specified and all of the parameters are set up correctly, then all of the measurement channels will measure in parallel.

Advanced time sampling features (B1500A/05A/06A)

We have spent a lot of time explaining how the SMUs make measurements in time sampling mode, but we have not discussed much about how they force voltage and current. In time sampling mode the B1500A, B1505A and B1506A SMUs actually support some very useful sourcing features that have some important measurement uses. To begin with, let us review the basic output of an SMU during time sampling.

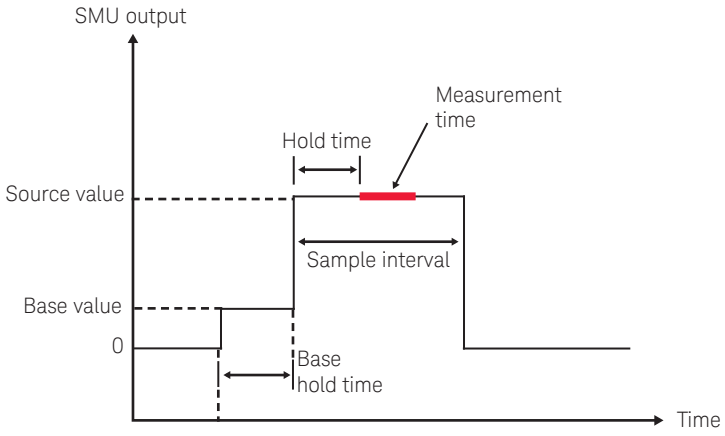


Figure 5.4. The basic SMU sourcing parameters for a time sampling measurement made on the B1500A/B1505A/B1506A.

Note that both a “Base Value” and a “Base Hold Time” can be specified. It should also be pointed out that in this diagram (Figure 5.4) as well as the following diagrams, the assumption is made that the measurement times are all less than the specified sampling interval (as is should be if everything is set up correctly).

The B1500A, B1505A and B1506A support a very useful feature (found under the “Advanced” settings) that permits SMU output to remain active after the measurement completes. This feature is referred to as “bias hold after measurement”, and it can be set to either the base or source value. The figure shown below illustrates the case where the SMU has been set to hold its base value.

Advanced time sampling features (B1500A/05A/06A) *(continued)*

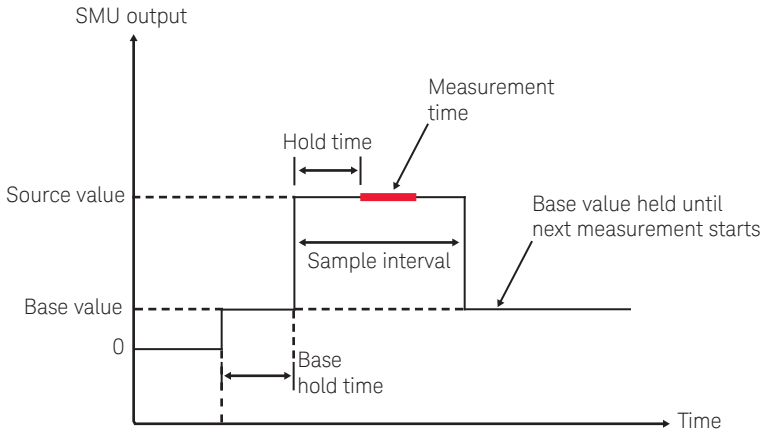


Figure 5.5. The B1500A and B1505A/06A permit you to hold the bias to a non-zero value after measurement and this can be set to either the base or source value.

Note: It is important to understand that the bias is held only under two conditions.

1. When running multiple tests within another application test.
2. When sequencing tests in “Quick Test” mode.

The bias hold feature is not the same as the SMU Standby Mode, which energizes the SMUs to a specified value at all times (even when a measurement is not being made).

Finally, there is one other very interesting feature supported by the B1500A and B1505A/06A in time sampling mode. This is the ability to specify a negative hold time. The figure shown below illustrates what is meant by a negative hold time.

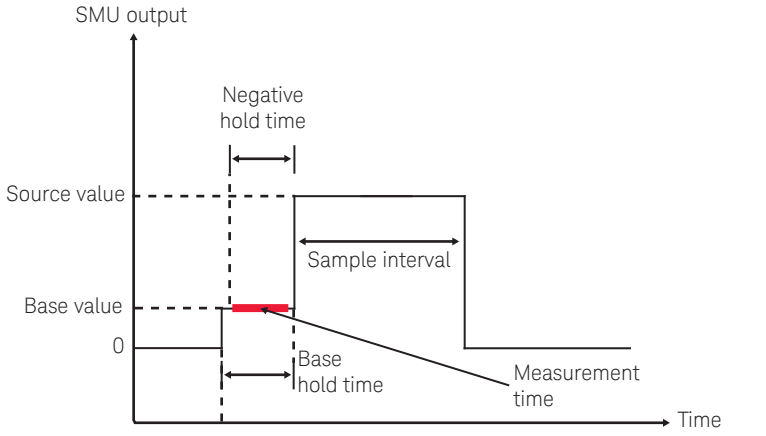


Figure 5.6. The B1500A and B1505A/06A support a negative hold time feature that permits the SMU to start measuring before the SMU starts to supply stimulus.

Advanced time sampling features (B1500A/05A/06A) *(continued)*

It is important to note that the negative hold time feature is only supported when the sampling interval is less than 2 ms, and that the maximum negative hold time supported is –90 ms.

It turns out that these advanced time sampling features can be used very effectively for certain types of reliability measurements. In particular, negative bias temperature instability (NBTI) measurements can be performed very fast and effectively with SMUs by employing the aforementioned measurement features. Essentially, in an NBTI measurement, we apply stress bias to a transistor gate and then periodically remove the stress to measure the transistor characteristics. However, if the stress drops to zero during the stress-to-measure transition, then the transistor can "recover" and the measurement results will be invalid.

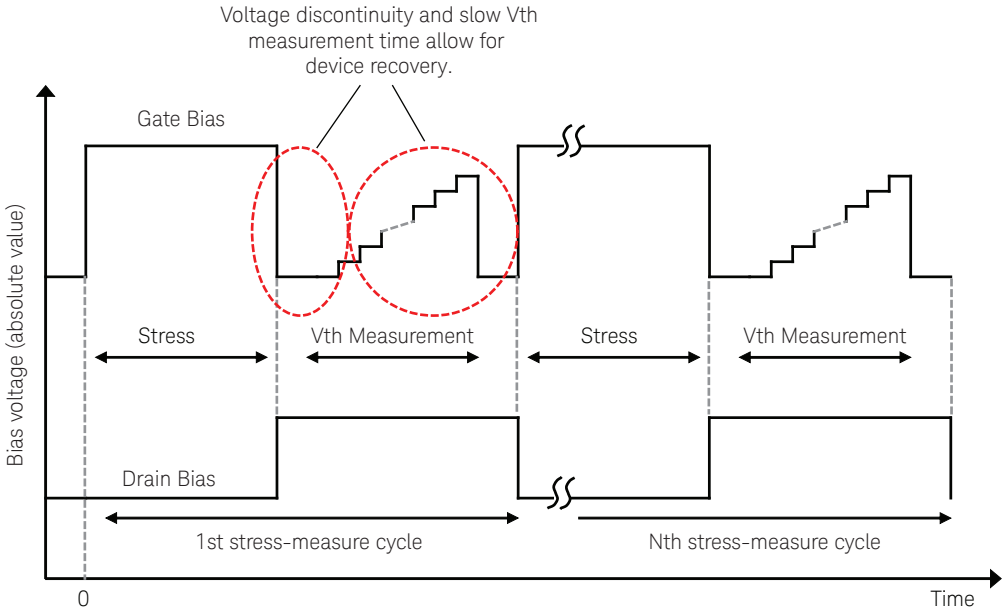


Figure 5.7. When making NBTI measurements it is essential that no discontinuities occur during the transition from stress to measurement.

Advanced time sampling features (B1500A/05A/06A) *(continued)*

By using the bias hold feature on the B1500A and B1505A/06A, we can make sure that this does not happen. In addition, by using the negative hold time feature, we can actually start measuring while still in the stress phase. This ensures that no data is lost during the stress to measure transition. This technique is illustrated in the figure shown below.

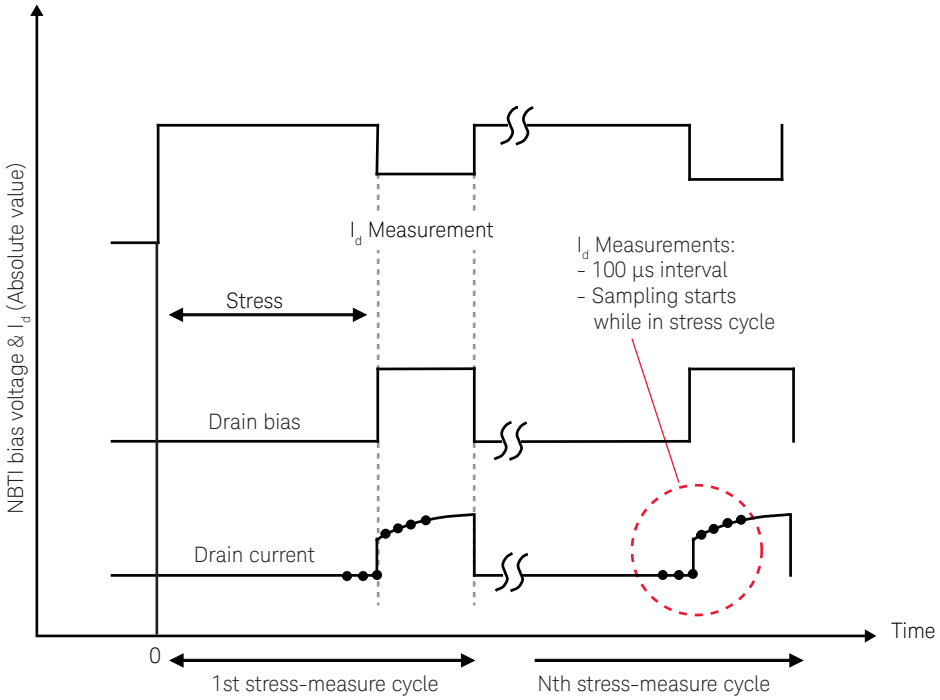


Figure 5.8. Making ultra-fast NBTI measurements using B1500A SMUs.

This technique is often referred to as “on-the-fly NBTI” measurement. The B1500A can make these measurements on a single SMU in 100 µs intervals. Using parallel measurement techniques, the B1500A can make parallel measurements using multiple SMUs in intervals given by the following equation:

$$t = 100\ \mu s + [20\ \mu s \times (n - 1)]$$

Here *n* is the total number of SMUs measuring in parallel. For more information, please refer to the relevant B1500A application notes on this topic.

Maintaining a constant sweep step

It is sometimes desirable to perform a sweep measurement with constant sweep steps. However, it is important to understand that in a standard sweep measurement, there is no certainty of the measurement time required to complete each point in the sweep. The following plot illustrates this situation.

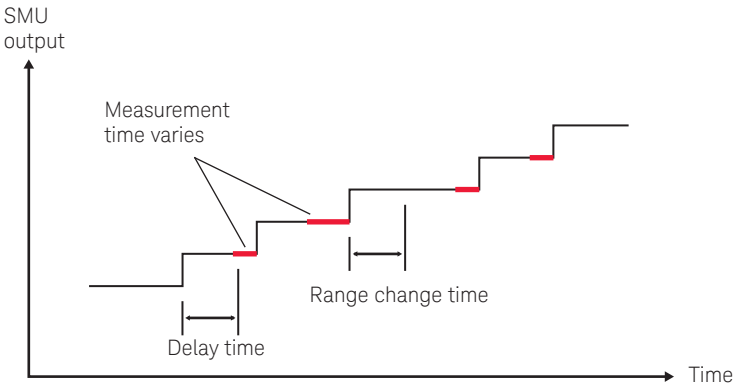


Figure 5.9. On a standard sweep measurement, the amount of time spent at each point in the sweep can vary greatly.

There are two factors that can impact the sweep measurement just shown. First, measurement time varies from one measurement point to the next. Second, if a range change occurs, then there will be a very long delay between measurement points.

There are ways to solve both of the above-mentioned problems. Parametric measurement instruments support a feature known as **Step Delay**. This feature allows you to specify the time that the instrument waits from the beginning of the actual measurement to when it increases to the next point in the sweep. Therefore, as long as the step delay is set to a value greater than that of the maximum expected measurement time, we can specify a constant total step time. Of course, this assumption is only valid if we do not encounter any range changes. To make sure that no range changes occur, fixed ranging should be used. The key point to remember is that the fixed range chosen must include the maximum expected value to be measured during the sweep.

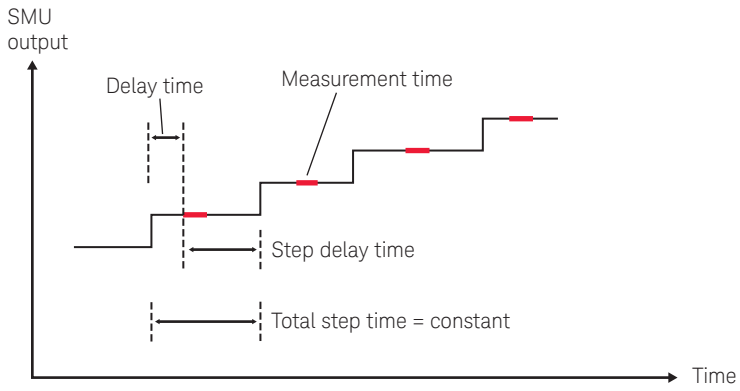


Figure 5.10. By using step delay and fixed measurement ranging, a constant sweep step time can be achieved.

Maintaining a constant sweep step *(continued)*

There are some reliability measurements where maintaining a constant sweep step time is very important. The most common of these measurements are the voltage ramp (VRAMP) and current ramp (JRAMP) tests. An example of a VRAMP test that uses the step delay feature to create uniform time steps for the voltage ramp is shown below.

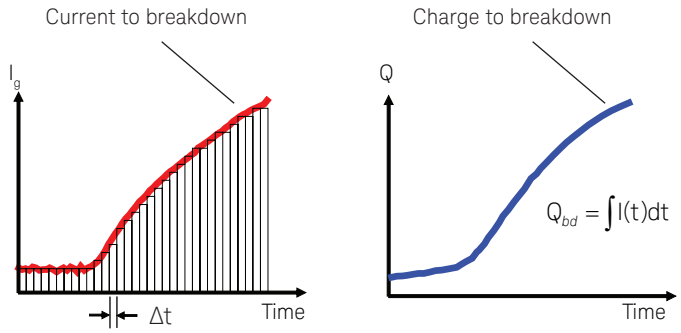


Figure 5.11. By using the step delay feature to create uniform time steps for a voltage ramp (VRAMP) sweep, it is very easy to calculate the total charge-to-breakdown (Q_{bd}).

By maintaining a constant time step for the voltage sweep, it is very easy to calculate out the charge-to-breakdown (Q_{bd}) from the measured gate current at each point in the sweep using a simple rectangular approximation:

$$Q_{bd} = \int I(t)dt \approx \sum I \cdot \Delta t$$

High speed test structure design

To succeed in making high speed measurements, you need more than just the correct measurement instrumentation. You also need to put sufficient forethought into the creation of the test structures that will be used to make the measurement. Attempts to make fast pulsed measurements with conventional DC test structures using DC positioners are unlikely to yield good measurement results. In general, fast pulsed measurements require test structures designed for a ground-signal (GS) or ground-signal-ground (GSG) measurement environment and RF positioners. The following figure illustrates this point.

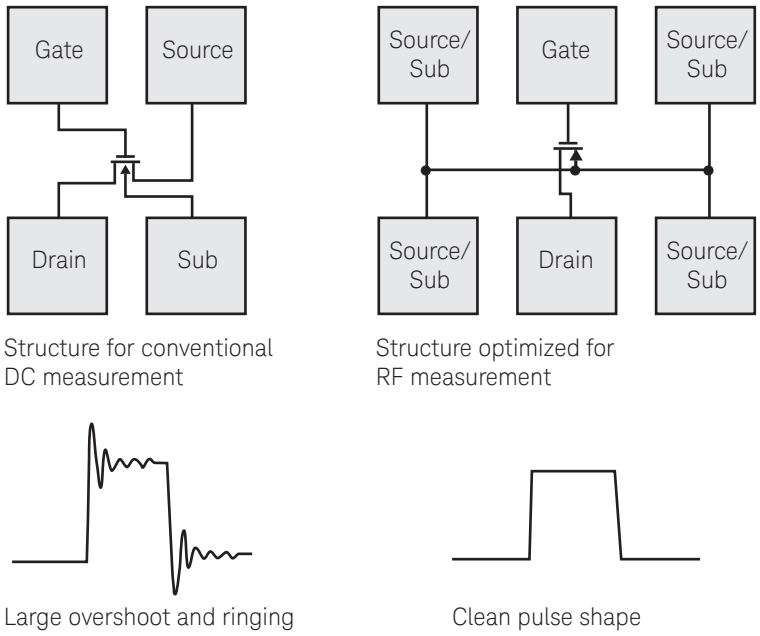


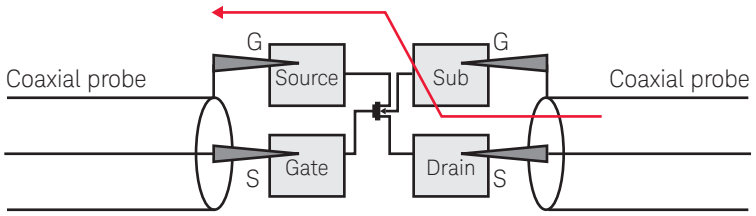
Figure 5.12. Illustration showing how conventional test structures using DC probes do not yield satisfactory results when attempting to make fast pulse measurements.

A photograph of a GSG RF probe tip is shown below.

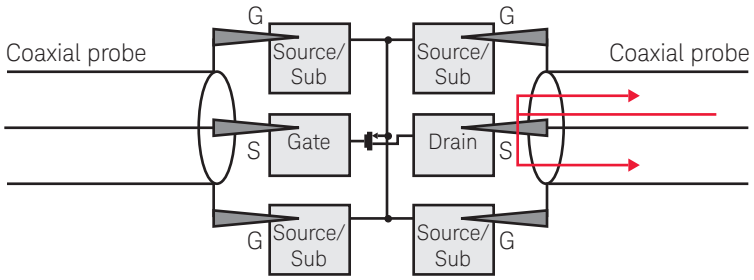
Figure 5.13. A GSG probe tip. Note that this probe tip uses an SMA style coaxial connection. Photo courtesy of FormFactor.

High speed test structure design *(continued)*

GSG probes do not necessarily produce superior measurement results to GS probes, since the results obtained also depend strongly on the pad layout and DUT structure. However, the following figure illustrates the reason why GSG probes can be superior to GS probes at high frequencies.



Some of the incoming signal energy must travel a long distance through the DUT to ground.
→ Large signal loss at high frequencies (not a concern for DC measurement)



All of the incoming signal energy has a short return path through the DUT to ground.
→ Signal loss over frequency is minimized.

Figure 5.14. GSG probes can provide superior performance at higher frequencies relative to GS probes due to their shorter return path for the incoming signal energy.

The basic limitation of GS probes is that the energy entering the DUT can have a potentially long path length to return to ground, which results in significant signal loss. However, GSG probes have a much shorter return path to ground, which minimizes the signal loss over frequency.

Another consideration is the bandwidth of the RF probes that you are using. Since the shortest pulse widths that we use in parametric tests are about 10 ns, at first thought it might seem that one or two hundred mega-hertz of bandwidth would be sufficient. However, this is actually not the case. Remember that a square wave (pulse) is actually the summation of an infinite series of sine waves (odd harmonics). You need to have enough bandwidth in your RF probes (and also in your overall system) to support the higher harmonics, or you will get a distorted pulse shape. Therefore, to produce a clean 10 ns pulse you typically need at least 1 GHz of bandwidth.

High speed test structure design *(continued)*

In addition to the type of RF probes chosen and their bandwidth, the physical layout of the DUT also has a big impact on the integrity of the waveform pulse. In particular, placing the drain and source connections on MOSFET on opposite sides of the layout creates extremely long signal paths that will cause lots of ringing and greatly distort the shape of the applied pulse as shown below.

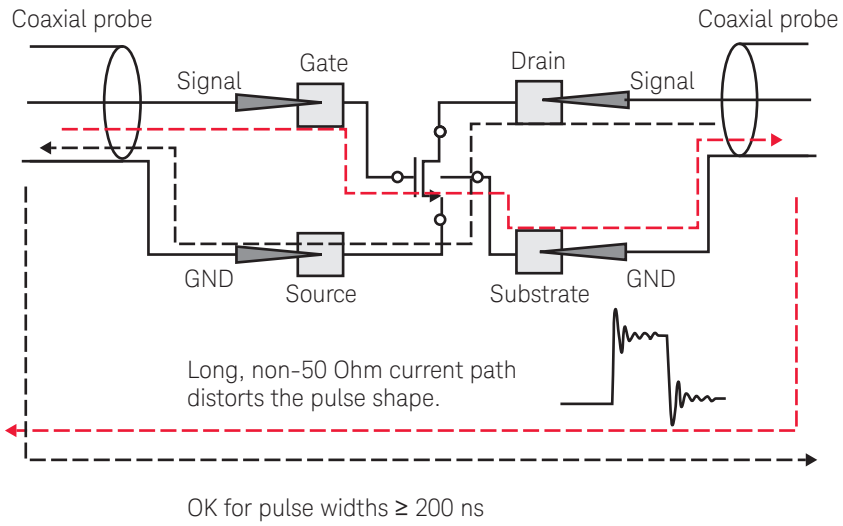


Figure 5.15. A GS probing arrangement with the drain and source on opposite sides can only support pulse widths of 200 ns or greater without causing pulse distortion.

Layouts with the drain and source terminals on opposite sides will generally only support pulse widths of 200 ns or greater. A much better arrangement is to place the source and drain on the same side of the layout as shown below.

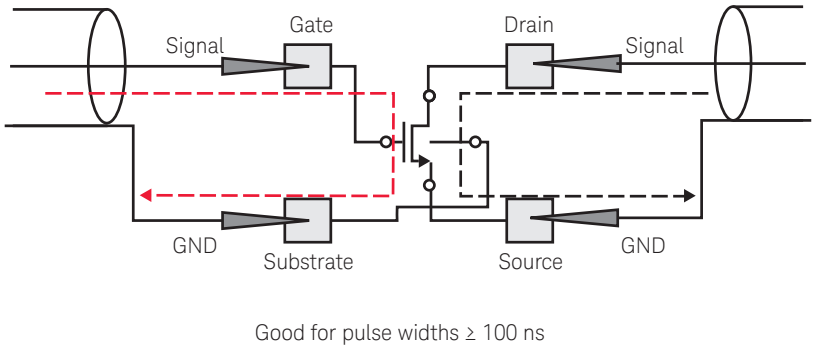


Figure 5.16. A ground-signal probing arrangement with the drain and source on the same side can support pulse widths down to around 100 ns.

Placing the source and drain on the same side creates a much shorter path for the signal energy, and a ground-signal pad layout can support pulse widths down to approximately 100 ns.

High speed test structure design *(continued)*

In order to achieve pulse widths down to 10 ns, both a ground-signal-ground pad layout and a structure design that shorts the source and substrate connections together are required. An example of this is shown below.

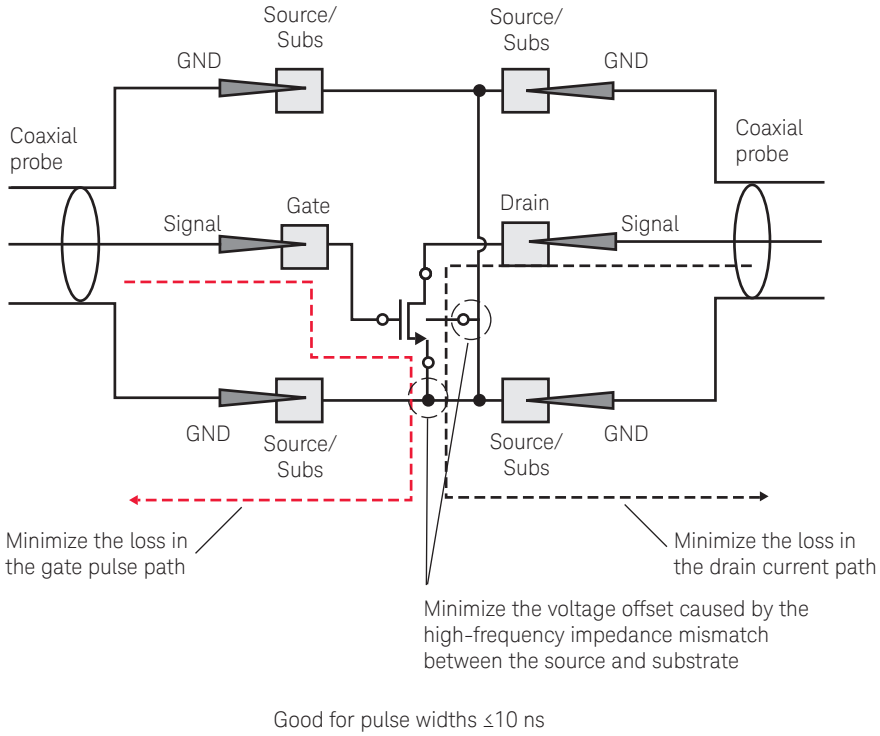


Figure 5.17. A ground-signal-ground probing arrangement with the source and substrate connections tied together can support pulse widths down to 10 ns or less.

Experimental results show that this layout scheme can reliably produce clean 10 ns pulses at the gate of the MOSFET.

Fast IV and fast pulsed IV measurements

Overview

Increasingly, many parametric measurements require some sort of fast IV or fast pulsed IV measurement capability. There are many factors driving this need, including new physical effects caused by decreasing lithographic and voltage scaling, the use of novel and exotic materials, and faster operating speeds that can generate large transient currents. Given this wide range of needs, it is difficult to create a universal solution that is both flexible and cost-effective. For example, some pulsed measurements may require large voltages but relatively slow pulse capability; others may require small voltages but relatively fast pulse capability. It is certainly possible to design a single solution that can do both but the cost of creating an all-in-one solution can become prohibitive (especially if it is important to support multiple measurement channels for parallel testing). In fact, it would be more efficient to support a range of solutions so that the user can select the one that best fits within their test needs and budget.

Until recently, solutions for fast pulsed IV measurements typically consisted of a precision pulse generator and oscilloscope as shown in the following illustration.

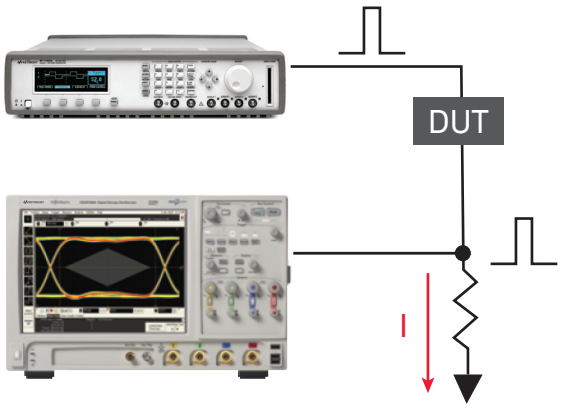


Figure 5.18. A conventional fast pulsed measurement solution using off-the-shelf instrument components.

In this solution, the current flowing through the DUT is calculated by measuring the voltage across a shunt resistor using an oscilloscope. However, this type of solution has several drawbacks.

1. It requires a precision resistor.
2. It requires some sort of software to integrate everything together.
3. It has limited accuracy.
4. It can only test one device at a time.
5. It requires a lot of external cabling.

While none of these factors are insurmountable, they can take many engineering hours of effort to overcome. It is typically more cost effective to use a solution that consists of self-contained hardware. In the following sections, we will look at different fast pulsed IV measurement solutions and discuss the types of measurement challenges for which they are best suited.

High-voltage semiconductor pulse generator unit (HV-SPGU)

The B1500A supports a single-slot, two-channel HV-SPGU module that can generate pulses with amplitudes of ±40 V (80 V peak-to-peak) into an open load and that can supply pulse widths from 5 µs to 10 s for pulsed IV measurement.

Note: Although the HV-SPGU module is capable of generating voltage pulse widths as narrow as 50 ns, when used in pulsed IV measurements the HV-SPGU's pulse width is limited by its minimum current measurement interval (which is 5 µs). A photograph of an HV-SPGU module (2 output channels total) is shown below. Up to five HV-SPGU modules can be installed in a single B1500A, for a maximum of ten channels. Synchronization between the modules is maintained using external connections between the "Ref Out", "Ref In", "Sync Out", and "Sync In" outputs and inputs.

Figure 5.19. The B1500A's HV-SPGU module has ±40 V output capability and supports two channels per module.

Although this pulse generator module is ideal for its many features for flash memory testing, it also can be used for fast pulsed measurement. In order to improve accuracy, the HV-SPGU has a built-in voltage monitoring capability. The output voltage monitoring capability permits the actual voltage at the output to be measured. A simplified schematic of the HV-SPGU module is shown below.

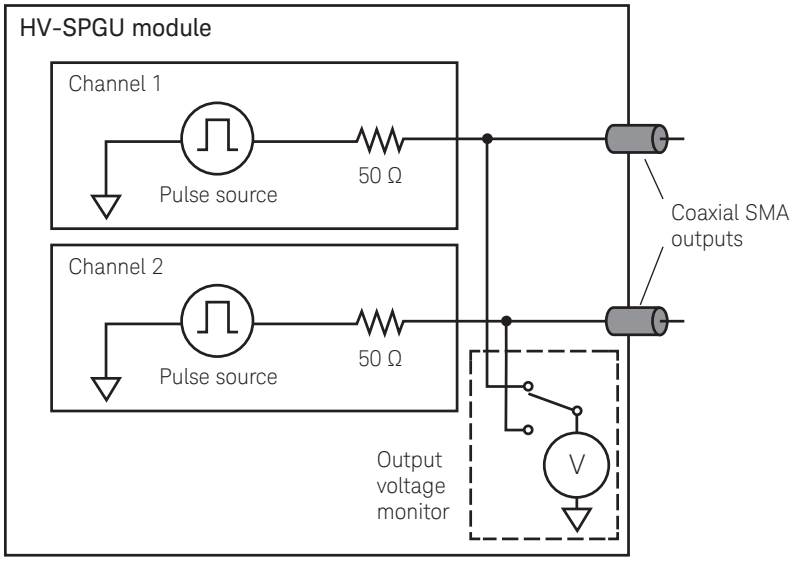


Figure 5.20. Simplified HV-SPGU schematic showing the shared output voltage monitor circuit.

High-voltage semiconductor pulse generator unit (HV-SPGU) *(continued)*

Basic pulsed IV measurement on a MOSFET can be performed using only a single HV-SPGU module by connecting the drain to one channel and the source to another as shown below.

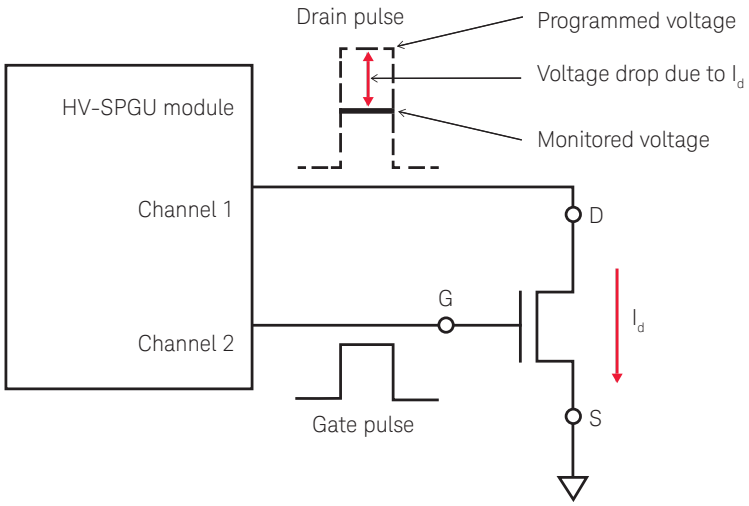


Figure 5.21. Performing pulsed IV measurement on a MOSFET using a single HV-SPGU module.

By determining the difference between the programmed voltage and the monitored (measured) voltage, and by making use of the fact that the output impedance is a well-controlled and known value (50 Ω), it is possible to determine the current flowing through the HV-SPGU output using a simple calculation.

Although this calculation is simple in principle, obtaining an accurate value for current drain, I_d, can be challenging in practice. To simplify this, Keysight supplies a library of application tests for performing pulsed IV measurements using the HV-SPGU. The most basic of these application tests permits you to specify single (spot) pulsed measurements on two HV-SPGU channels. The setup screen for this application test is shown below.

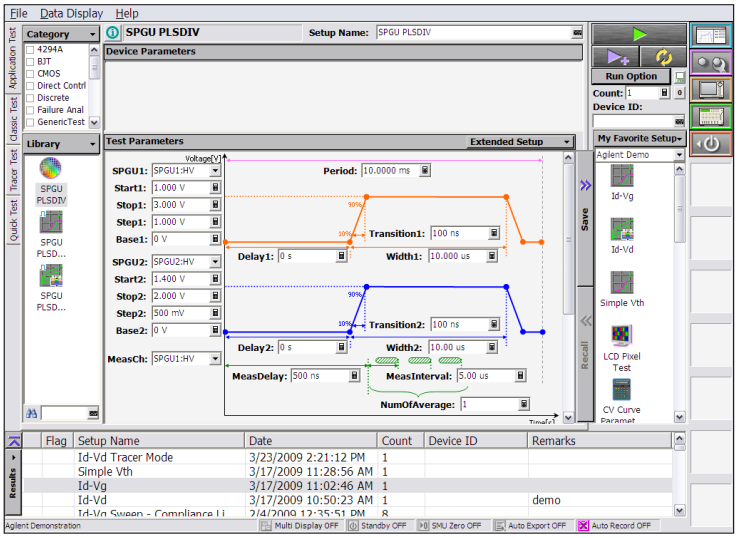


Figure 5.22. The basic B1500A EasyEXPERT application test for performing pulsed IV measurement using the HV-SPGU module.

High-voltage semiconductor pulse generator unit (HV-SPGU) *(continued)*

By itself, this application test is not particularly useful; however, it can be used within other application tests to create more complicated pulsed IV application tests. Using this technique, Keysight supplies the two most commonly needed application tests to perform pulsed I_d-V_g and I_d-V_d sweeps. An example of the input screen for the pulsed I_d-V_g application test is shown below.

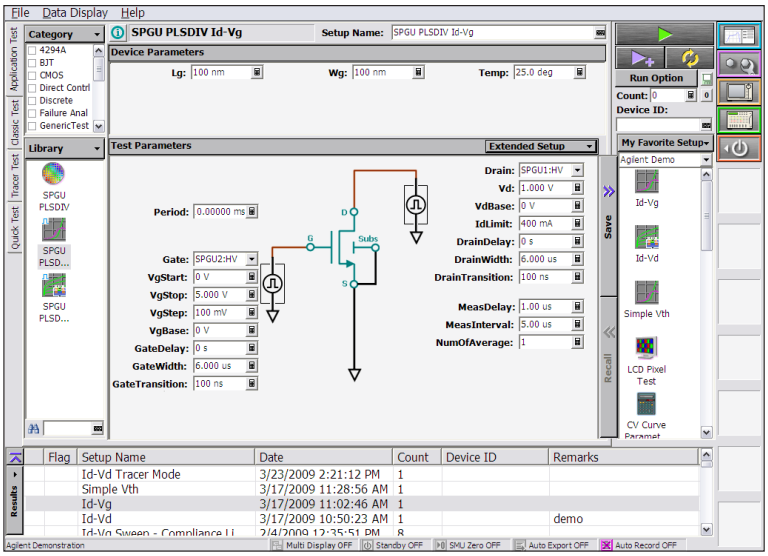


Figure 5.23. A sample B1500A EasyEXPERT application test to perform fast pulsed I_d-V_g measurement using the HV-SPGU module.

The input screen for the I_d-V_d application test is similar.

Due to its large voltage output capability, the HV-SPGU module is a good choice for testing devices in the "medium" power range (such as GaAs and HEMT devices used in RF applications). In addition, it is a good choice for measuring transient currents in circuits where more than 10 V of amplitude are required.

Waveform generator/fast measurement unit (WGFMU)

The B1500A supports a single-slot, two channel WGFMU module for pulsed IV measurement that can generate pulses with amplitudes of ±3 V, ±5 V, 0 V to 10 V or –10 V to 0 V and that can supply pulse widths from 50 ns to virtually any maximum length. Up to five WGFMU modules can be installed in a single B1500A, for a total of ten channels maximum. Similar to the case of the HV-SPGU module, synchronization between the WGFMU modules is maintained using external connections between the "Sync Out", and "Sync In" outputs and inputs. A photograph of the module outputs is shown below.

Figure 5.24. The B1500A's WGFMU module can output 10 V peak-to-peak and has a minimum pulse width of 50 ns.

Waveform generator/fast measurement unit (WGFMU) *(continued)*

The B1500A's WGFMU solution consists of the WGFMU module as well as two remote-sense and switch units (RSUs) connected to the WGFMU module via furnished cables. The WGFMU module contained in the mainframe generates the arbitrary waveforms, and these waveforms are then transmitted through the cables to the RSUs. The RSU, which performs the actual current or voltage measurement, is separate from the WGFMU module so that it can be placed near the device under test (DUT) to minimize cable lengths and guarantee accurate high-speed measurement.

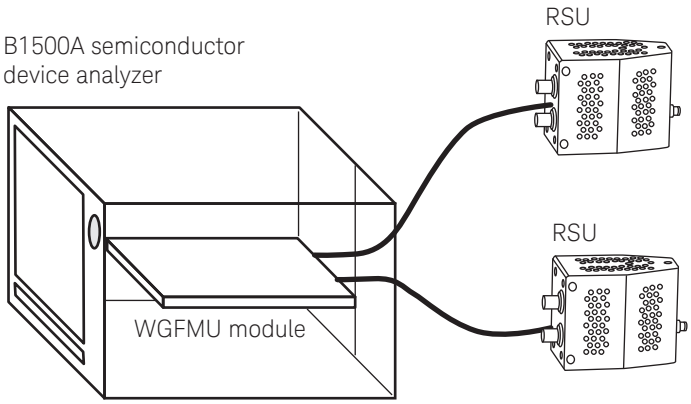


Figure 5.25. Each single-slot WGFMU module has two channels, and each channel has an RSU associated with it.

Each WGFMU channel has independent arbitrary linear waveform generator (ALWG) voltage generation capability and two user selectable operation modes: PG mode and Fast IV mode. The PG mode combines a very fast voltage measurement capability with 50 Ω output impedance to minimize waveform reflections. The Fast IV mode has a slightly slower measurement speed and slower waveform rise/fall times than the PG mode, but it can measure both current and voltage. The figure shown below shows a simplified circuit diagram of one channel of the WGFMU and RSU combination.

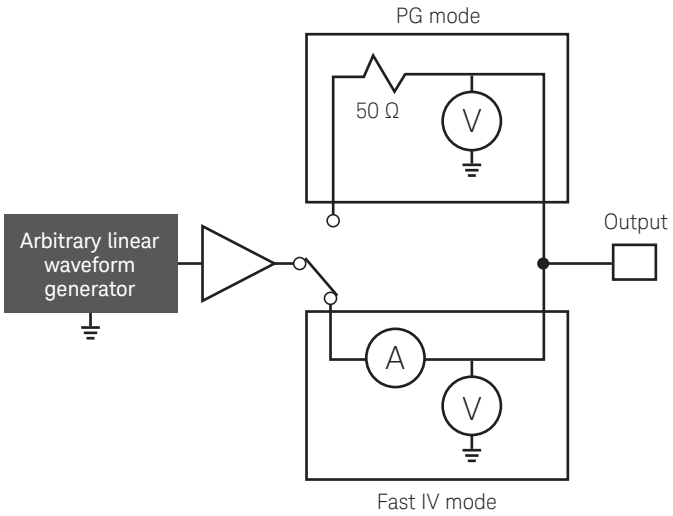


Figure 5.26. A simplified schematic of one channel of the waveform generator/fast measurement unit (WGFMU) and remote-sense and switch unit (RSU) combination.

Waveform generator/fast measurement unit (WGFMU) *(continued)*

In addition to its high-speed measurement capabilities, each RSU has the ability to multiplex its output between the WGFMU and an optional SMU. Even though the triaxial to coaxial cable conversion does not permit measurements to the full limit of the SMU specifications, this switching capability does permit the user to verify DC measurement results made using the WGFMU against those of an SMU. Moreover, each RSU also has a coaxial output that can be connected through an internal buffer circuit to the SMA coaxial output; this permits monitoring of the RSU output waveform with an external instrument such as an oscilloscope. A schematic illustrating these features is shown below.

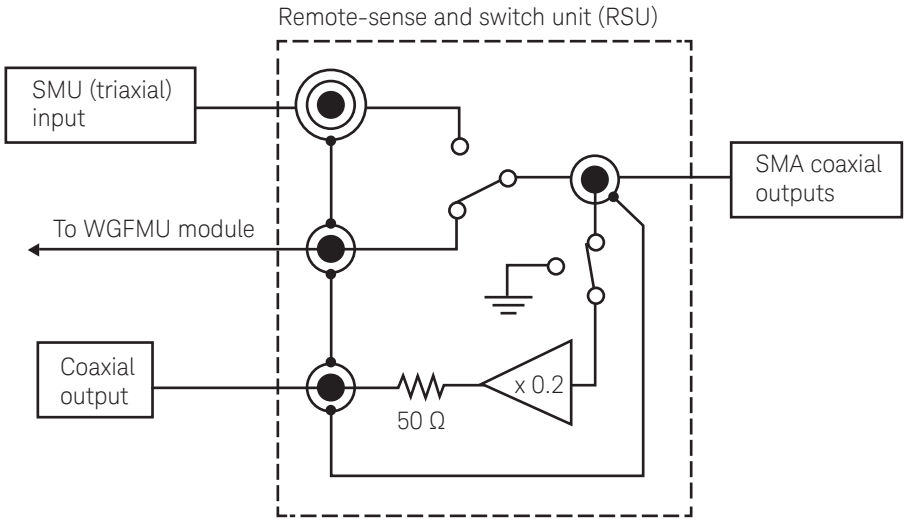


Figure 5.27. The RSU has features that allow it to work with an optional SMU and with an optional external monitoring device such as an oscilloscope.

A photograph of the RSU module inputs is shown below.

Figure 5.28. A photograph of the RSU module showing its voltage monitor (coaxial) output, SMU (triaxial) input and connector input for the cable that goes to the WGFMU module.

Waveform generator/fast measurement unit (WGFMU) *(continued)*

The WGFMU module has the ability to not only generate arbitrary waveforms, but also to measure either current or voltage very quickly at any point on those waveforms. In addition to this powerful measurement capability, the WGFMU module also allows the user to dynamically change the measurement range to optimize measurement accuracy. Of course, the user can also specify the sampling rate and averaging time. The following diagram illustrates all of these features.

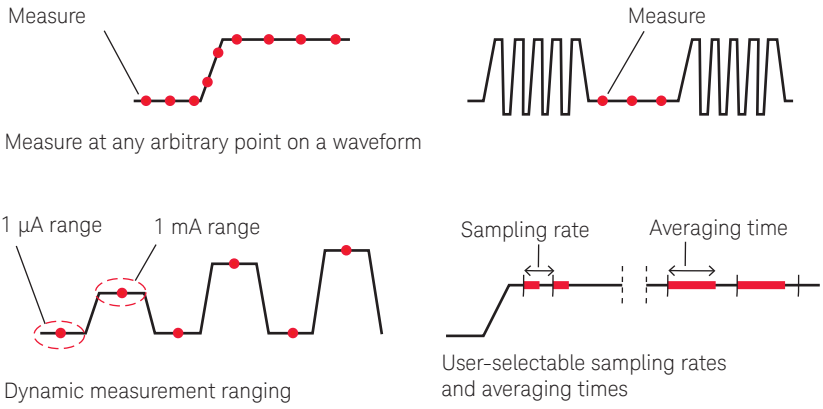


Figure 5.29. The key measurement feature of the WGFMU is its ability to measure at any point on an arbitrarily generated waveform.

An I_d-V_g measurement made with the WGFMU module using three different measurement conditions is shown below.

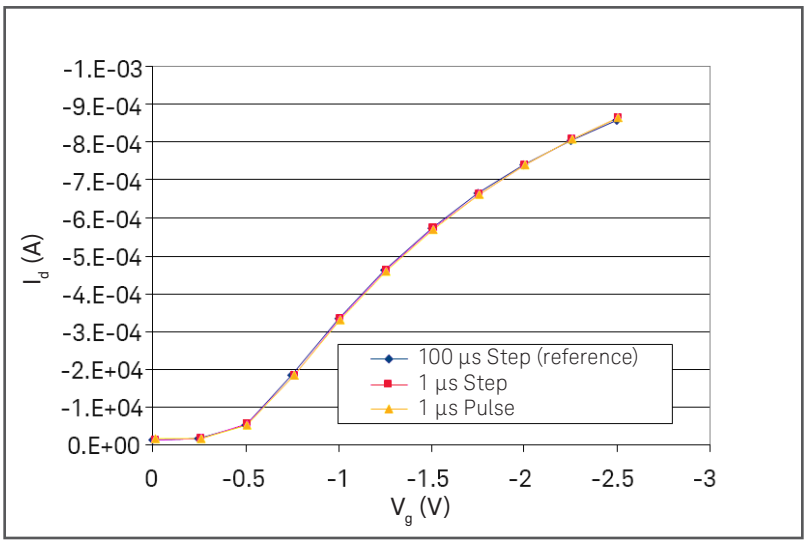


Figure 5.30. The fast pulsed I_d-V_g sweep measurement (~1 µs per point) made using the WGFMU module correlates well with the non-pulsed, slower measurements.

Waveform generator/fast measurement unit (WGFMU) *(continued)*

To create a baseline, a staircase sweep measurement with 100 µs per step (50 µs delay) was performed first. Next, a staircase sweep measurement with 1 µs per step (500 ns delay) was performed. Finally, a pulsed sweep measurement with a 1 µs pulse (100 µs rise/fall times, 500 µs delay, 2 µs period) was performed. These measurement results confirm that the WGFMU module can indeed make reliable high-speed pulsed measurements.

The WGFMU module is an ideal solution for many types of reliability test that require fast measurement capability. Not only can the WGFMU module measure very quickly, but its arbitrary linear waveform generation capability makes it very easy to create complicated AC stress conditions. The following figure shows examples of reliability measurements made with both DC and AC stress conditions.

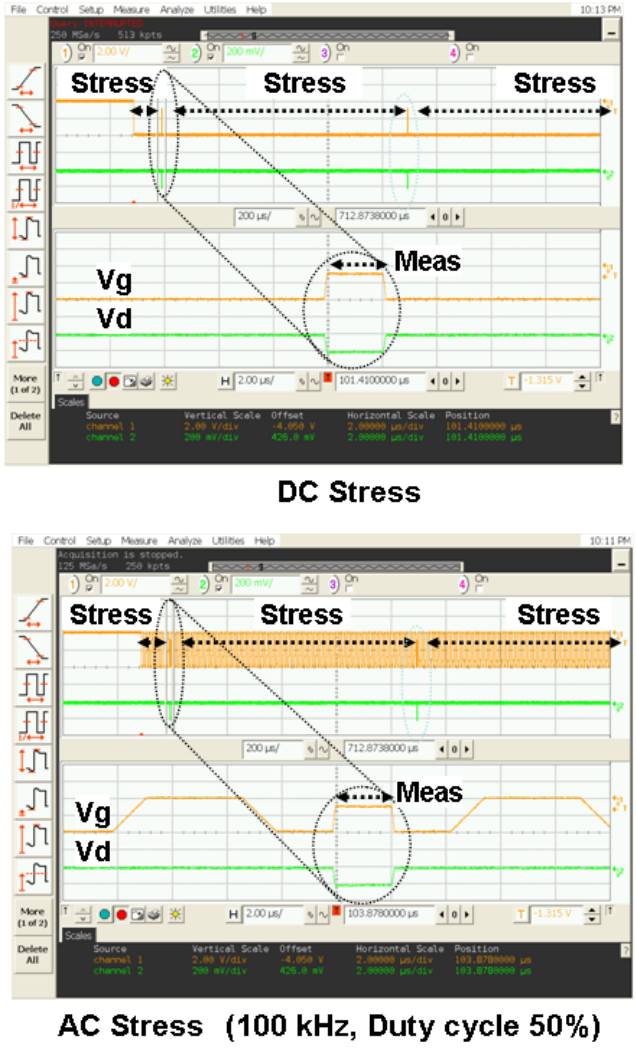


Figure 5.31. The WGFMU can easily apply both DC and AC stress, and there is no delay in transitioning from stress to measure and there is no glitching.

The advantage of the WGFMU module is that the transition from stress to measure and back to stress again is virtually seamless.

Waveform generator/fast measurement unit (WGFMU) *(continued)*

Similar to the HV-SPGU module, Keysight supplies a library of application tests to control the WGFMU module.

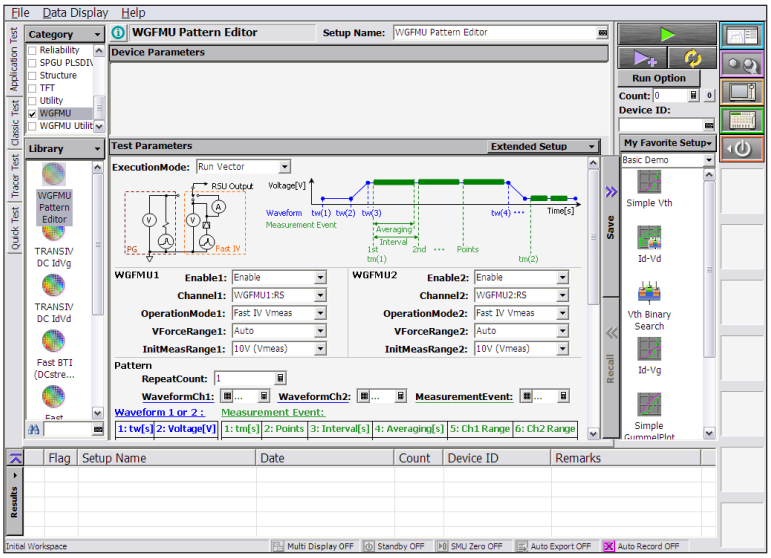


Figure 5.32. The WGFMU module pattern editor application test allows you to control the WGFMU from within EasyEXPERT.

An ideal application of the WGFMU module is ultra-fast NBTI testing.

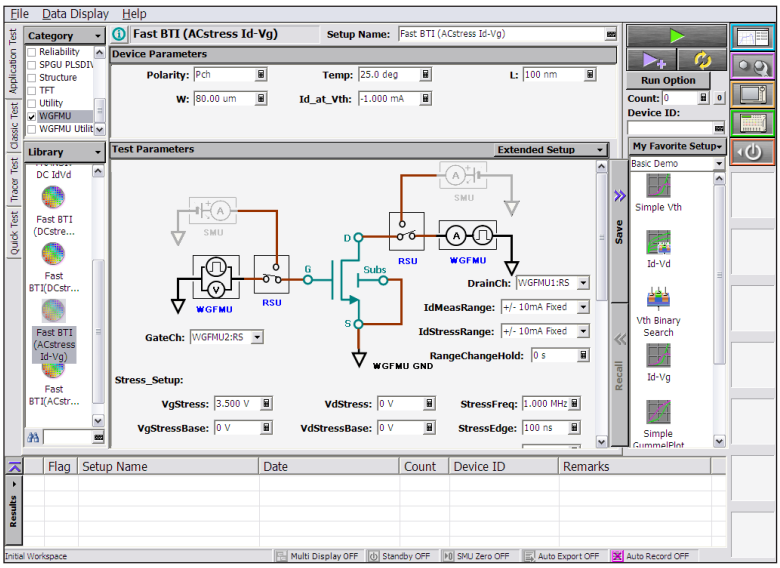


Figure 5.33. The fast BTI application test comes in versions that support both DC and AC stress (the AC stress version is shown here).

For complicated stress and measurement reliability testing, it is difficult to control the WGFMU using an application test due to all of the various parameters that need to be set. In these cases, it is better to control the B1500A remotely using an external controller and software. Keysight can supply the necessary API to control the WGFMU module using a variety of third party software.

Random Telegraph Noise (RTN) Measurement

The B1500A's WGFMU can make random telegraph noise (RTN) measurements. A random telegraph signal has the following properties.

1. It has two values (±1), i.e. X(t) = ±1.
2. The number of zero crossings in any interval (0, t) is described by a Poisson process.

The following plot shows an example of a random telegraph signal.

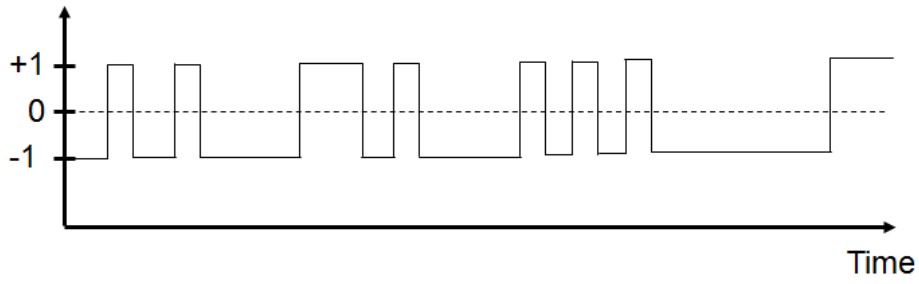


Figure 5.34. The general characteristic of a random telegraph signal.

In a semiconductor MOSFET, random telegraph noise manifests itself as distinctive variations in the drain current over time even while the gate and drain voltages remain constant. The cause of RTN is attributed to electron trapping and de-trapping, and while it has always been present in semiconductor MOSFET devices, its effects were not relevant until recently. The reason for RTN's current importance is that as lithography has shrunk and voltage margins decreased, RTN can now affect the stability of SRAM cells. This of course has a profound impact on circuit reliability.

Since the time constant for RTN electron trapping or de-trapping is very small, measurement equipment with the ability to sample current in nanoseconds is required. The WGFMU module is an ideal solution for this, as it has two independent channels so both the MOSFET gate and drain voltages can be biased while the drain current is measured. The following figure shows the RTN measurement setup.

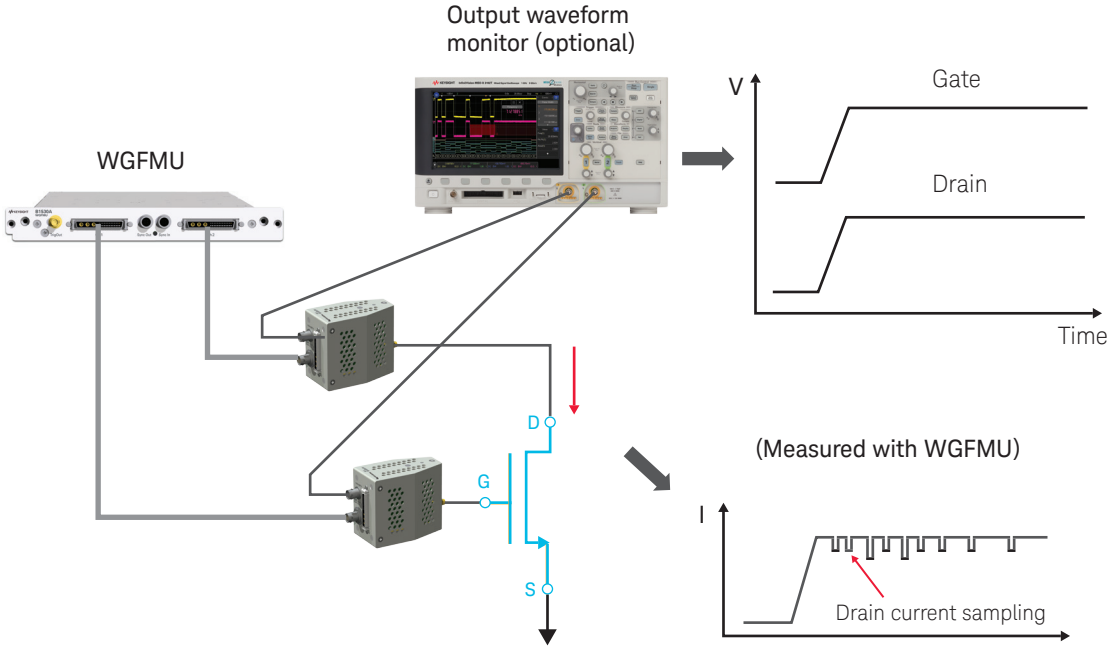


Figure 5.35. Measuring RTN using the B1500A's WGFMU module.

While RTN measurement is not fundamentally complex, it can be fairly difficult to detect because it only shows up in a fraction of devices and it is highly dependent on the gate to source biasing. A change in the gate to source biasing of 100 mV (or less) can often make RTN appear or disappear. The following figure shows an example of RTN measured on a CMOS transistor using the WGFMU module.

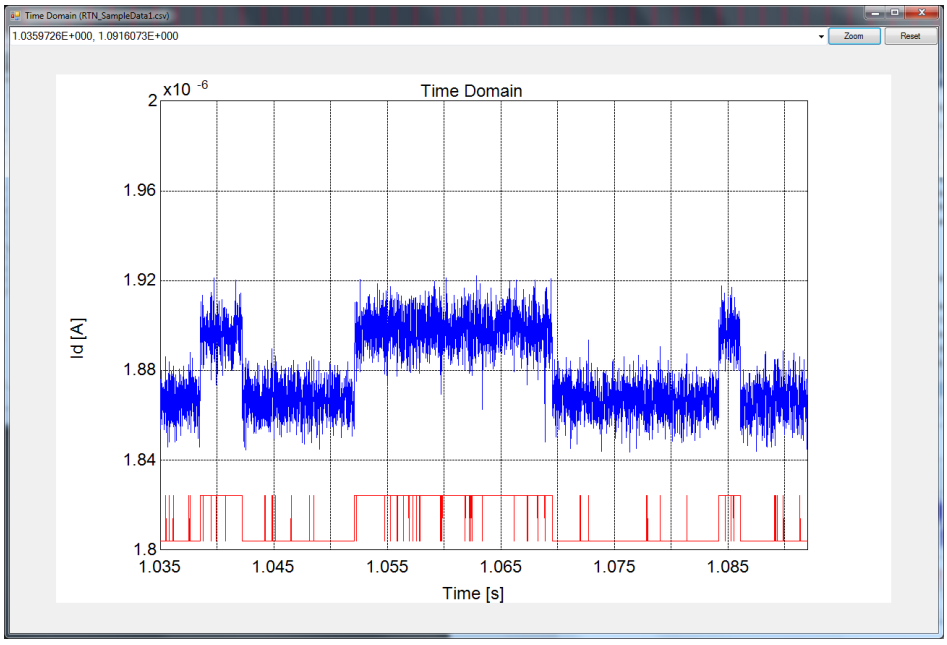


Figure 5.36. CMOS transistor RTN example data.

The best way to measure RTN is to measure a large number of devices across a wafer at different gate to source biases. The data can then be examined for the characteristic RTN signature.

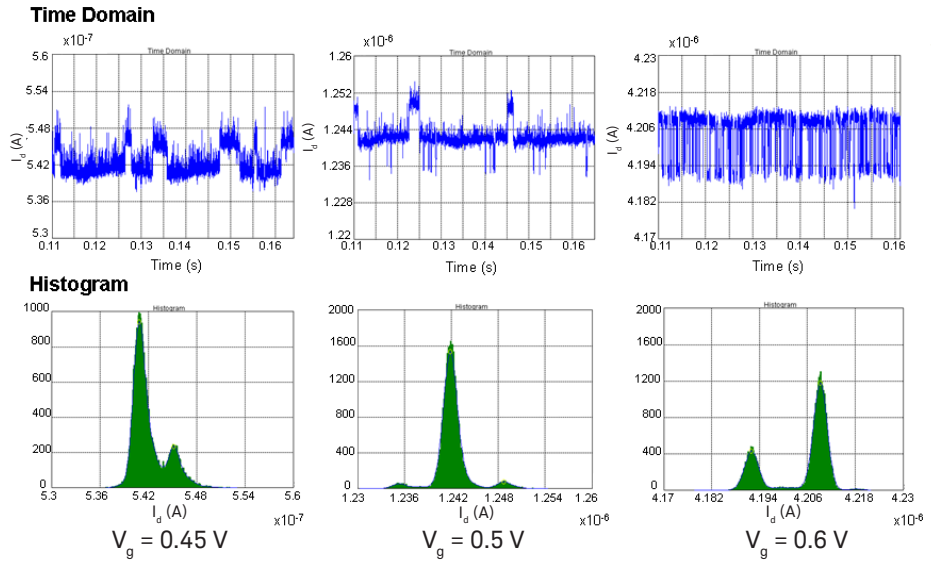


Figure 5.37. Variations in RTN characteristics as a function of gate to source voltage.

Summary of Keysight pulsed IV measurement solutions

Given the number of choices available, it is clear that a number of factors need to be weighed against one another when selecting a pulsed IV solution. The following steps are listed as a guide to help you with this process.

1. Make sure that you understand your measurement requirements.
   The key factors to consider are listed below:
   - Range of pulse widths
   - Required current measurement resolution
   - Maximum voltage and current output capability
   - Multiple pulsing capability

   Note: Make sure that you understand that some solutions only work for specific device types and configurations.
2. Determine which solution or solutions meet your pulse width requirements, taking into consideration both your future as well as current needs.
3. If more than one solution meets your pulse width requirements, then select among the solutions using the other measurement parameters (current measurement resolution, current/voltage output capability, etc.). Keep in mind that there may be some trade-offs among these various parameters (such as accuracy versus voltage/current output capability).
4. Once you have decided upon a solution, re-verify all of the specifications of that solution to make sure that it meets the measurement needs of your applications and devices.

As we discussed earlier in this chapter, SMUs can also make pulsed measurements. Although the SMU pulsing capabilities are slow relative to the three solutions discussed in this section, in certain cases (such as the on-the-fly NBTI measurement) SMU pulsed measurement can provide a very satisfactory and cost-effective measurement solution. Therefore, SMUs need to be included in any comparison of available pulsed measurement solutions. The following table summarizes the key specifications of Keysight's low-power pulsed measurement solutions.

Pulsed measurement solution	SMU	HV-SPGU	WGFMU
Supported instruments	4155, 4156, B1500A	B1500A	B1500A
Maximum voltage	100 V[1,2] /200 V[3]	±40 V	–10 V to 0 V or 0 V to 10 V
Maximum current	100 mA[1,2]/1 A[3,4]	400 mA	10 mA
Current resolution	1 fA[2]/10 fA[1,3]	40 µA	0.014% of range[5]
Pulse width range	50[4]/500 µs – 2 s	5 µs – 10 s	50 ns (min.)[6]
SMU measurement support	Yes	No	Yes
Waveform monitor function	No	No	Yes
Multi-channel pulsing	No	Yes	Yes
External equipment required	None	None	None
Connection	Direct	Direct	Direct

1. MPSMU
2. HRSMU
3. HPSMU
4. MCSMU
5. Requires averaging; see data sheet for more details.
6. The maximum pulse width is difficult to specify for this module, but it can easily be 20 million seconds (effectively infinite).

Figure 5.38. Table summarizing Keysight's low-power pulsed measurement solutions.

Chapter 6 Making Accurate Resistance Measurements

"It requires a very unusual mind to undertake the analysis of the obvious."
— Alfred North Whitehead

Resistance measurement basics

Upon first thought, it might seem strange to devote a chapter only to resistance measurement. After all, resistance is governed by the simplest of all rules (Ohm's law) that every electrical engineering student learns in their first week of class:

$$V = I \times R \qquad \text{(Equation 6.1)}$$

However, despite the simplicity of this equation, it turns out that measuring resistance accurately is actually one of the more challenging areas in parametric tests. The reason for this is that the above equation is overly simplistic, and thus ignores the fact that resistance generates heat, which in-turn affects the value of the resistance itself. Therefore, it would be more accurate to re-write the above equation as:

$$V = I \times R(T) \qquad \text{(Equation 6.2)}$$

In this equation the resistance (R) is a function of temperature (T). This phenomenon, whereby the actual value of the resistance being measured changes due to heat generated by the current flow, is commonly referred to as the Joule self-heating effect.

Another factor requiring consideration is the resistance of the cables used to make a resistance measurement. When attempting to measure very small resistance values, Kelvin measurement techniques must be employed. The basics of Kelvin measurements have been explained in earlier chapters, and it is straightforward to apply these techniques to measuring a resistor. However, it is worthwhile to point out that the combination of Joule self-heating effects and cable resistance make resistance measurements especially challenging. To reduce Joule self-heating, you need to reduce the current (power) flowing into the resistor being measured. However, small currents flowing through the cable resistance require measurement equipment with very accurate voltage measurement capabilities in order to distinguish the cable voltage drop from the voltage drop across the actual resistor. For these reasons resistor measurements can easily require a voltage measurement resolution capability of less than 1 millivolt.

The final factor to consider when making resistance measurements is electro-motive force (EMF). EMF is the technical name for the burst of electrical noise that is generated when a mechanical relay opens or closes. All SMUs have mechanical (reed) relays in their output paths that generate EMF when the SMU is activated. EMF can have significant impact upon measurement accuracy for all types of measurements, but it can especially impact low-level resistance measurements because they require extremely sensitive voltage measurement resolution.

Floating vs. grounded resistance measurements

When making a resistance measurement, the first fundamental choice you must make is whether to perform the measurement in a floating or grounded configuration. As you can see in the following figure, the connections for these two cases are quite different.

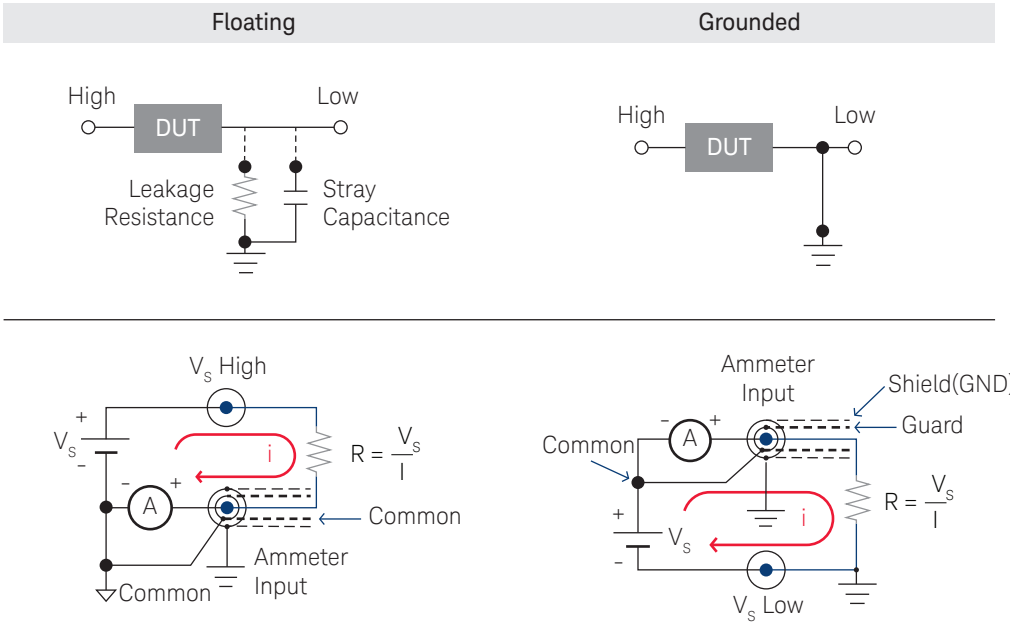


Figure 6.1. Floating and grounded resistance measurement configurations.

On the top left, you can see the case where the device under test (DUT) is floating with respect to earth ground. In this situation, the resistance between the high terminal and the low terminal is measured. Note that although the low side is floating, parasitic resistances and capacitances may provide a “sneak path” to ground. This is acceptable as long as a DC bias source is connected to the low terminal. The bottom left shows the circuit diagram corresponding to a floating device measurement. The test device is connected between the voltage source (V_s High) output and the ammeter input. Since the ammeter measures very low currents and is very noise-sensitive, it is located close to ground potential to shield the test device for better measurement results.

On the top right, you can see the case where the DUT is grounded. Since the low side is grounded, the applied test voltage and the current measurement must both occur at the DUT's high side terminal. The bottom right shows the circuit diagram corresponding to a grounded device measurement. In this configuration, the ammeter is connected to the voltage source (V_s) positive output, because the device is grounded on one side.

Note: Neither one of these configurations is necessarily “better” than the other, and you can obtain good high resistance measurement results using both setups.

Resistivity

Resistivity is a basic material property. If we measure the current flowing through a bar of homogeneous material with a uniform cross section when a voltage is applied to it, we can determine its resistance from the equation R = V/I. The related material property of resistivity can be calculated from the resistance measurement if the cross-sectional area (width x depth) and length of the sample are known as shown below.

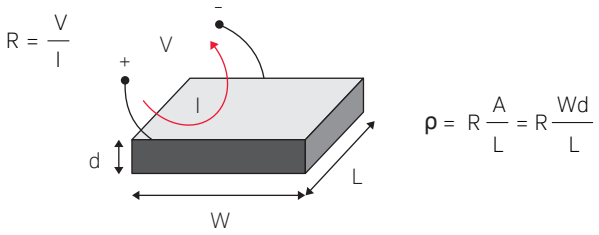


Figure 6.2. Calculating resistivity from a resistance measurement.

Sheet resistivity is typically measured for all the implantation layers as well as the metal interconnect layers. It is common to see resistivity expressed in “Ohms per square”, in recognition of the fact that the resistance from one side of a square of homogeneous material of uniform depth to the other side is always the same regardless of the size of the square. A little reflection on the effect of combining squares of material of uniform thickness together into larger squares should convince the reader of this fact.

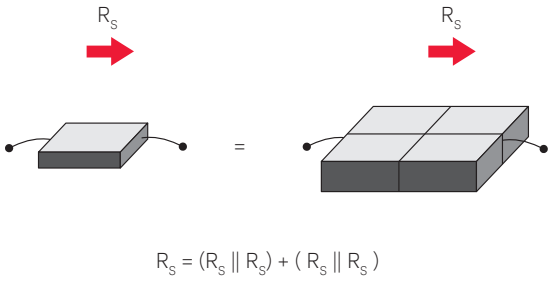


Figure 6.3. The Ohms per square of a homogenous material of uniform thickness is always the same regardless of the size of the square.

Van der Pauw test structures

The resistances measured when making resistivity measurements can be extremely small (on the order of milliohms), which mandates the use of Kelvin measurement techniques. Therefore, it is extremely convenient to be able to combine the Kelvin measurement technique with simple geometry to determine the sheet resistance of a given material. The configuration universally used to achieve this is known as a Van der Pauw structure, named after L.J. Van der Pauw who first proposed this measurement technique in 1958. The Van der Pauw structure consists of a square sample with an electrical contact at each of the corners as shown below.

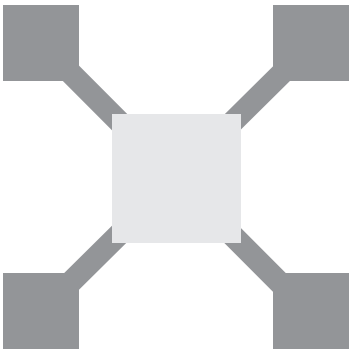


Figure 6.4. The Van der Pauw test structure for determining resistivity.

Van der Pauw showed in his original paper that the following relationship holds:

$$\exp\left(\frac{-\pi d R_{AB,CD}}{\rho}\right)+\exp\left(\frac{-\pi d R_{BC,DA}}{\rho}\right)=1$$

Here d is the thickness of the sample, p is the sheet resistance of the sample, and the resistance measurements are defined as shown below.

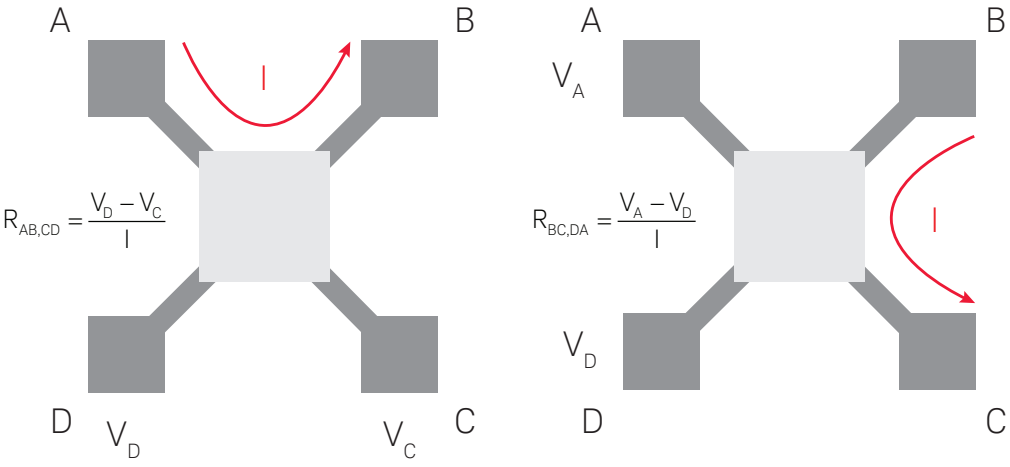


Figure 6.5. Calculating sheet resistance from a Van der Pauw test structure.

The calculation of the resistivity from this equation and these measurements is straightforward.

Accounting for Joule self-heating effects

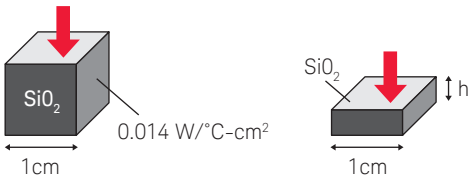
Joule self-heating limits the accuracy of resistance measurements because it changes the value of the resistor being measured. However, once you select the maximum level of resistance change that can be tolerated, you can translate this into maximum allowable power dissipation. We know that the equation for power dissipated in a resistor is given by:

$$P = V \times I = V \times \frac{V}{R} = \frac{V^2}{R}$$

By rearranging the above equation and solving for voltage, we can convert the maximum power dissipation into a minimum acceptable voltage resolution necessary to meet the resistance measurement requirements. The equation describing this relationship is given by:

$$V_{max} = \sqrt{P_{max} \times R(T)}$$

A simple example illustrates how this process works. The maximum allowed power dissipation is determined by the thermal resistance of the underlying silicon dioxide layer. Silicon dioxide (SiO_2) has a thermal conductivity of 0.014 W/°C-cm^2. We can use this to calculate the thermal resistance as shown below:



Thermal Conductivity

(h x 71) °C-cm^2/W, or
0.007 °C-cm^2/W for h = 1µm

Thermal Resistance

Figure 6.6. The thermal properties of silicon dioxide.

This means that if you apply 1 W to a 1 cm square metal structure of this thickness the temperature will rise by 0.007 °C; if you apply 1 W to a 10 µm square metal structure of this thickness then the temperature will rise by 7000 °C. Of course, in this example the material on top of the SiO_2 would vaporize long before a temperature of 7,000 °C is reached. This example is here to emphasize the fact that one cannot arbitrarily increase the power (=current) into a resistor in an attempt to compensate for insufficient accuracy in the measurement hardware. At or near room temperature, the resistance of a copper (Cu) or aluminum (Al) metal line changes by about 0.35%/°C. For the case shown above, we can compute the maximum amount of power that can be dissipated to produce a 0.1% change in resistance for a Cu or Al metal square 10 mm by 10 mm.

$$0.1\% = P_{max} \times 0.007°C - cm^2 / W \times 0.35\% / °C \times 1/(10\ mm)^2 \times (10\ mm / 1\ cm)^2$$
$$\therefore P_{max} = 0.04\ mW$$

To achieve 0.1% accuracy in a copper Van der Pauw structure with an equivalent resistance of 10 mΩ per square (1 mm thick film) we have the following:

$$V_{max} \sqrt{(0.04\ mW) \times (10\ m\Omega)} \approx 0.000632\ V$$

In other words, the measurement instrumentation must be able to resolve down to about 1 mV!

Note: The Van der Pauw structure is a relatively extreme example. A metal line 1 mm wide and 100 mm long typically requires the measurement instrumentation to be able to resolve to the tens of micro-volts in order to obtain 0.1% accuracy. Since we know that the maximum allowable voltage is proportional to the square root of (P_{max} x R), the required voltage resolution increases proportionally to the square root of the resistance value. Since this is a relatively weak dependence on R, 10 mV is a good estimate of the measurement resolution requirement for most resistance measurements.

Eliminating the effects of electro-motive force (EMF)

What is EMF?

Electro-motive force (EMF) is the name given to the transient voltage pulse created when reed relay switches open and close. Conventional reed relay switches, which can be obtained from a variety of sources, typically generate a thermo-EMF ranging from a few tens of micro-volts to a few hundreds of micro-volts after the relay activation current is turned on or off. This voltage drift, which can continue for several minutes before dying out, is not acceptable when making precision measurements such as those required for parametric tests. The figure shown below is an example of the thermal-EMF generated by a commercially available reed relay.

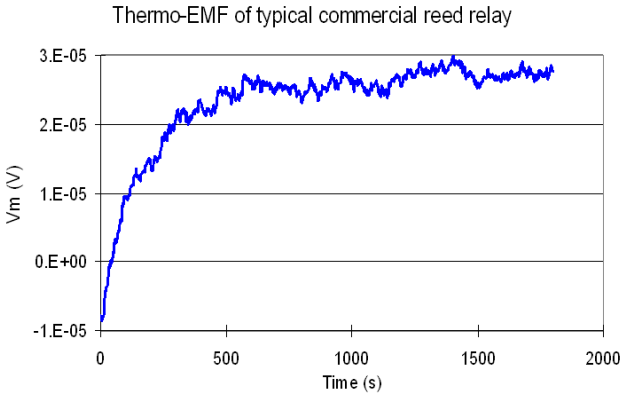


Figure 6.7. This graph shows the thermo-EMF generated by a typical reed relay.

It is important to note that the above graph is not characteristic of the reed relays used in parametric measurement instruments (for obvious reasons). Keysight Technologies has developed a proprietary reed relay technology known as the “Cool Guard” relay. These relays are used in parametric measurement equipment to minimize the effects of EMF, thereby improving both throughput and measurement accuracy. The characteristics of a Cool Guard relay are shown below.

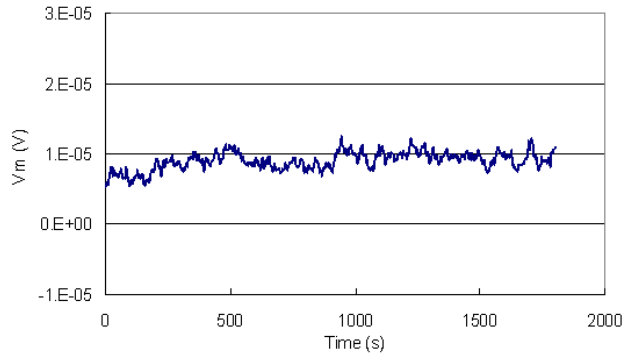


Figure 6.8. This graph shows the thermo EMF generated by a Keysight Cool Guard relay.

What is EMF? *(continued)*

The SMU schematic discussed previously in chapter 3 left out the output relay switch for purposes of simplicity. However, it is important to understand that all SMUs have a switch in the output path, and that this switch generates EMF when opened or closed. The revised SMU schematic is shown below.

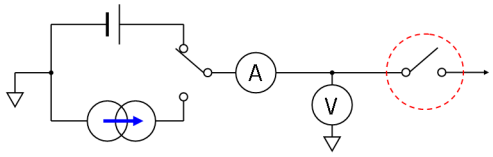


Figure 6.9. SMU schematic showing the output connection relay.

When using the SMUs from the instrument front panel, the software takes care of opening and closing this relay so its presence is invisible to the user. However, when using instrumentation under program control, the user must issue a connect (“CN”) command in order to close this relay, and the relay remains closed until a command is issued (such as “*RST”) to open the relay(s).

Mitigating the effects of EMF

Reducing the impact of EMF on parametric measurements is relatively straightforward. The data sheet specifications take thermo-EMF into account, so under normal measurement conditions you do not need to be concerned about thermo-EMF. However, when performing extremely sensitive measurements or when trying to extend the measurement accuracy of your parametric measurement equipment, the following guidelines can minimize or eliminate thermo-EMF effects:

1. Make sure that the output relay is closed and that the relay has stabilized to its final value. This means that the instrument should be warmed up and that the SMU output relay should remain closed.
2. Complete any measurements as quickly as possible (within less than 10 seconds). Doing this will keep the measurement drift to within a few microvolts.

One valuable resistance measurement technique that eliminates both the effects of EMF as well as any offset voltages in the measurement equipment is to force current and measure the resistance twice using the Kelvin measurement technique, reversing the polarity of the force current for the second measurement. If you work out the Kirchhoff's current law (KCL) and Kirchhoff's voltage law (KVL) equations for this circuit, you can see that by averaging these two measurements, the offset voltages and EMF voltages cancel out and you get the true value of the resistance.

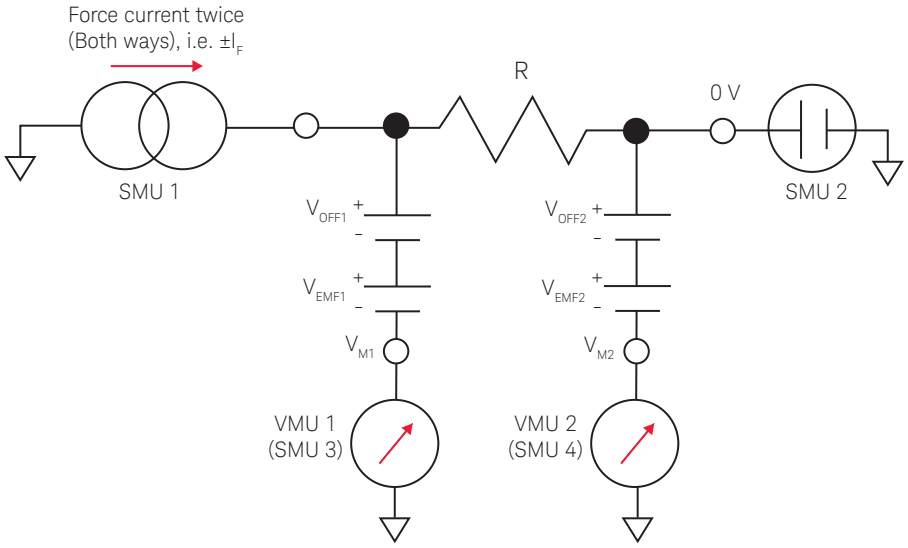


Figure 6.10. By measuring resistance twice and averaging the two resistor values, the effects of EMF and voltage offsets are eliminated.

$$R = \frac{(V_{M1} - V_{M2})}{I_F}$$

In summary, the key points of the above technique are:

1. Make sure that you keep the applied current small enough to eliminate any Joule self-heating effects.
2. Measure the resistance twice by reversing the polarity of the applied current and average the two results.
3. If available, use a voltmeter rather than an SMU to measure the voltage since voltmeters generally have superior voltage measurement accuracy.

Chapter 7 Diode and Transistor Measurement

"Choose a job you love, and you will never have to work a day in your life"
– Confucius

Introduction

It is not the intent of this handbook to teach a course on semiconductor device physics as there are already an abundance of excellent textbooks available on this subject. However, it is difficult to discuss making parametric diode and transistor measurements without first spending a little time understanding their operations. Therefore, we will give a brief review of pn junctions, diodes, and MOS and bipolar transistor operations with an emphasis on how we characterize them in parametric test as opposed to detailed theoretical derivations.

PN junctions and diodes

Review of PN diode operation

Intrinsic semiconductor materials (such as silicon) do not have an abundance of either electrons or electron holes. However, silicon can be doped with other materials such that it becomes either n-type (possessing excess electrons) or p-type (possessing excess electron holes). When considered individually, these materials are not particularly interesting. However, consider the case shown below when these two materials are brought into close contact.

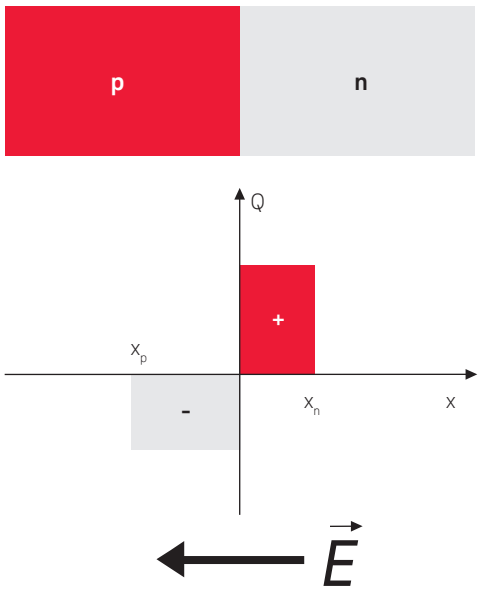


Figure 7.1. The cross section of a pn junction assuming an abrupt change from p-doped to n-doped material. The graph shows the fixed charge remaining after the mobile carrier diffusion has stabilized.

Assuming the extremely idealized case of an abrupt junction (i.e. one that instantaneously transitions from p to n material) as shown in Figure 7.1, we can see that something very interesting happens. The force of diffusion causes holes from the p-type material to flow into the n-type material (leaving behind a fixed negative charge), and similarly the force of diffusion causes electrons from the n-type material to flow into the p-type material (leaving behind a fixed positive charge). This diffusion process will continue until the electric field created by the fixed charge in what is normally called the space-charge region becomes strong enough to exactly balance the diffusion tendencies of the mobile carriers.

Review of PN diode operation *(continued)*

The one-dimensional (x-axis) equations defining current flow in a semiconductor are shown below.

$$J_n = q\mu_n E_x + qD_n \frac{dn}{dx} \quad \text{(Equation 7.1)}$$

$$J_p = q\mu_p E_x - qD_p \frac{dp}{dx} \quad \text{(Equation 7.2)}$$

Where
- J is the current density of electrons (n) and holes (p)
- q is the electron charge
- E_x is the electric field in the x-dimension
- μ is the mobility of electrons (n) and holes (p)
- D is the diffusion constant for electrons (n) and holes (p)
- n is the electron density
- p is the hole density

These equations basically state what was alluded to in the previous discussion of an abrupt pn junction. Namely, current flow in a semiconductor consists of two parts: a drift current proportional to the applied electric field and a diffusion current proportional to the spatial first derivative of the mobile carrier density. In addition to the above current flow equations, we also have the Einstein relationship which states the ratios of the mobility and diffusions constants as shown below.

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{kT}{q} \quad \text{(Equation 7.3)}$$

Where
- q is the magnitude of the electron charge (1.602×10^{-19} Coulomb)
- k is Boltzmann's constant (1.38×10^{-23} J/K)
- T is the absolute temperature [deg K]

The general form of Poisson's equation relates the second derivative of the electric potential to the total space charge density (ρ). Since we know that in a semiconductor this has to be related to the densities of mobile and fixed charge, we can derive the following equation.

$$\frac{d^2\varphi}{dx^2} = -\frac{q}{\varepsilon_{Si}}\left(p - n + N_d - N_a\right) \quad \text{(Equation 7.4)}$$

Where
- N_d is the donor density concentration
- N_a is the acceptor density concentration
- ε_{Si} is the permittivity of silicon

In the case of the abrupt junction shown in Figure 7.1, we make what is known as the depletion approximation, which assumes that the semiconductor is divided into distinct regions which are either completely neutral or completely depleted of mobile carriers. Therefore, in the depletion region we can rewrite the above equation as follows.

$$\frac{d^2\varphi}{dx^2} = -\frac{q}{\varepsilon_{Si}}\left(N_d - N_a\right) \quad \text{(Equation 7.5)}$$

Review of PN diode operation *(continued)*

Using the depletion approximation, we can integrate the above equation to get the electric field in both the p and n regions as shown below.

$$E_x(x) = -\frac{qN_a}{\varepsilon_{Si}}\left(x + x_p\right) \quad -x_p \le x \le 0 \qquad \text{(Equation 7.6)}$$

$$E_x(x) = -\frac{qN_d}{\varepsilon_{Si}}\left(x_n - x\right) \quad 0 \le x \le x_n \qquad \text{(Equation 7.7)}$$

Where

x_p is the width of the space charge in the p region (see Figure 7.1)
x_n is the width of the space charge in the n region (see Figure 7.1)

On a graph, these equations will appear as shown below.

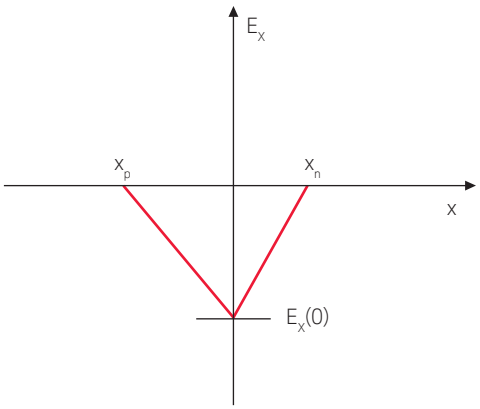


Figure 7.2. The electric field in an abrupt pn junction under the depletion approximation assumption.

We know that the electric field has to be continuous, so at x = 0 we can write the following.

$$E_x(0) = -\frac{qN_d x_n}{\varepsilon_{Si}} = -\frac{qN_a x_p}{\varepsilon_{Si}} \qquad \text{(Equation 7.8)}$$

This gives us the result shown below.

$$N_d x_n = N_a x_p \qquad \text{(Equation 7.9)}$$

Equation 7.9 shows an important characteristic of pn junctions: the width of the depletion region varies inversely with the magnitude of the dopant concentration. In other words, higher dopant concentrations result in narrower space charge regions.

Review of PN diode operation *(continued)*

When no voltage is applied to the pn junction, a barrier exists to current flow and the diode acts as an open circuit. The derivation of the current flow equations are involved and beyond the scope of this text. However, it should be somewhat intuitive that as we apply a positive voltage (i.e. electric field) to the p-region we are acting to reduce the built-in electric field of the pn junction. At some point the electric field is reduced enough to allow current to flow through the pn junction.

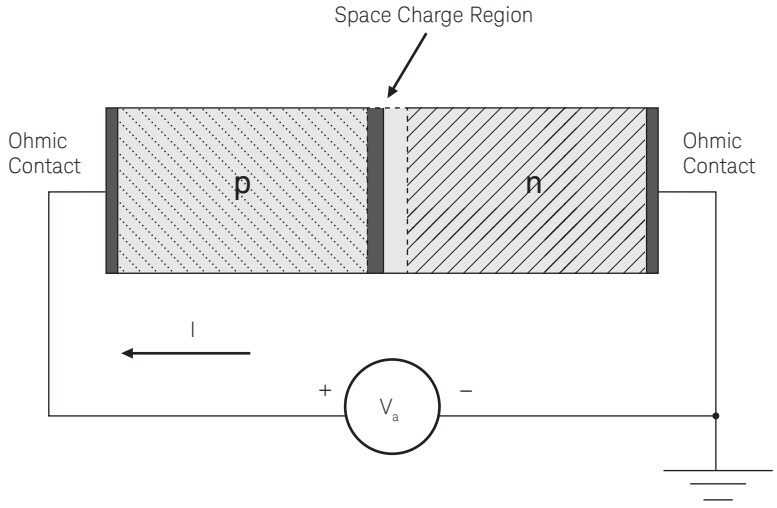


Figure 7.3. The behavior of a pn junction under positive applied bias.

Therefore, without any detailed derivations we will simply state that the current flow through a pn diode. exhibits exponential dependence upon applied voltage (V_a) and is given by the equation shown below.

$$I = I_o\left(\exp\left(\frac{qV_a}{kT}\right) - 1\right) \qquad \text{(Equation 7.10)}$$

This is sometimes called the ideal diode equation. It predicts a saturation current of $-I_o$ for negative values of V_a and an exponentially rising current for positive values of V_a. To emphasize that a diode only conducts current in one direction, it has the circuit symbol shown below.

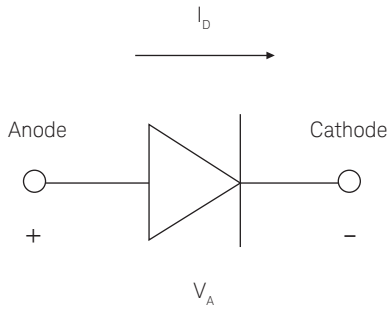


Figure 7.4. The circuit symbol for a diode.

Review of PN diode operation *(continued)*

The p-doped region is denoted as the anode, and the n-doped region is denoted as the cathode. Diodes can conduct current in both directions. However, larger voltages need to be applied to the cathode (relative to the anode) in order for current flow to occur in the reverse direction. In this condition, the diode is said to "breakdown", which is a logical term for this phenomenon, since it is an aberration from normal diode behavior. The physics of semiconductor junction breakdown will not be discussed in this handbook, but some practical measurement examples will be explored later.

The Ohmic contacts shown at the ends of the diode in Figure 7.3 simply mean that the semiconductor material is heavily doped enough such that the metal to semiconductor contact does not present any sort of barrier to the flow of current. If the semiconductor material is lightly doped, the metal to semiconductor contact can actually behave as another form of diode known as a Schottky barrier diode. Current flow in a Schottky barrier diode has a dependence on applied voltage, shown below.

$$I = I'_o\left(\exp\left(\frac{qV_a}{nkT}\right) - 1\right) \qquad \text{(Equation 7.11)}$$

In equation 7.11, n is a constant usually ranging between 1.02 and 1.15. The prime symbol is present on I_o to emphasize that this constant is different in value from that for the case of a pn junction. Schottky diodes typically have an effective "turn-on" voltage that is several hundred millivolts less than that of a pn junction diode, which makes them essential in the design of bipolar logic circuits since they can keep the base to collector junction from forward biasing and therefore keep the transistor out of saturation.

One important point to note about pn junctions is that they behave as voltage-dependent parallel plate capacitors. As external voltage is applied, the charges in and around the space-charge region are modified. Therefore, junction capacitance is an important parameter that must be characterized for all semiconductor devices, as it impacts the speed at which the devices will switch when used in an integrated circuit. However, since capacitance measurement is much more challenging to perform correctly than simple current and voltage (IV) measurements, we will defer a detailed discussion of semiconductor capacitance measurement until chapter 8.

Basic diode characterization

Diodes are relatively simple devices to characterize. From equation 7.10, we can see that a plot of the log of the diode current (I_d) should be linear with respect to applied voltage. A plot of diode current and the log of the diode current for a “typical” diode are shown below.

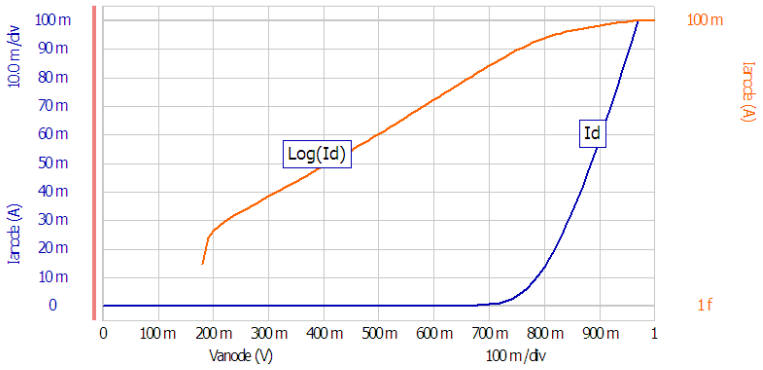


Figure 7.5. A pn diode sweep in the forward direction plotting both I_d (blue) and Log(I_d) (orange).

Of course, another important parameter is the reverse breakdown characteristics of the pn diode. A plot of this is shown below.

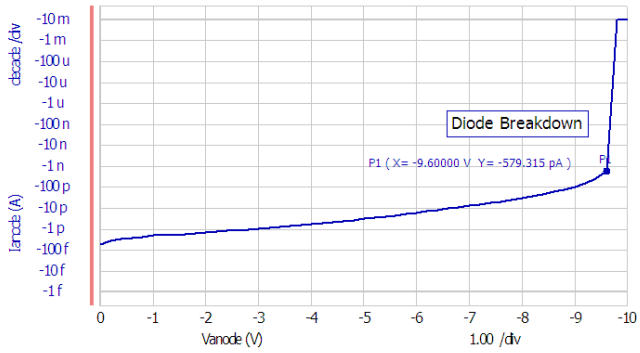


Figure 7.6. The reverse breakdown characteristics of a diode.

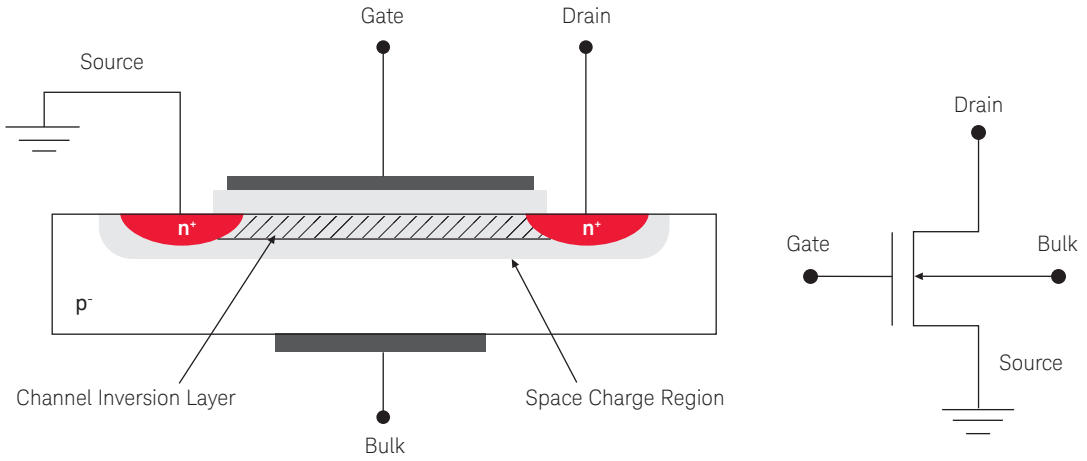
It is possible to control the reverse breakdown characteristics of certain classes of diodes very precisely using various processing techniques. Diodes with these sorts of precisely controlled breakdown characteristics are known as zener diodes. The ability to control the breakdown voltage has many beneficial uses in circuit design, since it permits the zener diodes to be used as voltage clamps within the circuit.

MOS transistor measurement

Silicon-based MOS transistors are the dominant technology for digital electronics for many reasons, foremost of which is their ability to scale relatively easily (at least until recently) with improvements in lithography. The operation of a MOS transistor is more intuitive and easier to explain than that of a bipolar transistor. Essentially, in a MOS transistor we have two highly doped contact regions (the source and the drain) separated by a channel region of opposite doping type. By applying a voltage to the gate, which is located above the channel region, we can induce charge in the channel region such that we create a conductive path between the source and the drain.

Review of MOSFET operation

A simplified cross-section of an n-channel type MOS (NMOS) transistor with positive bias on the gate and the source grounded is shown below. Note: In this example, we assume that the voltage applied to the bulk is zero volts.



$V_d \approx 0, V_g > V_t$

Figure 7.7. An n-channel MOS transistor with gate bias greater than the threshold voltage but with negligible drain to source voltage.

In this example, the voltage applied to the drain (V_d) is small relative to the gate voltage (V_g). As long as V_g exceeds the threshold voltage necessary to invert the channel (V_t) a conductive channel exists between the source and drain. In this situation the drain current (I_d) is given by the charge in the channel (Q_{ch}) divided by the time it takes the carriers to transit across the channel (T_{tr}).

$$I_d = \frac{Q_{ch}}{T_{tr}} \qquad \text{(Equation 7.12)}$$

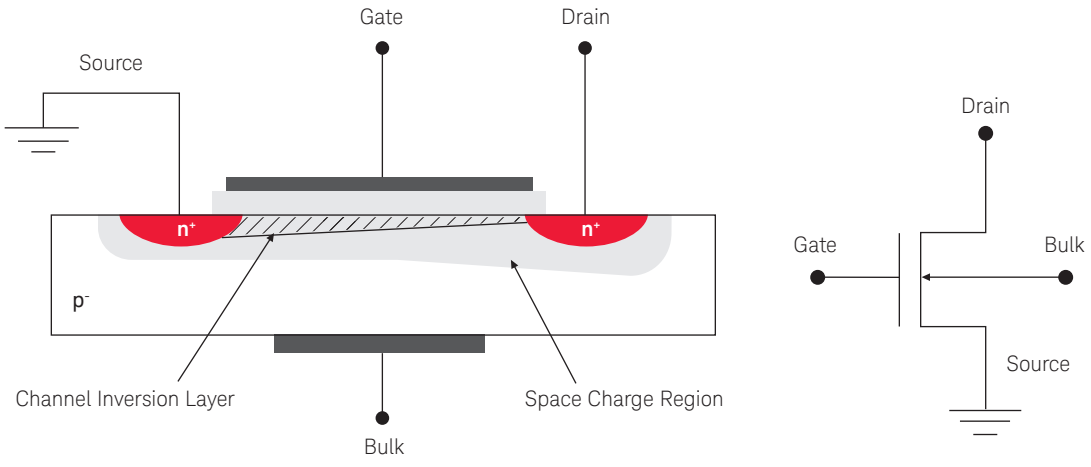
Review of MOSFET operation *(continued)*

The channel charge is dependent on the channel capacitance, area and applied gate voltage. The transit time is dependent on the drift velocity of the carriers and the channel length. In this regime, the value of the drain current is approximated by the equation shown below.

$$I_d = \mu_n \frac{W}{L} C_{ox} \left(V_g - V_t \right) \bullet V_d \qquad \text{(Equation 7.13)}$$

Where
μ_n is the electron mobility
W is the channel width
L is the channel length
C_{ox} is the capacitance per unit area in accumulation

If we increase the drain voltage (V_d) such that it is no longer negligible relative to the gate voltage, then the space charge region and the channel are no longer spatially uniform as shown below.



$0 < V_d < (V_g - V_t), V_g > V_t$

Figure 7.8. An n-channel MOS transistor with gate bias greater than the threshold voltage but with non-negligible drain to source voltage.

Review of MOSFET operation *(continued)*

The simple assumptions used to derive the previous equation for I_d are no longer valid, since the drain voltage acts to reduce the inversion layer near the drain. In essence, this means that the inversion layer becomes more trapezoidal than rectangular. We can approximate this by assuming the average voltage in the channel to be $(V_g - V_d/2)$. This leads to the revised equation for I_d shown below.

$$I_d = \mu_n \frac{W}{L} C_{ox} \left(V_g - V_t - \frac{V_d}{2} \right) \bullet V_d \qquad \text{(Equation 7.14)}$$

This equation defines the behavior of I_d in the non-saturation region of operation.

Plotting out the previous equation for various values of V_g results in a series of downward-facing parabolas with maxima at the point where $V_d = (V_g - V_t)$ as shown below.

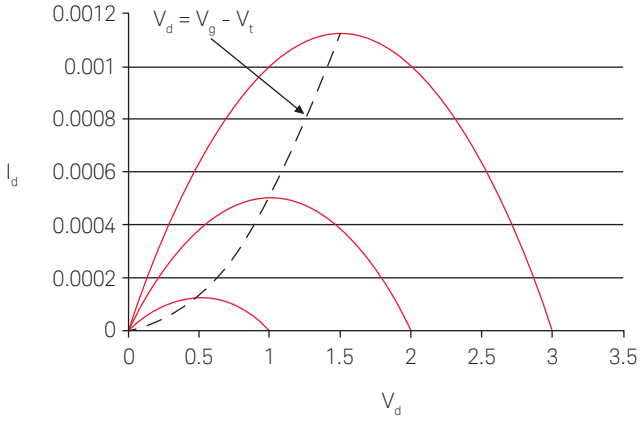
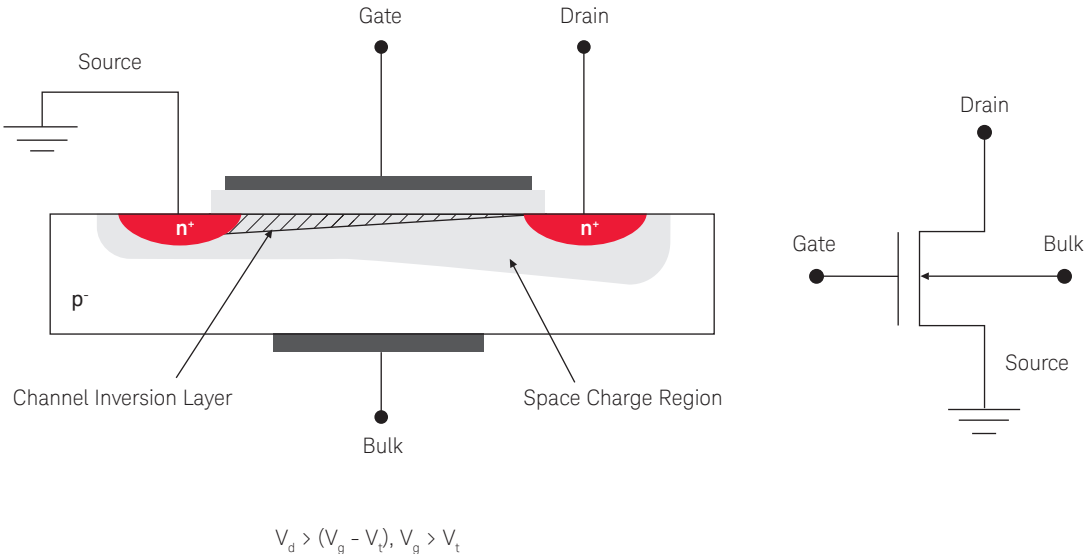


Figure 7.9. Plotting out the value of I_d using the previous equation for different values of V_g yields a series of downward facing parabolas.

Review of MOSFET operation *(continued)*

It should be obvious that this does not make any physical sense for values of V_d greater than V_g - V_t, since this would imply a negative transconductance (transconductance being the first partial derivative of V_d with respect to V_g). To understand what is going on, we need to consider what happens when V_d becomes greater than (V_g - V_t).



$V_d > (V_g - V_t), V_g > V_t$

Figure 7.10. An n-channel MOS transistor with gate bias greater than the threshold voltage and with V_d greater than (V_g – V_t).

When V_d becomes greater than (V_g - V_t) the inversion layer becomes pinched off at the drain. However, as the electrons approach the drain, there is no barrier to stop them and they are rapidly accelerated by the high electric field into the drain region. In this situation, the current is determined by the rate at which electrons reach the edge of the depletion region, which is in-turn determined by the maximum electron drift velocity. Since the maximum drift velocity is (at least to first order) insensitive to V_d, once V_d reaches the value of (V_g - V_t), the drain current becomes saturated (constant). The equation for the drain current when in saturation is given below.

$$I_d = k\frac{W}{L}\left(V_g - V_t\right)^2 \qquad \text{(Equation 7.15)}$$

Review of MOSFET operation *(continued)*

A plot of the I_d versus V_d characteristics for an n-channel MOSFET is shown below.

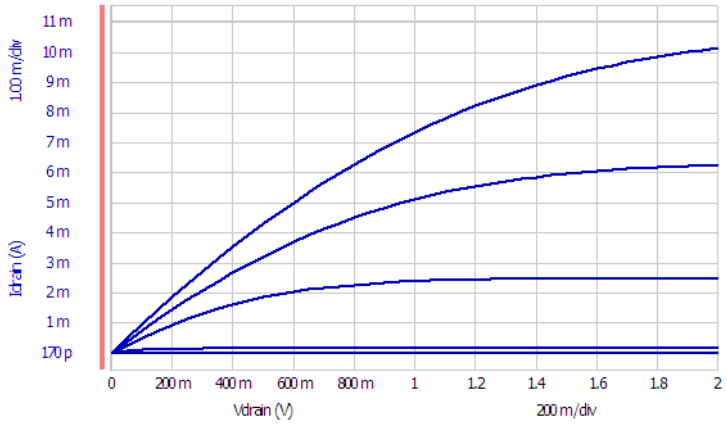


Figure 7.11. A "typical" I_d versus V_d characteristic for an n-channel MOSFET.

The value of the constant k is proportional to the product of μ_n and C_{ox}. However, it is not exactly the same. In actuality, the drain current exhibits a slight positive dependence on the drain voltage, since the effective channel length (L) decreases slightly for larger values of V_d. Most of the time, the k, W and L factors are simply absorbed as another constant, shown in the equation below.

$$I_d = \beta \bullet \left(V_g - V_t\right)^2 \qquad \text{(Equation 7.16)}$$

Of course, the previous discussion of MOS transistor behavior was extremely simplified. However, the modification to the previous equations for other conditions is relatively straightforward. In the case, where the transistor source terminal is not at ground, it is a correct assumption that (at least to first order) an increase in V_s will have the effect of reducing the influence of the gate voltage. Therefore, it is reasonable to modify the equation for I_d in saturation for the case where $V_s \neq 0$ as follows.

$$I_d = \beta \bullet \left(V_g - V_s - V_t\right)^2 \qquad \text{(Equation 7.17)}$$

Basic MOSFET characterization

From the equations we developed for MOSFET behavior, we can see that a plot of the square root of I_d versus V_g should be linear.

$$\sqrt{I_d} = \sqrt{\beta} \bullet \left(V_g - V_s - V_t\right) \qquad \text{(Equation 7.18)}$$

In addition, the intercept of this line with the x-axis (V_g) should equal the threshold voltage (when the source is grounded). Therefore, a simple technique for determining the threshold voltage (V_t) of a MOS transistor is to plot the square root of I_d and extrapolate this to its intersection with the x-axis.

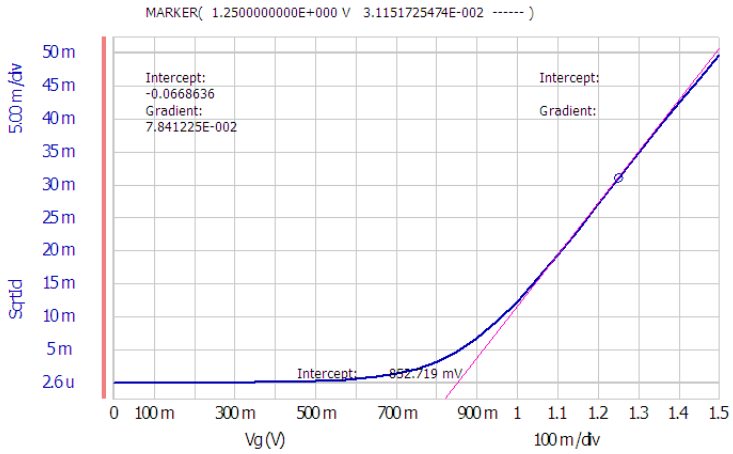


Figure 7.12. Determining the Vth of a MOSFET by plotting the square root of I_d and extrapolating the slope back to the x-axis.

Note: In reality there is some subthreshold leakage current flowing in the transistor for V_g values less than V_t, so a tangent line needs to be extrapolated to the x-axis from a point on the square root of I_d plot where $V_g \geq V_t$.

The effect of a non-zero bulk voltage can be understood by examining its impact upon the charge within the channel. Consider the case shown below where a negative voltage is applied to the bulk.

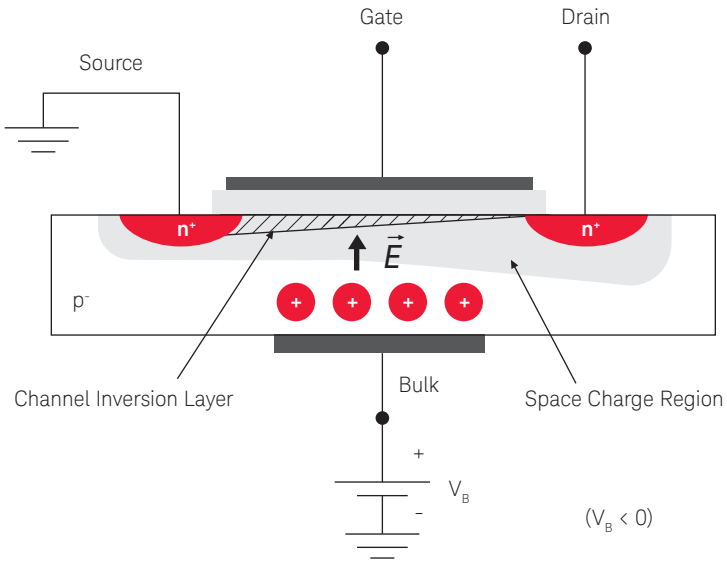


Figure 7.13. The effect of applying a negative bias to the n-channel MOS transistor bulk terminal.

Basic MOSFET characterization *(continued)*

Applying a negative charge on the bulk will induce positive charge to accumulate near the bulk contact. The effect of this positive charge is to act in opposition to the voltage applied to the gate, effectively requiring more gate voltage to achieve channel inversion than for the case where the bulk is at zero volts. Therefore, we would expect to see V_t increase with decreasing values of V_b (which is commonly known as the "body effect").

Similarly, applying a positive charge on the bulk will cause V_t to decrease. The equations that describe this behavior are beyond the scope of this basic handbook, but a plot of I_d versus V_g for several values of V_b shows that physical reality matches our intuition.

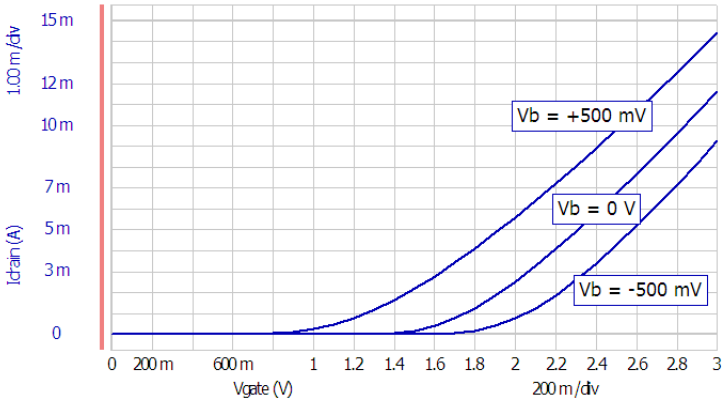


Figure 7.14. The effect of the bulk voltage on threshold voltage (Vt).

As can be seen from this plot, negative values of V_b shift the curve to right which implies an increase in the threshold voltage (V_t). Similarly, positive values of V_b shift the curve to the left which implies a decrease in the threshold voltage.

On-resistance, R_{ds}(on), is an important parameter for power MOSFETs. R_{ds}(on) is defined as the forced drain voltage (V_{ds}) divided by the drain current (I_d) in the region, where the MOSFET is in the on state. The B1505A and B1506A can easily calculate the R_{ds}(on) at hundreds of amps of current. In the following example, the R_{ds}(on) of a MOSFET is measured up to 300 A. Note that while we are measuring using drain currents in the hundreds of amps, the value of the on resistance is quite small (about 900 micro-Ohms).

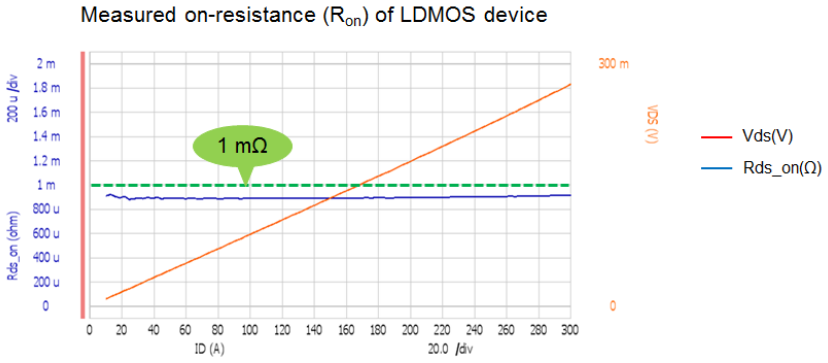


Figure 7.15. The B1505A and B1506A can easily measure sub-milliohm on resistances at hundreds of amps of current.

Basic MOSFET characterization *(continued)*

The drain-source breakdown voltage (BV_{dss}) and drain cutoff current (I_{dss}) are also very important parameters for high-power MOSFETs. To determine BV_{dss}, the gate and source are both shorted to ground and the drain voltage is swept from zero volts to a very large value. Although the voltage required to reach breakdown is typically over 1000 V for power MOSFETs, the HVSMU of the B1505A and B1506A can easily make this measurement. EasyEXPERT's analysis functions make it easy to automatically display the BV_{dss} once a current level (1 µA in this case) to define the onset of breakdown has been selected. Similarly, once a voltage level (1500 V in this case) has been selected to define I_{dss}, it too can be automatically displayed. The following plot illustrates this measurement.

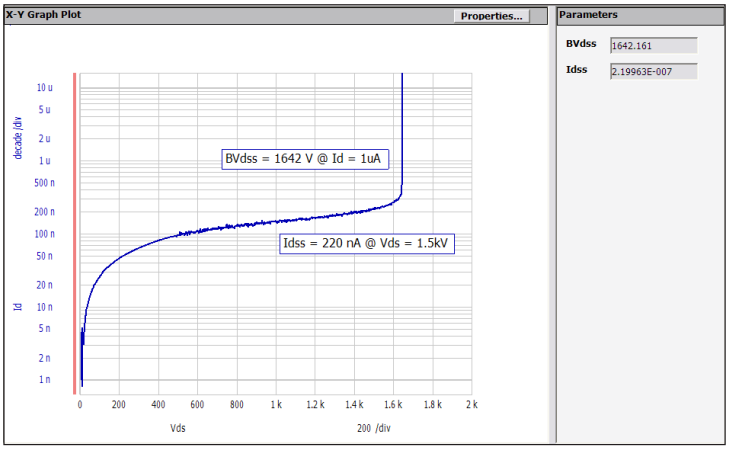


Figure 7.16. The B1505A/B1506A's HVSMU module and EasyEXPERT software make it easy to measure the BV_{dss} and I_{dss} of a power MOSFET.

Bipolar transistor measurement

With the predominance of CMOS technologies in integrated circuits, bipolar transistors have largely been relegated to specialized (albeit not necessarily unimportant) niche applications. The most common uses for bipolar transistors today are in microwave circuitry, high-power applications and analog circuitry.

Review of bipolar transistor operation

The operation of a bipolar transistor is actually quite a bit more complicated to explain (and less intuitive) than that of a MOS transistor. The configuration of a bipolar transistor consists of three dopant regions of alternating type (either NPN or PNP). The following discussion will be centered on the operation of an NPN transistor which has the circuit symbol shown below.

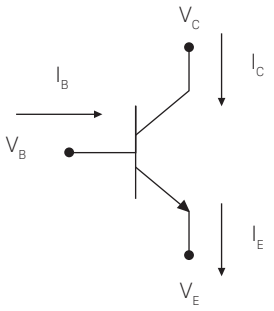


Figure 7.17. The circuit symbol for an NPN bipolar transistor.

Note that the base to emitter junction contains an arrow similar to that for the representation of a pn diode, and the base and emitter do form a pn diode. Similarly, the base to collector junction also forms a pn diode. However, while it is true that any bipolar transistor will function in both the forward and reverse directions, due to intentional asymmetrical doping of the junctions, optimal performance is obtained when the base to emitter junction is forward biased.

We will now consider the case where we have the emitter of the NPN transistor at ground and we apply a positive voltage to the base. Let's assume that the transistor collector is at a voltage much greater than that of the base (V_C » V_B). If the base to emitter voltage (V_{BE}) is sufficiently large enough to forward bias the base to emitter junction, then current will flow in the transistor as shown below.

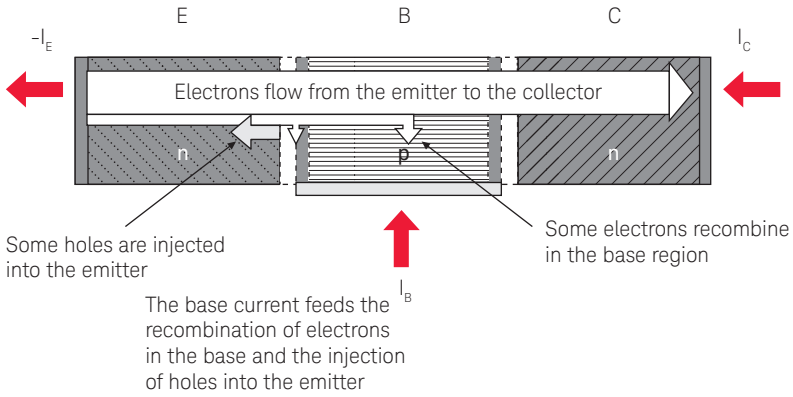


Figure 7.18. The basic operation of a bipolar transistor (NPN type).

Review of bipolar transistor operation *(continued)*

Of course, electron flow from the emitter to the collector is equivalent to a positive current flow from the collector to the emitter.

Similar to the case of a pn diode, there is an exponential relationship between current flow and the base to collector (V_{BC}) and base to emitter (V_{BE}) voltages. Assuming that V_{BC} is zero or negative, the collector current is approximately described by the equation shown below.

$$I_c \approx I_o\left(e^{\frac{qV_{BE}}{kT}}\right) \qquad \text{(Equation 7.19)}$$

As can be seen from this equation, the base to emitter voltage does not need to change much in order to have large changes in the collector current. Therefore, once the NPN transistor has a sufficiently large V_{BE} to become actively biased, the magnitude of the collector current is limited by the amount of current entering into the base of the transistor. As the previous diagram showed, this base current is necessary to feed both the recombination of electrons in the base region as well as the injection of holes into the emitter. There is a very simple linear relationship between the base and collector currents given by the equation shown below.

$$I_c = \beta \bullet I_B \qquad \text{(Equation 7.20)}$$

Basic bipolar transistor characterization

Probably the single most important bipolar transistor parameter is its current gain or beta. This parameter is also often written in terms of h-parameters as h_{FE}, which stands for forward gain through the emitter. A Gummel plot is usually used to determine the value of beta for various values of collector current. One method to perform a Gummel plot on an NPN transistor is to set the collector and base to zero volts and to sweep the emitter negatively. The values of base current and collector current are then measured and plotted on a log scale (since by equations 7.19 and 7.20, the natural logarithm of the collector and base currents should be straight lines).

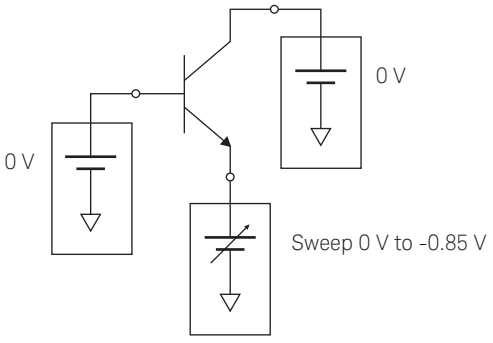


Figure 7.19. Typical setup for a bipolar transistor Gummel plot.

Basic bipolar transistor characterization *(continued)*

The results of an actual Gummel plot are shown below.

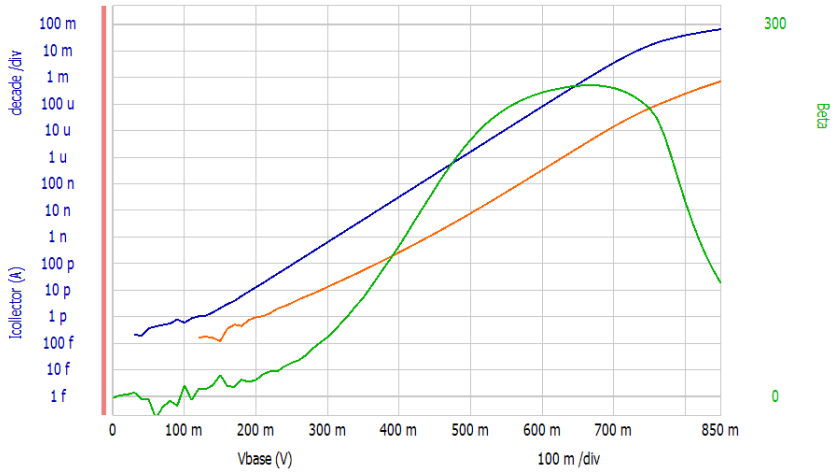


Figure 7.20. A Gummel plot made on an NPN bipolar transistor using the B1500A. Ic, Ib and the calculated beta are all shown on the same graph.

As this plot shows, the value of the transistor beta is actually dependent upon the large-signal value of the collector current, although it is approximately constant over a fairly wide range of collector current values.

Another important bipolar transistor parameter is the emitter resistance (R_e). The flyback method is the most common means used to determine this parameter. In this measurement, the SMU connected to the collector terminal is placed in current force mode with a very small force value (almost zero amps), making it essentially a high-impedance voltmeter. The SMU connected to the base terminal is also set to current force mode and the collector-emitter voltage (V_{ce}) is measured and plotted versus the base current (I_b).

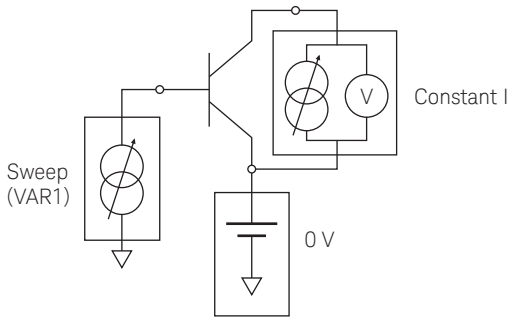


Figure 7.21. The configuration for measuring R_e on a bipolar transistor using the flyback method.

Basic bipolar transistor characterization *(continued)*

The slope of the line in the flyback region can be determined by creating a regression line at predefined points using the built-in EasyEXPERT analysis functions. The slope of this line is in units of amps/volt, which is the inverse of resistance (which has units of volts/amp). Therefore, the inverse of the regression line slope gives us the value of emitter resistance (R_e). A plot of I_b versus V_{ce} showing the emitter resistance is shown below.

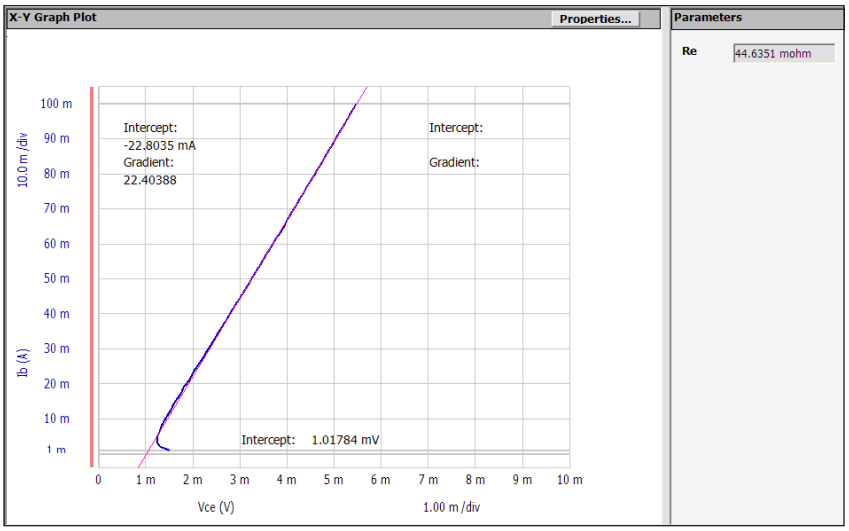


Figure 7.22. Extracting the emitter resistance (Re) using the flyback method.

As in the case of power MOSFETs, the breakdown voltages of power bipolar devices are also very important parameters. The junction breakdown voltages of power BJTs are defined as follows.

BV_{cbo} – The collector-base breakdown voltage with the emitter open

BV_{ceo} – The collector-emitter breakdown voltage with the base open

BV_{ebo} – The emitter-base breakdown voltage with the collector open

Basic bipolar transistor characterization *(continued)*

The B1505A/B1506A's HVSMU can measure breakdown voltages for most power BJTs, and it can also measure the leakage current with picoamp resolution. EasyEXPERT's graphical capabilities make it easy to superimpose these different breakdown measurements onto a single plot and to add comments to the graph for use in reports and presentation. An example of this is shown below.

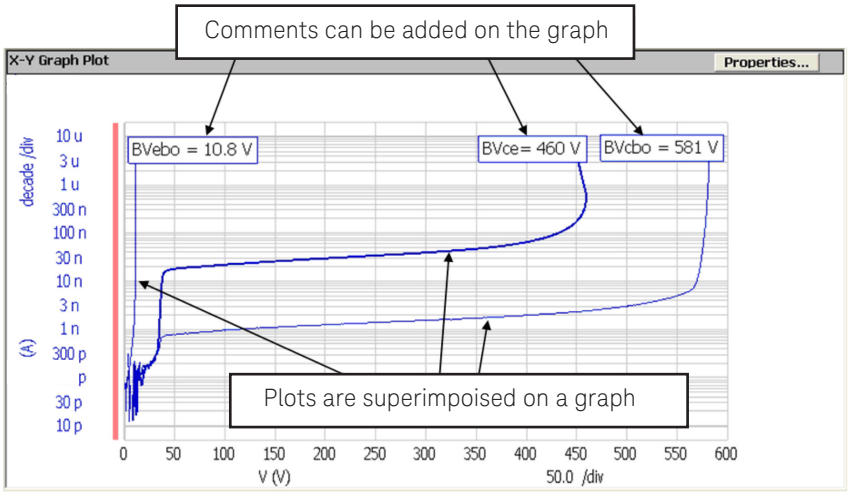


Figure 7.23. High power bipolar transistor breakdown voltages displayed on a single graph.

The insulated gate bipolar transistor

The insulated gate bipolar transistor (IGBT) is a power device that combines a MOSFET with a bipolar device. A generic version of the IGBT along with equivalent circuit diagrams is shown below.

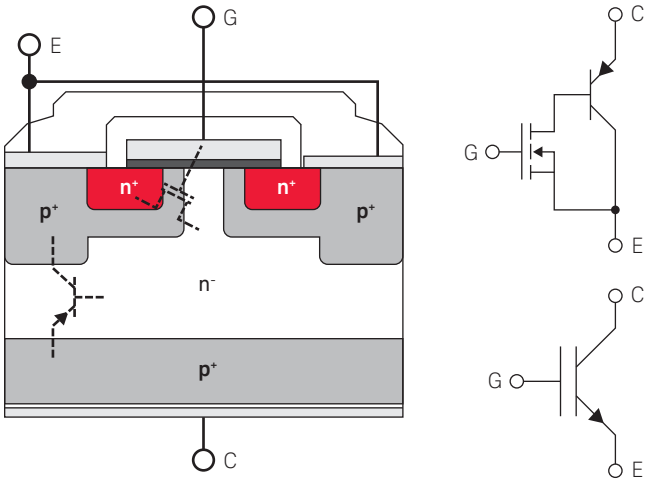


Figure 7.24. The structure and circuit symbols for the insulated gate bipolar transistor (IGBT).

The IGBT combines the high input impedance of a MOSFET with the low saturation voltage of a bipolar transistor. This makes the IGBT excellent for applications requiring high current and high voltage, although it cannot switch as fast as a power MOSFET due to the minority carriers associated with its bipolar component.

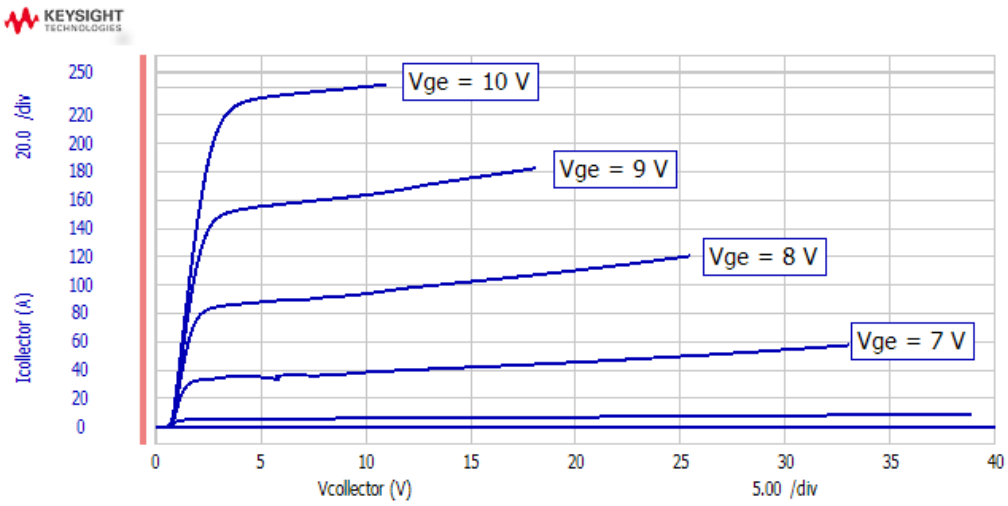


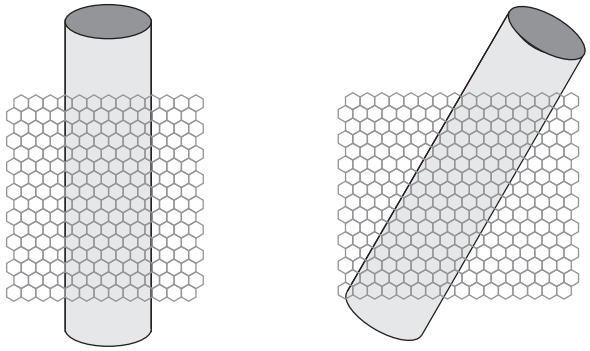
Figure 7.25. An Ic-Vge family of curves characterized on an IGBT device up to approximately 250 A of collector current.

Nanotech devices

A significant amount of research has gone into the development of transistors and other devices based on nanotechnology. While the testing of these devices does not present any fundamental new challenges, it is worthwhile to give a brief overview of their properties and how to characterize them.

Carbon nanotubes

Carbon nanotubes (CNTs) were first discovered in 1991 by Sumio Iijima. The CNT is essentially one or more sheets of graphitic carbon rolled up into tubes. The electrical properties of a CNT are strongly determined by the orientation of the carbon atoms in the tube. This orientation is referred to as “chirality”. As the following figure shows, a CNT with straight chirality generally acts as a conductor, while a CNT with non-straight chirality generally acts as a semiconductor.



Straight Chirality (Conductor)

Non-straight Chirality (Semiconductor)

Figure 7.26. The chirality of a carbon nanotube determines its electrical properties.

Carbon nanotube transistors

Graphite is a semi-metal with no bandgap. However, a few electrons can reach the conduction states from the Fermi energy level. By varying the chirality, energy states around the Fermi level can be added or removed, and this in-turn modifies the “effective bandgap”. In conductive mode, CNTs are believed to conduct current through ballistic transport, giving them very low resistance and excellent current carrying capability (1000 times better than copper wire). Most CNT FETs use either a back gate or a side gate structure as shown in the following figure.

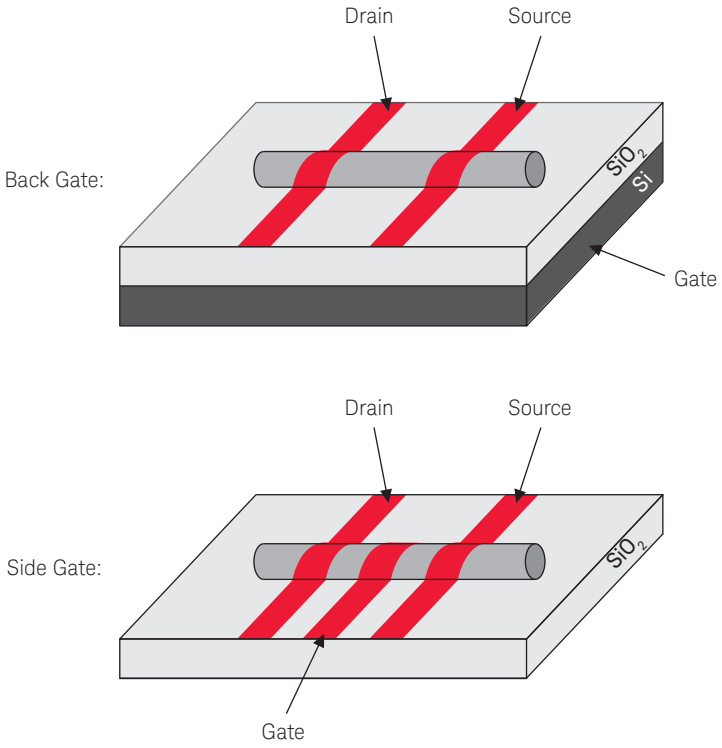


Figure 7.27. Illustration showing the two most common methods of creating a gate connection on a carbon nanotube transistor.

Often, the most challenging part of testing nanotech devices is making the physical connections to the DUT. For example, the outputs of SMUs are triaxial (to support low-current measurement) but some probers do not support triaxial probes. In this case, in order to maintain the instrument's low-current measurement performance, you need to tie the guard of the triaxial cable to the outer shield of the coaxial cable (which must be isolated from ground). The following figure shows how to properly connect a B1500A to a wafer prober with coaxial probes when probing a CNT FET.

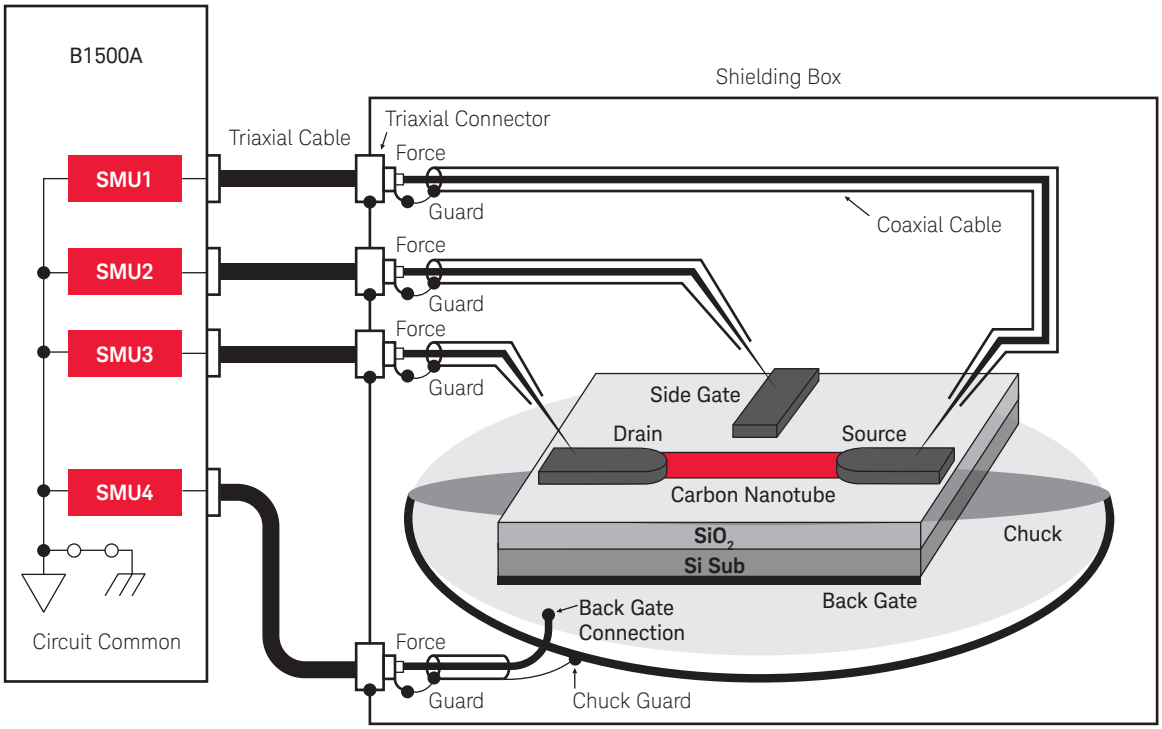


Figure 7.28. The proper connections to probe a CNT FET using the B1500A and a wafer prober with BNC probes.

EasyEXPERT software has built-in libraries to simplify the testing of CNT FETs and other types of nanotech devices. The following figure shows an application test that can characterize a CNT FET with both back and side gates.

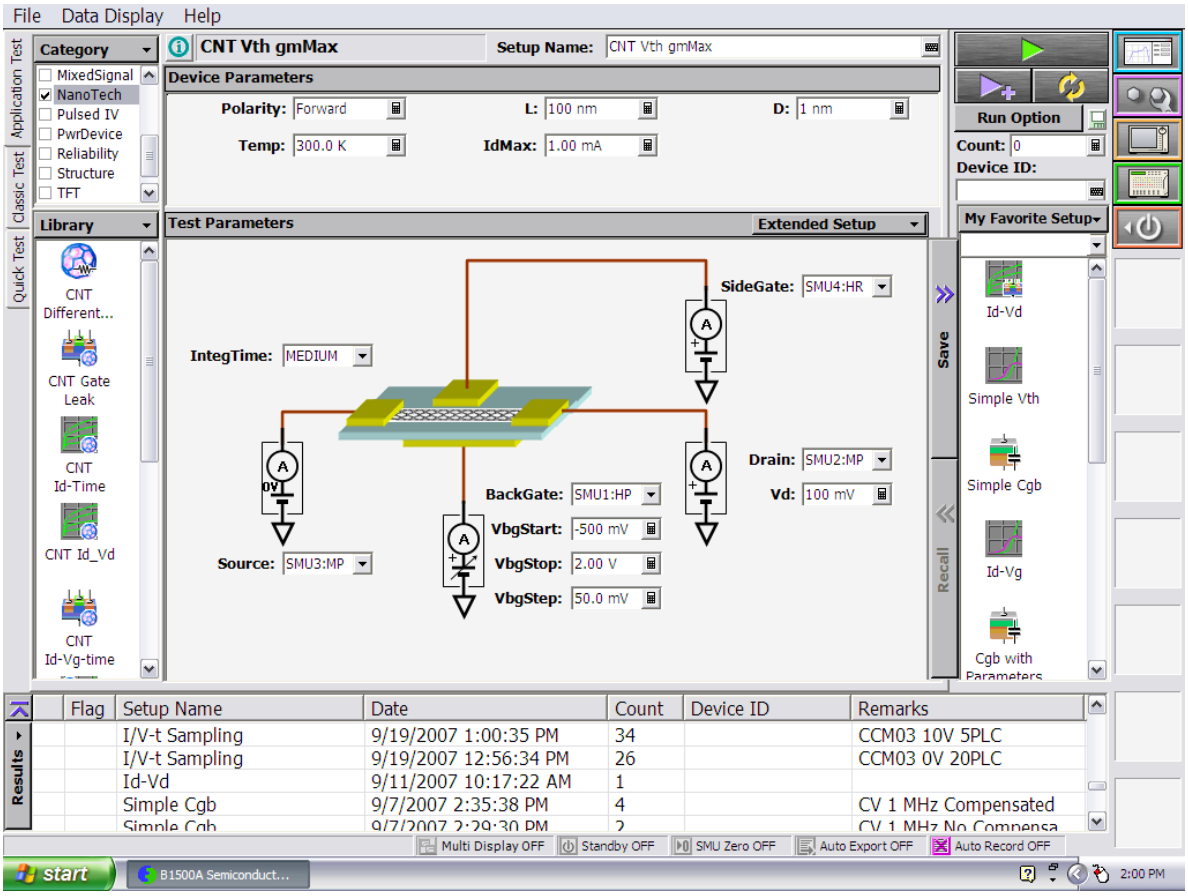


Figure 7.29. Example of a nanotech transistor application test furnished in the EasyEXPERT nanotech device library.

Chapter 8 Capacitance Measurement Fundamentals

"Furious activity is no substitute for understanding." – H. H. Williams

Introduction

Capacitance measurement is one of the subjects in parametric testing where many easily preventable measurement mistakes are often made. The reason for this is not lack of intelligence on the part of the user, but rather a lack of fundamental training on capacitance measurement theory and how to make good capacitance measurements. Unfortunately, this type of information is typically not taught in universities and can usually only be learned either through (rather painful) experience or by reading about it in a publication such as this.

Why do engineers make so many mistakes when measuring capacitance (especially on-wafer)?

The most common reasons are:

1. Capacitance measurement requires compensation to remove parasitic inductance and capacitance from the measurement cables and fixtures. This is often done improperly or not done at all.
2. An induced current flows through the outer shield of the BNC connectors (if used) on a capacitance meter, and this current is necessary to balance the measurement current of the capacitance meter. If the outer shield is grounded, then the induced current flow is shorted to ground and the bridge may not be able to balance. Many users are unaware of this issue.
3. Measuring capacitance on a semiconductor wafer on a wafer chuck is very different from measuring capacitance on a discrete device. The effects of the wafer prober chuck on the measurement cannot be ignored.
4. For higher measurement frequencies (>5 MHz), test structure (layout) design has a major impact on the success or failure of the measurements.

As we progress through this (rather lengthy) chapter, we will cover all of the above issues in detail.

MOSFET capacitance measurement

Review of MOSFET capacitance behavior

Before delving into capacitance measurement theory, it is useful to first review MOSFET device operation to remind ourselves as to why we make these measurements in the first place. While the capacitance measurement techniques discussed in this chapter are general and can be used on a variety of different device types, the dominance of MOSFETs in modern electronics has elevated their importance enough to justify devoting a section to review their operating characteristics.

MOSFETs are voltage-dependent capacitors. The MOSFET gate-to-substrate capacitance depends upon the applied DC voltage (which we measure using an AC voltage of much smaller magnitude that rests on top of the DC voltage). The following set of diagrams shows the behavior of an NMOS transistor as the voltage applied to the gate is varied from negative to positive.

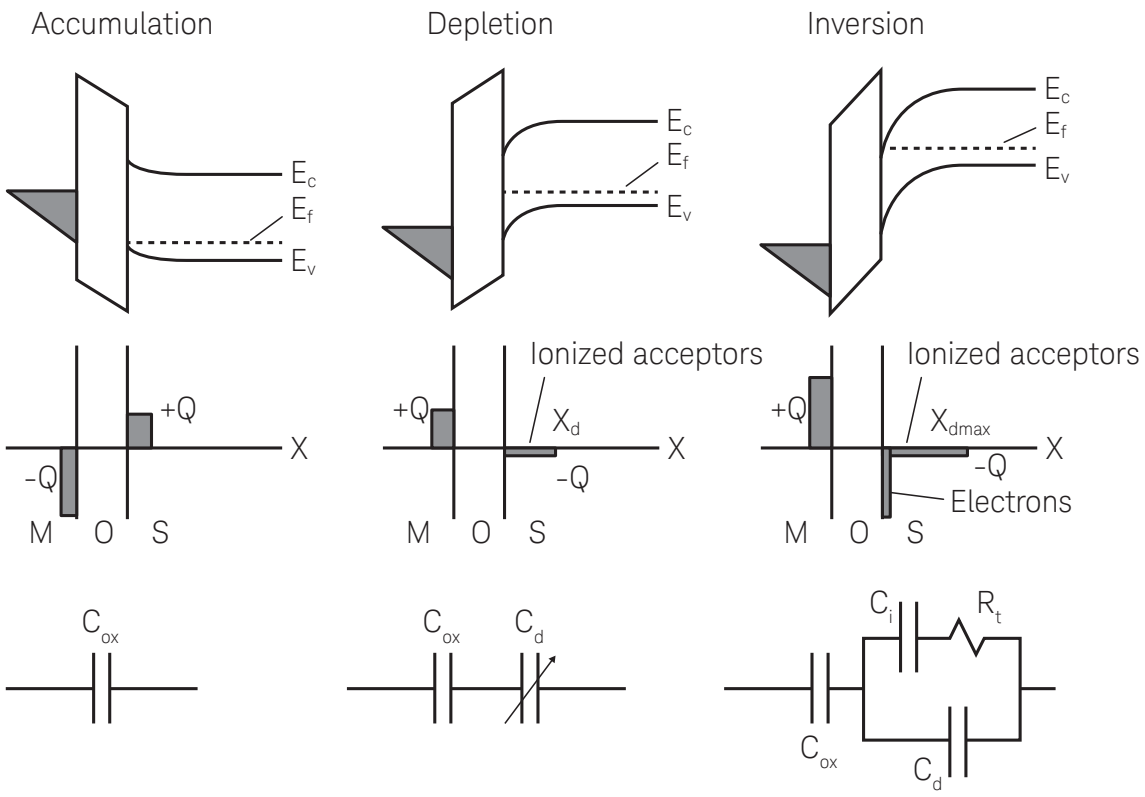


Figure 8.1. Capacitive behavior of an NMOS transistor to changes in the voltage applied to the gate.

If the silicon is held at ground and a negative voltage is applied to the gate, the MOS capacitor will begin to store positive charge at the silicon surface. The surface has a greater density of holes than Na (the acceptor density), and this condition is known as surface accumulation. In this condition, the mobile charge on both sides of the oxide can respond rapidly to changes in applied voltage, and the device looks just like a parallel plate capacitor of thickness t_{ox}. Since it is a pure gate oxide capacitance, we denote its value as C_{ox}.

Review of MOSFET capacitance behavior *(continued)*

If a positive gate voltage is applied to the gate relative to the silicon, the built-in positive voltage between the gate and the silicon is increased. The silicon surface becomes further depleted of carriers as more acceptors become exposed at the surface, resulting in the condition known as surface depletion. In this condition, electrostatic analysis shows that the total MOS capacitance consists of the series combination of C_{ox} and the capacitance across the surface depletion region, C_d. Note that C_d depends upon the applied voltage.

If the positive gate voltage is further sufficiently increased, then the energy bands bend away considerably from their levels in the bulk of the silicon. The depletion region reaches a maximum width, x_{dmax}, and all of the electron acceptors within this region are fully ionized. In the surface region, generation of carriers exceeds recombination, and the generated electrons are swept by the electric field into the oxide-silicon interface where they remain due to the energy barrier between the conduction bands of the silicon and the oxide. Thus, the total charge in the silicon consists of the sum of these two charges. Electrostatic analysis again shows that the total MOS capacitance can be modeled as the oxide capacitance in series with the parallel combination of the depletion capacitance and the series combination of surface charge capacitance, C_i, and the depletion resistance, R_t.

Sample MOSFET parameter calculation

The measurement of the gate to substrate capacitance of a MOSFET device is highly important because it is the only measurement able to determine many important device parameters such as substrate impurity concentration (N_{sub}) and flat band voltage (V_{fb}). A sample parameter calculation is illustrated to show how to extract these factors from a capacitance versus voltage (CV) curve. All of the following calculations are based upon the (CV) curve shown below:

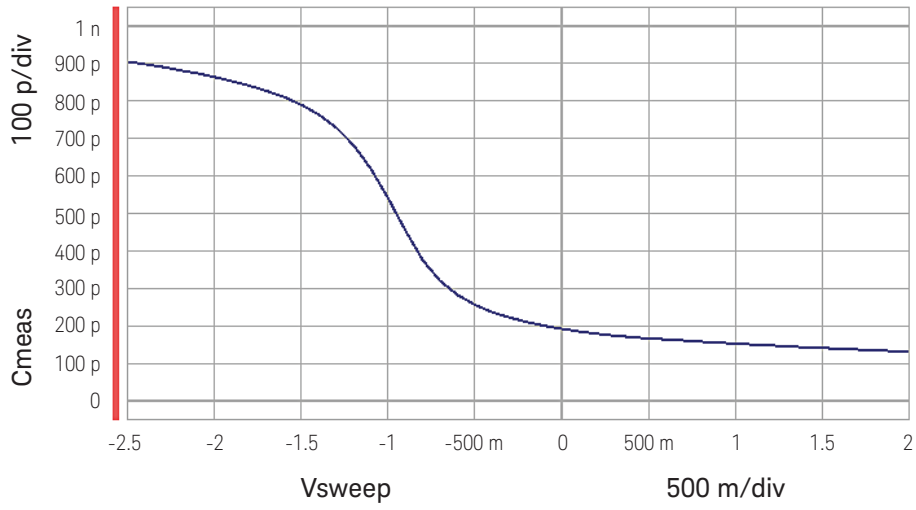


Figure 8.2. Capacitance versus voltage (CV) plot.

Sample MOSFET parameter calculation *(continued)*

The gate oxide thickness of a MOSFET capacitor (t_{ox}) can be calculated from the standard equation for a parallel plate capacitor:

$$t_{ox} = \frac{A \cdot 10^8 \cdot \varepsilon_o \cdot \varepsilon_d}{C_{ox}} \quad \text{[Angstroms]} \qquad \text{(Equation 8.1)}$$

Where:

A is the capacitor gate area [cm^2]
ε_o is the free space permittivity (8.854×10^{-14} F/cm)
ε_d is the dielectric constant of SiO_2 (3.9)
C_{ox} is the measured capacitance in heavy accumulation (V_g bias = V_{dd}) [F]

For $C_{ox} = 9.040 \times 10^{-10}$ F and A = 0.001 cm^2, we have that:

t_{ox} = 37 Angstroms

The next two parameters that we must calculate are the substrate impurity concentration and the Fermi potential. These are given by the following two equations:

$$N_{sub} = \frac{4 \cdot |\varphi_f|}{q \cdot \varepsilon_o \cdot \varepsilon_{si}} \left(\frac{Cs_{min}}{A} \right)^2 \quad \text{[1/cm}^3\text{]} \qquad \text{(Equation 8.2)}$$

$$\varphi_f = \pm \frac{kT}{q} \cdot \ln\left(\frac{N_{sub}}{n_i} \right) \quad \text{[Volts]} \qquad \text{(Equation 8.3)}$$

Where:

N_{sub} is the impurity concentration of the substrate
n_i is the intrinsic carrier concentration [$1/cm^3$]
Φ_f is the Fermi potential [Volts]
Cs_{min} is the minimum depletion layer capacitance [Farads]
ε_{Si} is the dielectric constant of Si (11.7)
q is the magnitude of the electron charge (1.602×10^{-19} Coulomb)
k is Boltzmann's constant (1.38×10^{-23} J/K)
T is the absolute temperature [deg K]

Note: The sign of the Fermi potential is determined by the doping in the channel. It is plus (+) for p-doped channels (NMOS transistors) and minus (-) for n-doped channels (PMOS) transistors. These equations do not have a closed form solution; they must be solved iteratively. Using the value of $Cs_{min} = 2.01 \times 10^{-10}$ F in the above example, through repeated iterations on a computer, we arrive at the values for N_{sub} and Φ_f shown below:

$N_{sub} = 1.812 \times 10^{17}$ cm^{-3}
Φ_f = 0.4315 V

Where A = 0.001 cm^2 and T = 300 K

The Debye length can be calculated from the following equation:

$$\lambda = \sqrt{\frac{2k \cdot T \cdot \varepsilon_o \cdot \varepsilon_{si}}{q^2 \cdot N_{sub}}} \qquad \text{(Equation 8.4)}$$

Sample MOSFET parameter calculation *(continued)*

The value of the depletion layer capacitance under flat band conditions (C_{sfb}) is given by:

$$C_{sfb} = \frac{\sqrt{2} \cdot A \cdot \varepsilon_o \cdot \varepsilon_{si}}{\lambda} \qquad \text{(Equation 8.5)}$$

Plugging the value of N_{sub} obtained above into these two equations yields:

$$C_{sfb} = 1.0789 \times 10^{-9} \text{ F}$$

Since the flat band capacitance (Cfb) is the series combination of C_{ox} and C_{sfb}, we know that:

$$C_{fb} = \frac{C_{ox} \cdot C_{sfb}}{C_{ox} + C_{sfb}} \qquad \text{(Equation 8.6)}$$

We have calculated the value of C_{ox} earlier, and we can combine this with the value of C_{sfb} just calculated, to arrive at a value for C_{fb} of:

$$C_{fb} = 4.9185 \times 10^{-10} \text{ F}$$

To get the value of the flat band voltage (V_{fb}) we need to take this value of C_{fb} and perform a linear interpolation on our capacitance plot. The two points on either side of the value of C_{fb} that we have just calculated are:

V = –0.9 V, C = 454 pF
V = –1.0 V, C = 540 pF

Thus, we can determine the value of the flat band voltage to be:

$$V_{fb} = -0.9440 \text{ V}$$

The next useful parameter to calculate is the surface charge density (Q_{ss}). Typically, this is divided by the electron charge (q) and expressed as Q_{ss}/q:

$$\frac{Q_{ss}}{q} = \frac{C_{ox}}{q \cdot A} \cdot \left| \phi_{ms} - V_{fb} \right| \qquad [1/cm^3] \qquad \text{(Equation 8.7)}$$

In this example,

$$\phi_f = 0.4315 \text{ V}$$
$$\phi_{ms} = -0.6 - \phi_f = -1.032 \text{ V}$$

where Φ_{ms} is the difference in work functions of the semiconductor (Si) and the gate poly (poly-Si).

Therefore, plugging in numbers we get that:

$$\frac{Q_{ss}}{q} = 4.9375 \times 10^{11} \text{ cm}^{-3}$$

Sample MOSFET parameter calculation *(continued)*

We can also calculate the fixed charge in the depletion layer (Q_b) and the threshold voltage V_t using the following equations:

$$Q_b = \pm q \cdot N_{sub} \cdot \frac{A \cdot \varepsilon_o \cdot \varepsilon_{si}}{Cs_{min}} \qquad [\text{Coulomb/cm}^2] \qquad \text{(Equation 8.8)}$$

$$V_t = V_{fb} + \left(2 \cdot \phi_f - \frac{A \cdot Q_b}{C_{ox}}\right) \qquad [\text{Volts}] \qquad \text{(Equation 8.9)}$$

Note: The sign of Q_b is determined by the doping in the channel. It is plus (+) for n-doped channels (PMOS transistors) and minus (-) for p-doped channels (NMOS) transistors. Plugging in values previously obtained, we have:

$$Q_b = -2.2785 \times 10^{-7}\ \text{C/cm}^2$$

$$V_t = 0.1711\ \text{V}$$

It should be obvious that it is best to create a program to automate this parameter extraction process.

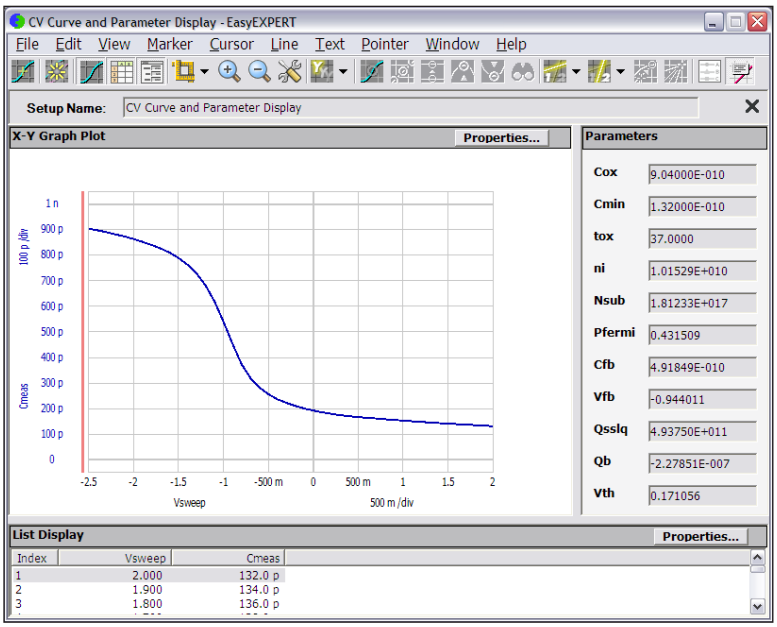


Figure 8.3. Automatic calculation of MOSFET capacitor parameters using Keysight EasyEXPERT software.

Quasi-static capacitance measurement

It is important to understand that a quasi-static capacitance versus voltage (QSCV) MOSFET device response to a stimulus is not a measurement technique. While it is true that certain measurement techniques are more likely to result in a QSCV response from a MOSFET than others, whether or not a stimulus results in a QSCV response depends upon a variety of process, device, and layout related factors in addition to the type of stimulus applied to the MOSFET. Therefore, as we discuss the step voltage QSCV measurement technique in the following section, it is worthwhile to keep in mind that in certain cases, QSCV device responses can also be observed using more traditional oscillator-based capacitance measurement techniques. Also, simply using the step voltage QSCV measurement technique does not guarantee a QSCV device response.

Obtaining a QSCV response

Before proceeding with an explanation of the voltage step QSCV measurement technique, it is useful to review the difference between a QSCV and a high-frequency capacitance versus voltage (HFCV) response.

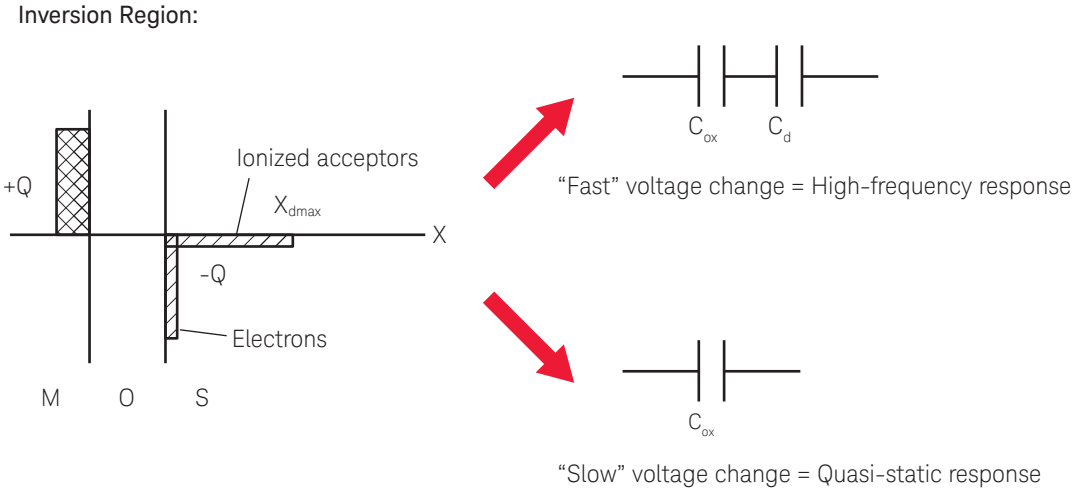


Figure 8.4. Quasi-static versus high-frequency capacitance versus voltage response for an NMOS transistor.

When the semiconductor channel is fully inverted (as shown above), the only sources for the free electrons are generation and recombination. If the applied voltage is modified gradually, then these mechanisms can supply mobile electrons causing the capacitor to act like a parallel capacitor with a capacitance value approximately equal to that of C_{ox}. This results in a classic QSCV response. However, if the voltage changes more rapidly than generation and recombination can respond, then the ionized acceptor region has to modulate to maintain charge balance. In this case, the total capacitance will be the series combination of the oxide capacitance and the depletion region capacitance (C_{ox} + C_d). This acts to reduce the overall value of the capacitance below that of C_{ox} alone, resulting in what is usually known as an HFCV response.

While many textbooks have attempted to supply equations to define what exactly is a sufficiently "slow" rate of voltage change to produce a QSCV response, measurement experience has shown that obtaining a QSCV response from a MOSFET is about equal parts art and science. In practice, it is usually best to start a sweep in the inversion region, and then proceed to sweep through the depletion and accumulation regions. In addition, it may be necessary to supply a DC bias to the capacitor for several seconds while it is inverted prior to the start of the sweep. Sometimes, it is also useful to shine a light onto the device to stimulate the generation of minority carriers before the start of the sweep. However, in this case it is essential to turn the light off before the start of the sweep, since the light disturbs the thermal equilibrium of the capacitor and distorts the measurement results.

The voltage ramp QSCV measurement technique

The classical method to perform a QSCV measurement involved using a voltage ramp created with an instrument such as the Keysight 4140B. The 4140B could generate a voltage ramp with a constant slope, which allowed the capacitance to be determined using the basic relationship between current, voltage and capacitance.

$$i = C\frac{dV}{dt} \Rightarrow C = \frac{i}{\frac{dV}{dt}} \qquad \text{(Equation 8.10)}$$

A typical QSCV measurement using this technique is shown below.

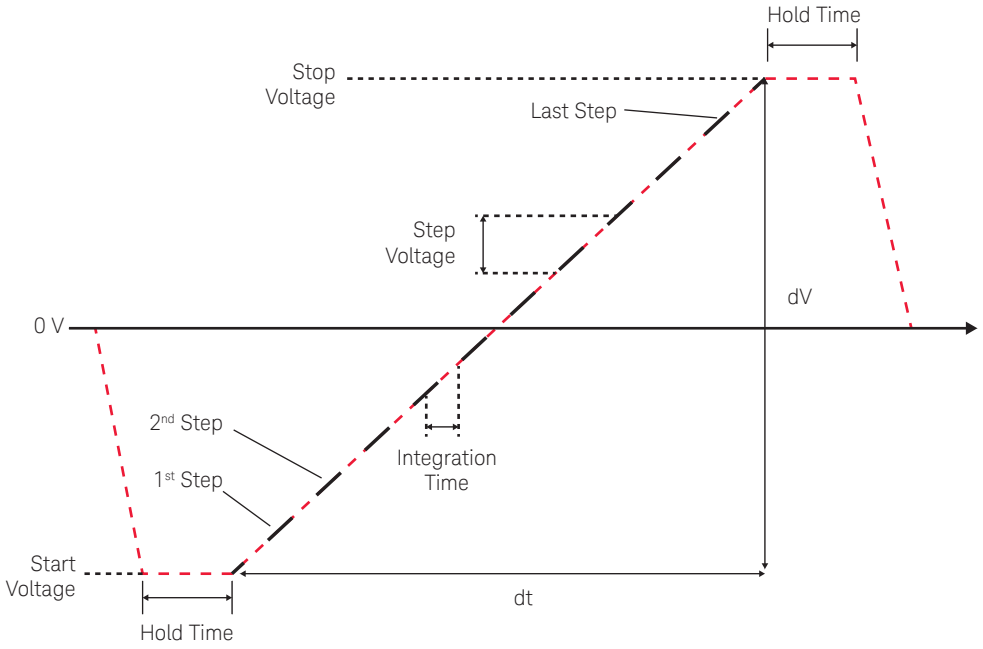
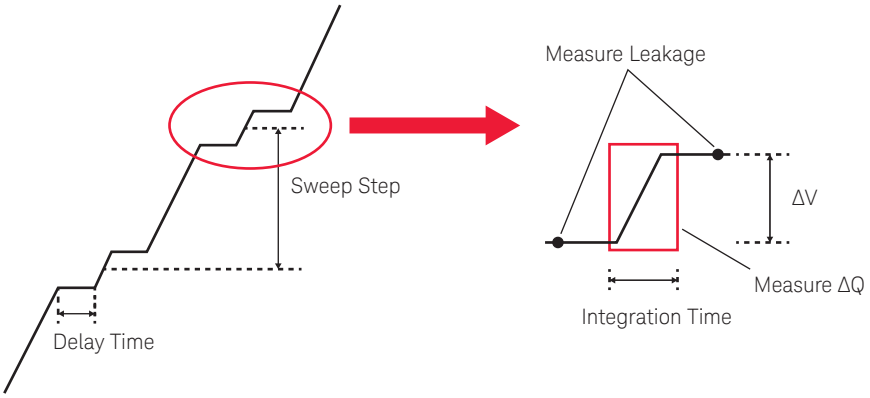


Figure 8.5. The classical voltage ramp QSCV measurement technique using a Keysight 4140B.

Since the 4140B is an obsolete instrument that is no longer sold or supported, this QSCV measurement technique has been replaced by the voltage step QSCV measurement technique.

The voltage step QSCV measurement technique

The step voltage QSCV measurement technique does not use a voltage ramp or oscillator to measure CV. Instead, it uses SMUs and performs a measurement that it very similar to a standard IV sweep. The key difference is that, at each point on the sweep, a user-specified ΔV is applied. The corresponding charge ΔQ that results can be calculated by measuring the current induced by the ΔV and numerically integrating the area under the current versus time curve. This then allows the capacitance at each point along the sweep to be calculated by dividing ΔQ by ΔV. The figure shown below illustrates this technique:



$$C = \frac{\Delta Q}{\Delta V} \qquad \Delta Q = \int i dt \sim \sum i \Delta t$$

Figure 8.6. The step voltage QSCV measurement technique.

Obviously, the value for ΔV chosen has to be less than or equal to the size of the sweep step.

In addition to measuring capacitance at each point in the sweep, the step voltage QSCV algorithm also supports the ability to remove leakage currents caused by electron tunneling through the gate dielectric. This is done by measuring the leakage current before and after applying the ΔV, and then subtracting out the leakage current before calculating the capacitance. Unfortunately, this technique is ultimately limited by the current measurement range in which the capacitance sweep is being made. If the leakage current exceeds the current measurement range, then an error will occur. For extremely leaky gate dielectrics (with oxide thicknesses of approximately 25 Angstroms or less), the step voltage QSCV measurement technique cannot be used. In this case, the only solution is to make extremely high frequency measurements using a conventional capacitance meter technique. This technique will be discussed later in this chapter.

A valid question to ask is: how well does the voltage ramp QSCV technique correlate with the voltage ramp QSCV technique? The answer is: Quite well, as long as you pick the measurement parameters correctly.

- The same start, stop and step values can be used
- The voltage step can be determined by multiplying the 4140B ramp rate by the 4140B integration time
- The integration time should be set to the same value as the 4140B integration time
- The leakage compensation feature is disabled

The voltage step QSCV measurement technique *(continued)*

The following graph shows actual measurement data taken on a device using the two different QSCV measurement techniques.

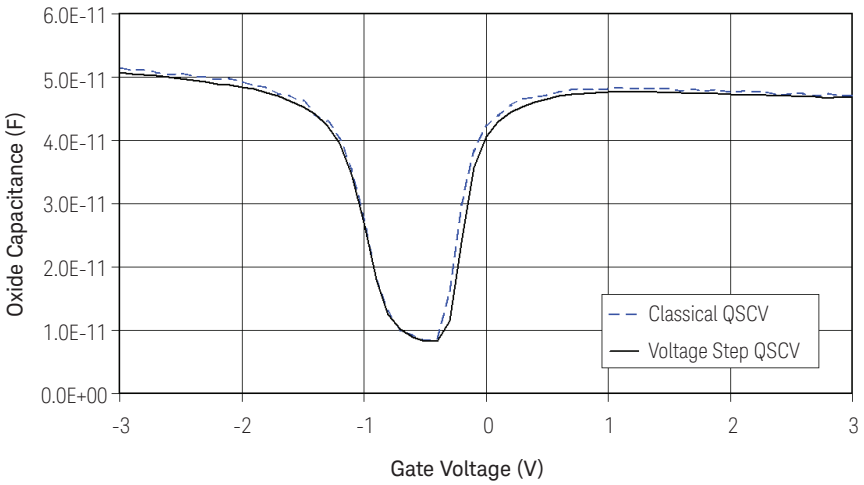


Figure 8.7. Correlation between the classical (voltage ramp) and step voltage QSCV measurement techniques.

As this graph shows, the correlation between the two techniques is quite good. If you have trouble correlating the two techniques, then check whether one or more of the following conditions applies:

- The 4140B ramp rate (dV/dt) is too steep
- The **delay time** or **integration time** is too short
- The leakage current compensation feature is turned on

QSCV offset compensation

The main parasitic element affecting the step voltage QSCV measurement technique is the offset capacitance between the triaxial cables. Therefore, it is sufficient for the QSCV measurement technique to support simple offset compensation. With the probe tips open, the instrument performs a capacitance measurement to determine the offset capacitance. This value is stored and automatically subtracted from subsequent QSCV measurements.

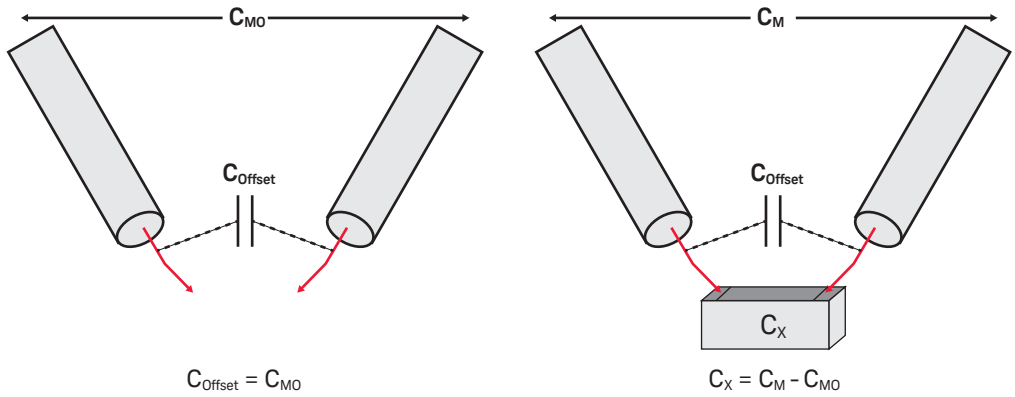


Figure 8.8. The QSCV measurement technique can automatically subtract the innate offset capacitance of the triaxial cables.

This compensation works well for the QSCV measurement technique and gives excellent correlation with oscillator-based capacitance measurement techniques.

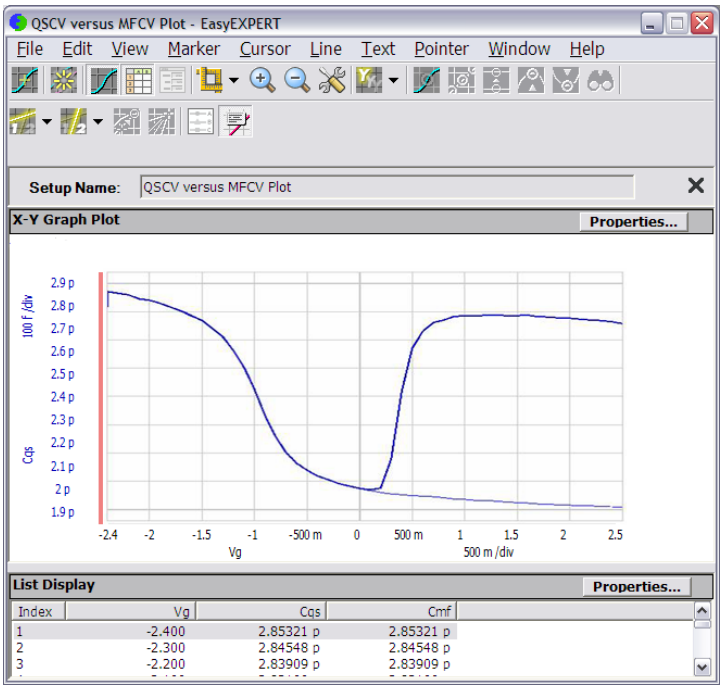


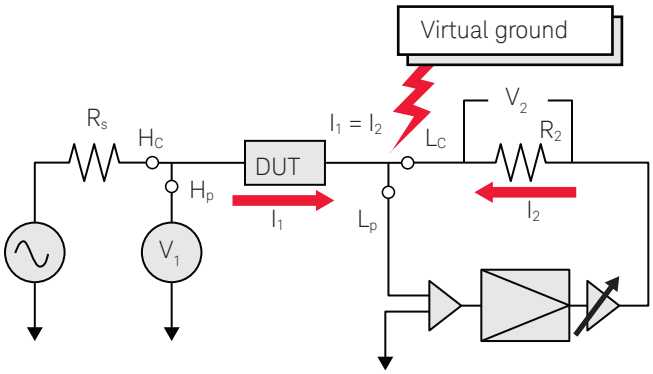
Figure 8.9. This superposition of quasi-static CV and high-frequency CV plots shows that you can get excellent measurement correlation between the two techniques.

Low frequency (< 5 MHz) capacitance measurement

For the purposes of this discussion we will consider low frequency measurements to be anything that uses an oscillator that has a frequency from a few Hertz up to 5 MHz.

Capacitance meter operation

For many people, the operation of a capacitance meter is a great mystery. However, the basic principal of the auto balancing bridge circuit is not difficult to understand. It is illustrated in the circuit schematic shown below.



$$V_2 = I_2 \times R_2$$

$$Z = \frac{V_1}{I_2} = \frac{V_1 R_2}{V_2}$$

Figure 8.10. The auto balancing bridge capacitance measurement method.

The auto balancing bridge can be conceptualized as an operational amplifier (op amp) circuit. Ohm's law applies: V = I x R. The device is stimulated by an AC signal, with the actual voltage applied to the device being monitored at the H (high) terminal. The L (low) terminal is driven to 0 volts by the virtual ground of the op amp. The current, I_2, through the range resistor, is equal to the current through the DUT. Therefore, the output voltage is proportional to the current through the device. The voltage and current are automatically balanced, thus giving rise to its name. To cover a wide frequency range, a null-detector and a modulator are used instead of an amplifier in practical circuits. The four inputs to the capacitance meter are defined below:

H_c: Signal source
H_p: Potential meter
L_c: Current meter
L_p: Potential meter to lock the phase of measurement signal

The 4TP measurement method

The most common measurement technique used with the auto balancing bridge capacitance meter is the 4 terminal pair (4TP) measurement method. In this procedure, the Hc and Hp terminals and the Lc and Lp terminals are shorted together as shown below.

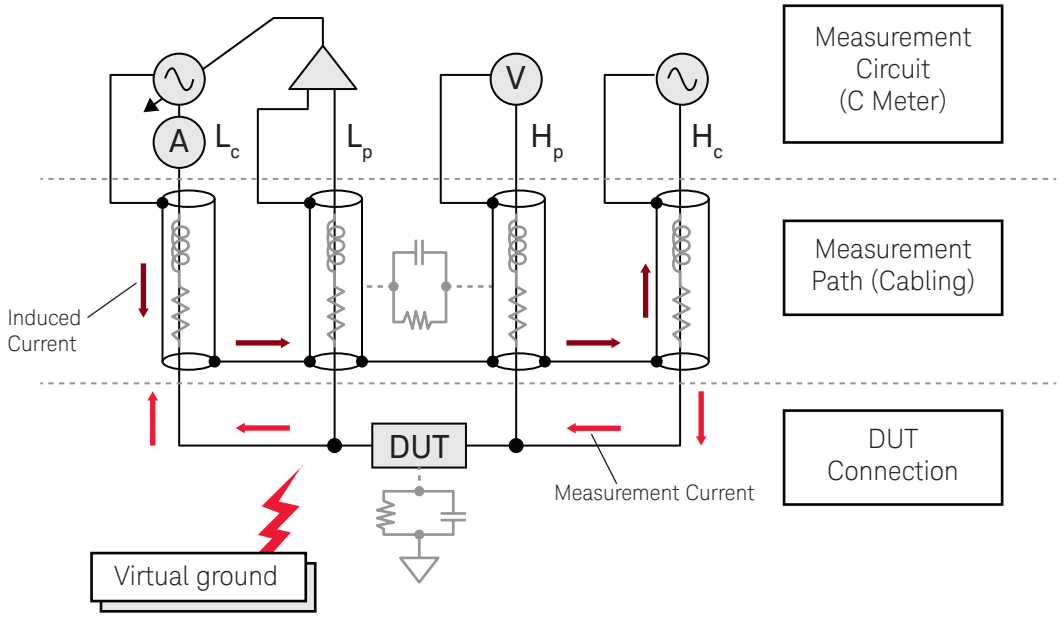


Figure 8.11. The 4 terminal pair (4TP) measurement method (including parasitic elements).

The H_p and H_c terminals are commonly referred to as the CMH (capacitance meter high) connection and the L_p and L_c terminals are commonly referred to as the CML (capacitance meter low) connection. There is residual inductance and resistance in the measurement path (cabling) and stray capacitance both between the cables and between the DUT and ground (which are also shown in Figure 8.11). The user must perform compensation through the measurement path (cables) in order to eliminate the effects of these parasitic elements or the accuracy of the measurement will be seriously degraded.

The CML terminal of the 4TP scheme is a virtual ground and it is very important not to tie this point to earth ground. If the CML terminal contacts earth ground, then the auto balancing bridge circuit will not be able to balance, resulting in measurement errors. The outer conductor (shield) of the capacitance cables has the same potential as the virtual ground, and it needs to be floated. In addition, the four outer shields should be connected together just before the probe tips with cables that are short in length. This creates a return path for the current that is induced in the shield, which stabilizes the series inductance of the cables. Unless the outer shields are shorted and the series inductance stabilized, the mere act of changing the cable separation will cause large variations in cable inductance that immediately invalidate any cable compensation that has been performed.

The importance of proper compensation

For capacitance meters, the “calibration plane” defines the point (normally the BNC outputs of the capacitance meter) at which the user can obtain the specified measurement accuracy of the capacitance meter. Of course, for parametric tests, we need to use cabling to test our DUT. In the case of the 4TP measurement method, this introduces a variety of sources of error due to residual parasitic elements in the test fixturing (cables).

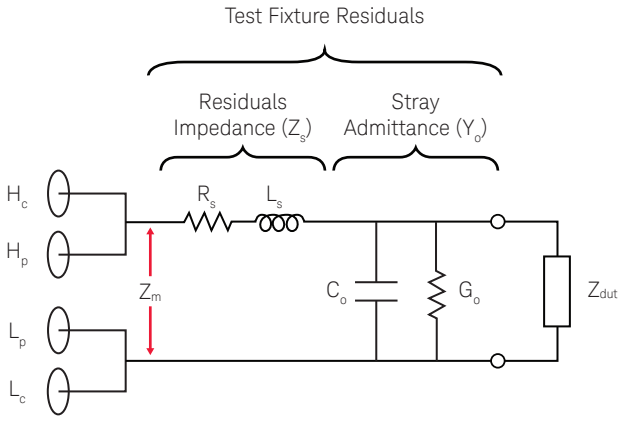
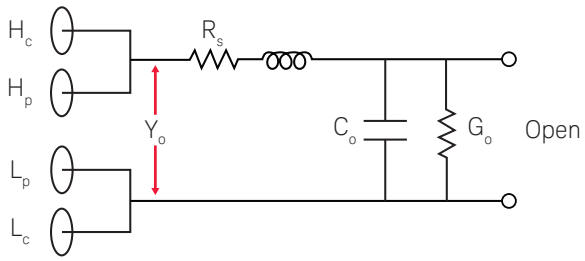


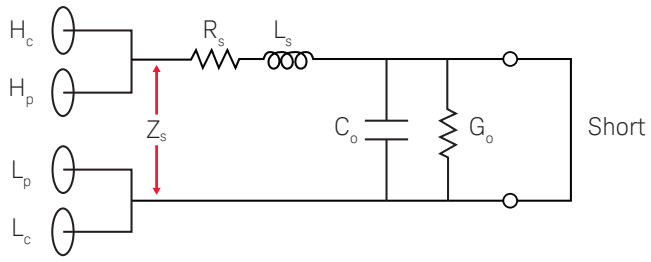
Figure 8.12. The test fixture residuals that you must account for in order to make accurate capacitance measurements through a test fixture (cable).

To remove the effects of these test fixture residuals, we perform two measurements: one with the test terminals open and one with the test terminals shorted as shown below.



$$Y_o = G_o + j\omega C_o$$

$$\left(R_s + j\omega L_s << \frac{1}{G_o + j\omega C_o} \right)$$

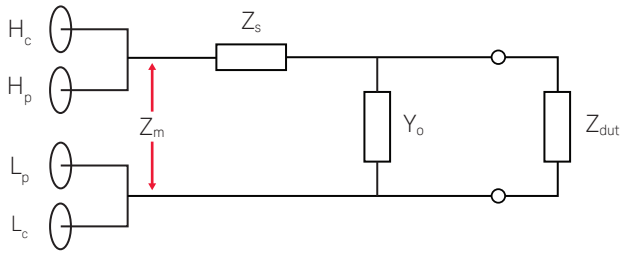


$$Z_s = R_s + j\omega L_s$$

Figure 8.13. Using Open/Short compensation to determine test fixture residuals.

The importance of proper compensation *(continued)*

Using these two values, we can calculate the true value of Z_{dut} from the measured value (Z_m) using the equation shown below.



$$Z_{dut} = \frac{Z_m - Z_s}{1 - (Z_m - Z_s)\,Y_o}$$

Figure 8.14. The true value of Z_{dut} can be calculated using the Open and Short values.

In addition to Open/Short cable compensation, phase compensation should always be performed before making measurements for the first time (when this feature is supported). Phase compensation improves the bridge balance stability at high frequencies and minimizes phase shift effects due to variations in frequency and cable length. Phase compensation should be performed before performing the Open/Short cable compensation, and it generally does not need to be done again unless you change your measurement configuration. The L_c and L_p terminals need to be shorted together during the phase compensation. The procedure to follow to perform phase compensation varies by instrument, so you need to refer to your instrument's manual to understand exactly how to do this (if it is supported).

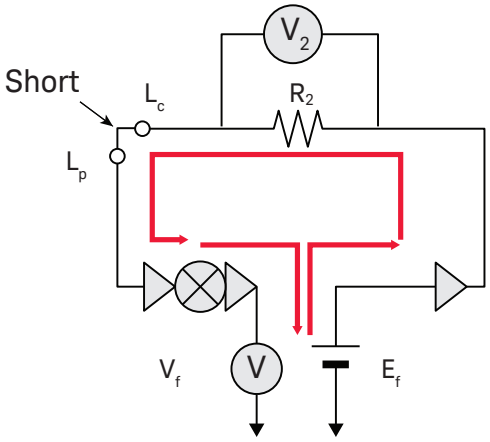


Figure 8.15. Phase compensation calculates the characteristic of the null loop circuit, which improves repeatability and phase accuracy.

For measurements at or below 5 MHz, load compensation is not always necessary; its effect on measurement accuracy at these frequencies depends on the exact fixturing that you are using. However, for measurements above 5 MHz you should always perform Open/Short/Load compensation or significant measurement error will occur. This requires the on-wafer measurement of some kind of calibrated resistor, and it is normally done with RF wafer probes. We will return to this topic when we discuss high-frequency capacitance measurement.

On-wafer capacitance measurement

There are some important measurement issues to take into account when making on-wafer capacitance measurements. The key here is how the H_c and H_p and L_c and L_p terminals connect to the device under test and the wafer prober. Many engineers making on-wafer capacitance measurements know from experience that they get much better measurement results by having the L_c and L_p terminals contact the wafer (DUT) and having the H_c and H_p terminals contact the wafer chuck when using the 4TP cabling method.

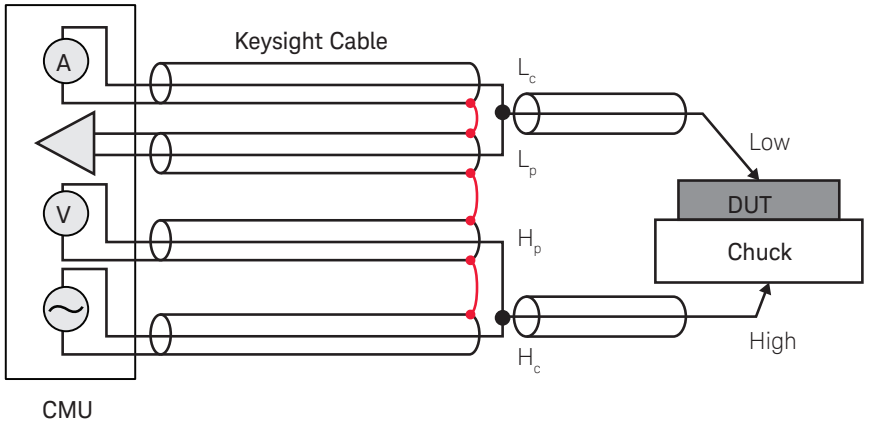


Figure 8.16. The proper way to make on-wafer capacitance measurements using the 4TP method.

However, most engineers do not know why this provides better measurement results. The reason that better measurement results can be obtained using this technique can be understood by realizing that on a wafer prober, you have to be concerned with chuck-to-ground capacitance. When making measurements at frequency, the chuck capacitance acts as a sneak path to ground. Consider the two different cases shown below.

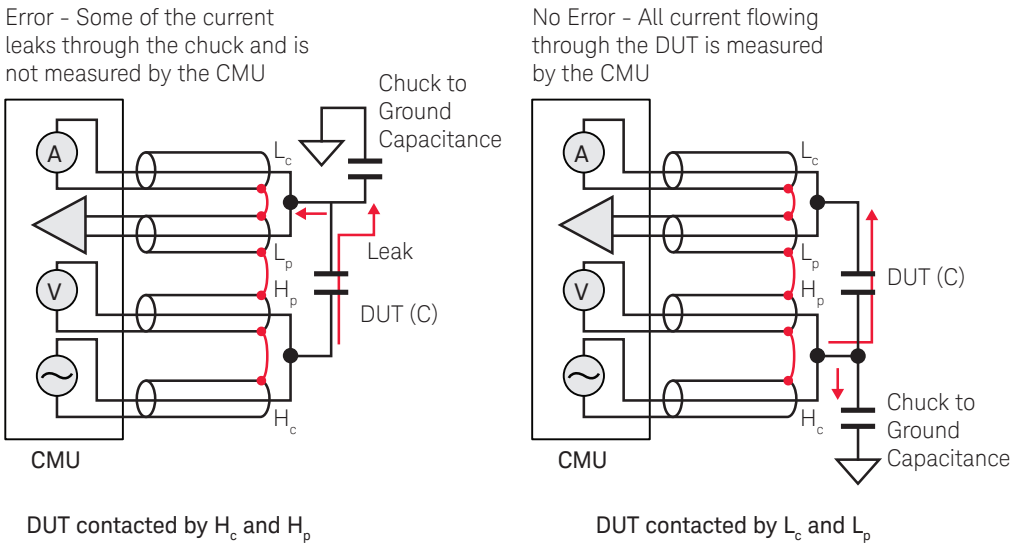


Figure 8.17. The reason why measuring from the Lc and Lp terminals eliminates measurement error.

The actual capacitance measurement is performed by the L_c and L_p terminals, so if these terminals directly contact the DUT then all of the current flowing through the DUT is measured by the capacitance meter and the current flowing to ground through the wafer chuck is irrelevant. Also, when measuring in this fashion, you need to reverse the applied DC bias, but this is a rather trivial consideration that can easily be automated.

On-wafer capacitance measurement *(continued)*

The other reason you are advised to connect the H_p and H_c terminals to the wafer chuck when making on-wafer capacitance measurements is noise. The wafer chuck acts like a large antenna, collecting random electromagnetic noise. If the L_p and L_c terminals are connected to the wafer chuck, then this noise will flow directly into the capacitance meter. In extreme cases, the capacitance meter may give an error message and return an unbalanced status error (UNBAL). However, if the H_p and H_c terminals are connected to the wafer chuck, then noise from the wafer chuck is attenuated when it passes through the DUT.

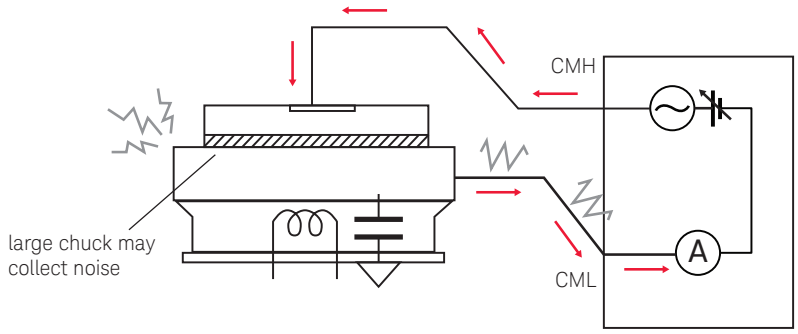
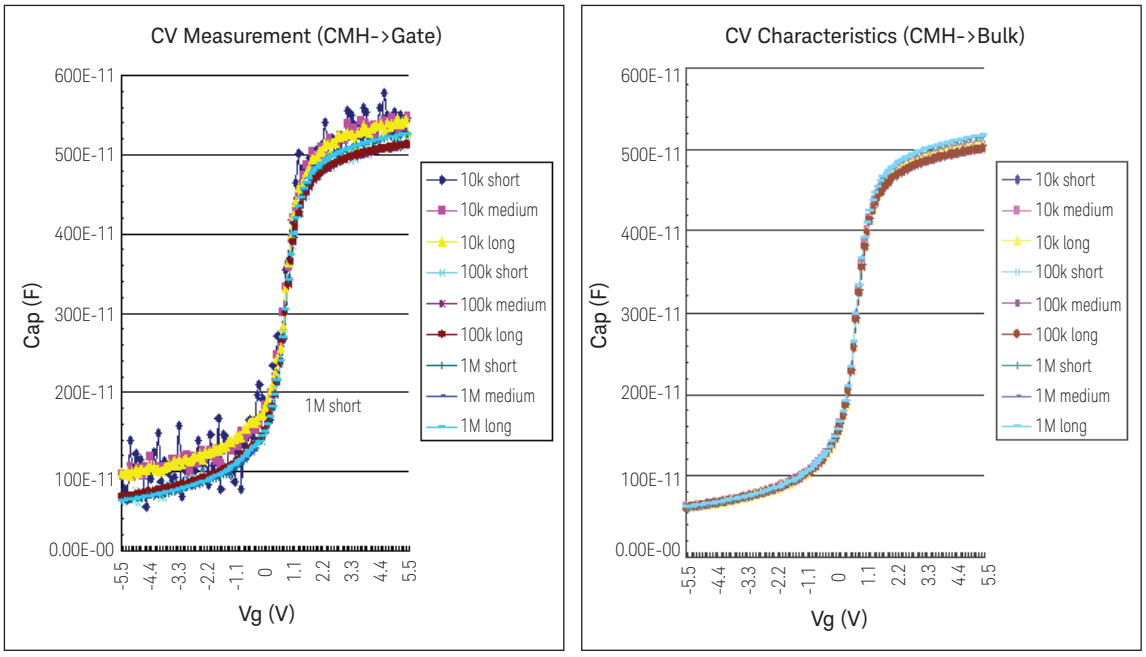


Figure 8.18. A large chuck can inject noise directly into the ammeter if the Hc and Hp terminals are connected to the DUT.

The following graphs compare the measurement results obtained for different frequencies and integration times when connecting the CMH terminals to both the gate and to the bulk.



CMH connected to gate

CMH connected to bulk (wafer chuck)

Figure 8.19. Measurement data comparing the results obtained connecting the CMH terminal to the gate versus connecting the CMH terminal to the wafer chuck (bulk).

On-wafer capacitance measurement *(continued)*

Almost all B1500A EasyEXPERT application tests for capacitance measurement are designed to have the CML terminals contact the DUT and the CMH terminals contact the wafer chuck. A useful feature is incorporated in these application tests. The application tests automatically take care of inverting all of the signs on the measurement inputs so that you do not have to convert them manually. The only exceptions to this feature are the few application tests labeled “simple”, since these are designed to be used with packaged devices for product demos.

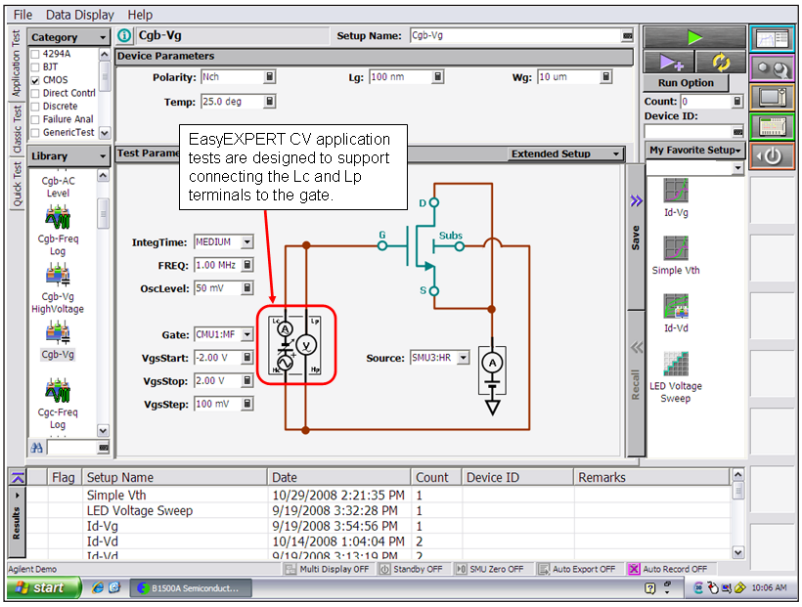


Figure 8.20. The terminal connections on a B1500A EasyEXPERT CV application test.

High frequency (>5 MHz) capacitance measurement

The choice of selecting 5 MHz as a dividing point between "low-frequency" and "high-frequency" measurement is not arbitrary. It is in fact a natural dividing line between these two measurement regimes. Above 5 MHz, new measurement techniques and schemes need to be employed in order to obtain satisfactory measurement results. For example, at frequencies above 5 MHz it becomes increasingly difficult or impossible to obtain good measurement results with standard DC probes; instead RF probes that require a ground-signal (GS) or ground-signal-ground (GSG) pad layout become mandatory. Also, in addition to having a pad structure that supports these types of RF probes, the actual layout of the device takes on critical importance. In fact, the most common reason for failure when trying to achieve satisfactory capacitance measurement results above 5 MHz is not due to the measurement equipment or cabling, but improper test structure design. Therefore, it is important that you read and understand this section before you undertake the design of your capacitance test structures.

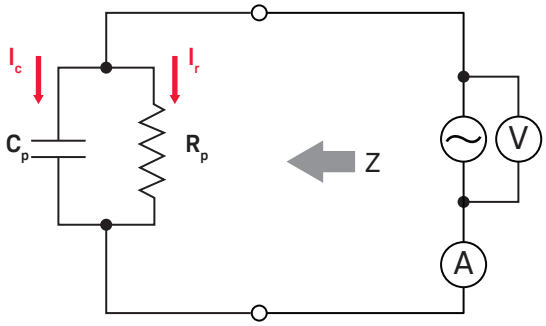
Thin Gate Dielectric Fundamentals

Many years ago, as transistor gate dielectrics became extremely thin (~25 Angstroms or less), the issue of gate leakage current due to electron tunneling moved to the forefront of parametric testing. While this is not a difficult issue to deal with for standard current versus voltage (IV) parametric measurement, it presents some challenges for capacitance versus voltage measurement (CV). To measure thin gate (leaky) dielectric MOSFET capacitors successfully, three essential requirements must be met:

1. A high measurement frequency (>5 MHz)
2. A device model suitable for thin-gate dielectrics
3. The elimination of wafer chuck effects

In reality, it is difficult to achieve all of the above requirements simultaneously, making on-wafer measurement of thin gate dielectrics one of the most challenging areas of capacitance measurement. We will examine each of these requirements in-turn.

To start with, you have to consider the basic two-element capacitance model shown below:



If I_r is large → R_p is small

Figure 8.21. The two-element capacitance model.

The equivalent impedance of this circuit is given by:

$$Z = \frac{\left(\frac{1}{C_p j\omega} \bullet R_p\right)}{\frac{1}{C_p j\omega} + R_p} = \frac{R_p}{1 + R_p C_p j\omega} = \frac{R_p - R_p^2 C_p j\omega}{1 + R_p^2 C_p^2 \omega^2}$$

(Equation 8.11)

Thin Gate Dielectric Fundamentals *(continued)*

We can rewrite this equation to show the real and imaginary parts of this impedance:

$$\mathrm{Re}\{Z\} = \frac{R_p}{1+R_p^2C_p^2\omega^2} \qquad \mathrm{Im}\{Z\} = \frac{-R_p^2C_p\omega}{1+R_p^2C_p^2\omega^2}$$

The dissipation factor (D) is given by the ratio of the real part to the imaginary part:

$$D = \frac{\mathrm{Re}\{Z\}}{\mathrm{Im}\{Z\}} = \frac{\dfrac{R_p}{1+R_p^2C_p^2\omega^2}}{\dfrac{-R_p^2C_p\omega}{1+R_p^2C_p^2\omega^2}} = -\frac{1}{R_pC_p\omega} \qquad \text{(Equation 8.12)}$$

For thin gate dielectrics, large leakage currents imply that the value of the equivalent parallel resistor is small. Therefore, the current flowing through the capacitor is decreased, making it harder to be measured accurately. To increase the current flow through the capacitor (decrease the D), a higher measurement frequency is required to reduce the impedance of the capacitor (which is equal to 1/Cj**ω**). Looking at this in the impedance plane, a smaller R_p implies a larger D (dissipation). This means that the accuracy of the capacitance measurement is degraded and that the influence from external noise is increased. For these reasons, thin gate oxide measurement requires higher measurement frequencies. The relationship between R_p, C_p and D is shown below.

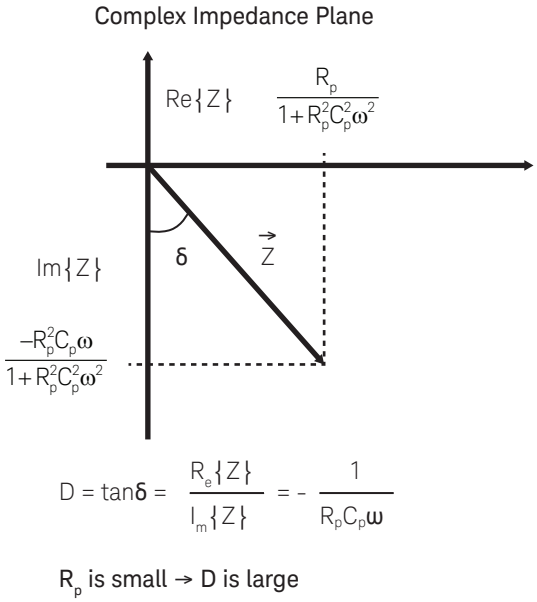


Figure 8.22. The relationship between parallel resistance (R_p) and dissipation (D).

Note: The above results imply that the dissipation can be made arbitrarily small by increasing the frequency. However, as we will see later when we discuss the three-element capacitance model, this is not necessarily true for thin gate dielectric devices.

Thin Gate Dielectric Fundamentals *(continued)*

Most capacitance meters or impedance analyzers have the ability to plot the value of D in real time. For the case of thick gate dielectrics, it is quite common to see values of D much smaller than 1, indicating that most of the measurement current is flowing through the capacitor rather than the parasitic resistor. However, for thin gate dielectrics, the user's expectations for an acceptable value of D need to be adjusted. Often, even increasing the frequency does not reduce the value of D to a value less than 1. It is difficult to come up with a clear "rule of thumb" for an acceptable dissipation factor for thin gate dielectrics, but a value of less than 5 is normally sufficient to achieve a satisfactory capacitance measurement.

Although it is evident that higher frequencies are required to measure thin gate dielectric transistor gate capacitance, the two-element model cannot explain the results observed at higher frequencies.

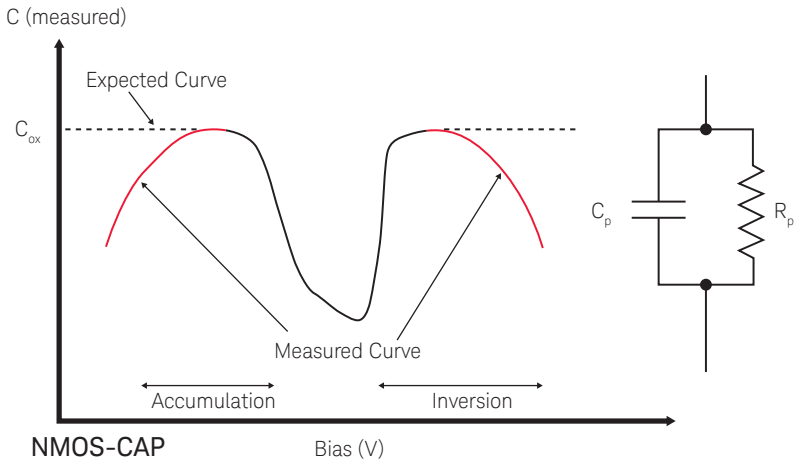


Figure 8.23. The two-element capacitor model cannot explain the results seen when measuring the gate capacitance on thin gate dielectric transistors at higher frequencies.

In order to correctly characterize thin-gate transistor dielectrics, we will proceed to address the three-element capacitor model shown below that includes contact and substrate resistance.

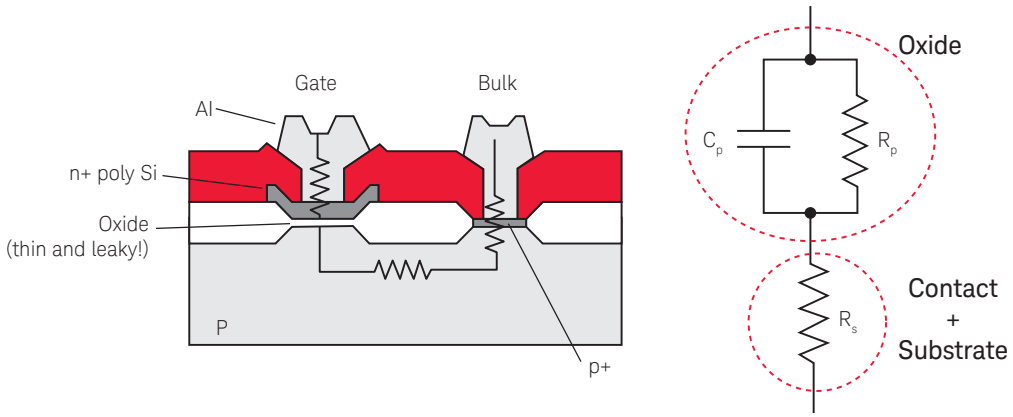


Figure 8.24. The three-element capacitor model which includes contact and substrate resistance.

Thin Gate Dielectric Fundamentals *(continued)*

For conventional thick gate dielectrics, the equivalent parallel resistance R_p is much larger than the series resistance R_s. This means that R_s can be disregarded when measuring at low frequency ranges under 1 MHz, which also explains why the parallel model was valid in the past for the thicker gate dielectrics.

In the past, many engineers noticed what seemed to be “negative capacitance” when measuring thin gate oxides. An example of the observed measurement results is shown below.

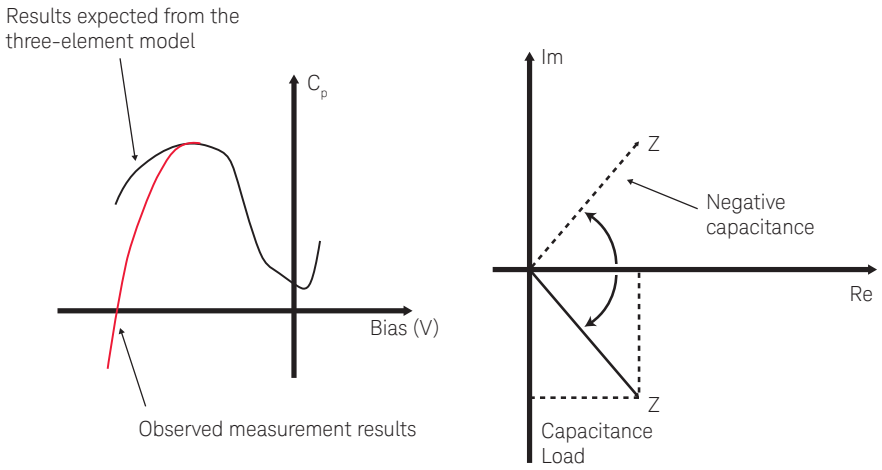


Figure 8.25. Illustration of negative capacitance effect observed when measuring thin gate dielectric capacitors.

Thin Gate Dielectric Fundamentals *(continued)*

Negative capacitance is implied when the measured impedance has a positive value for its imaginary part. In general, positive imaginary impedance implies inductance. This led many to try to explain the observed negative capacitance as an inductance caused by wiring in the wafer prober or even by the generation and recombination of holes and electrons. However, none of these approaches were successful.

Through studies, Keysight learned that the observed negative capacitance is caused by the four-terminal pair measurement technique used for on-wafer measurement, not from the device itself. The following circuit diagram shows the whole measurement system including the wafer chuck.

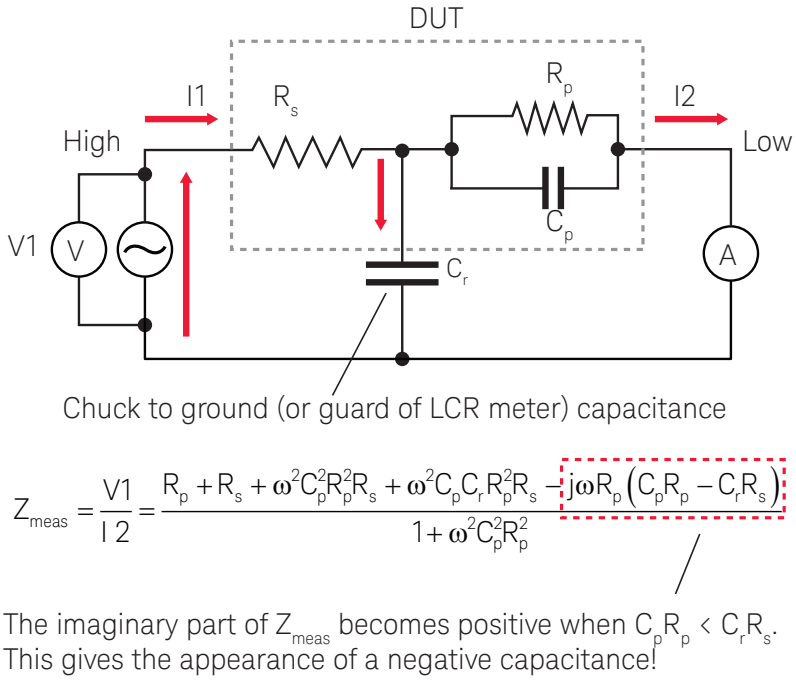


Figure 8.26. The explanation of the negative capacitance effect seen in thin gate dielectrics.

The key takeaway here is capacitive coupling occurs between the wafer chuck and the return path of the signal (outer shield of the BNC cable or earth ground). When using a four-terminal pair, part of the signal flowing through the DUT leaks through the wafer chuck via capacitive coupling. In this case the measured impedance is described by the equation shown above. From this equation, if the product of C_p and R_p is smaller than the product of C_r and R_s, the imaginary part of the measured impedance becomes positive. Because R_p comes from the direct tunneling of electrons through the gate dielectric, R_p rapidly becomes smaller with increasing gate bias. This means that a negative capacitance is observed when the gate bias goes over a value that satisfies the condition of $C_p \times R_p < C_r \times R_s$. This analysis explains the observed measurement results when performing the CV measurements on thin gate oxide devices. Note: The best way to eliminate this effect is through structure design (minimizing R_s), rather than through improved measurement techniques. More information on this can be found in a paper by Y. Okawa, et al, from ICMTS 2003 “The negative capacitance effect on the C-V measurement of ultra-thin gate dielectrics induced by the stray capacitance of the measurement system”.

Optimizing CV measurement results through proper structure design

In order to improve measurement performance, we want to minimize the contact resistance (R_s) as much as possible. From our prior discussion, we found that measurement accuracy deteriorates when R_s becomes comparable to R_p. Therefore, reducing R_s through prudent layout design should help to improve our measurement results. Also remember that when $C_p \times R_p < C_r \times R_s$, a "negative capacitance" will be observed. Therefore, by minimizing R_s we can eliminate this effect. The following figure is a good example of a test structure designed to minimize R_s.

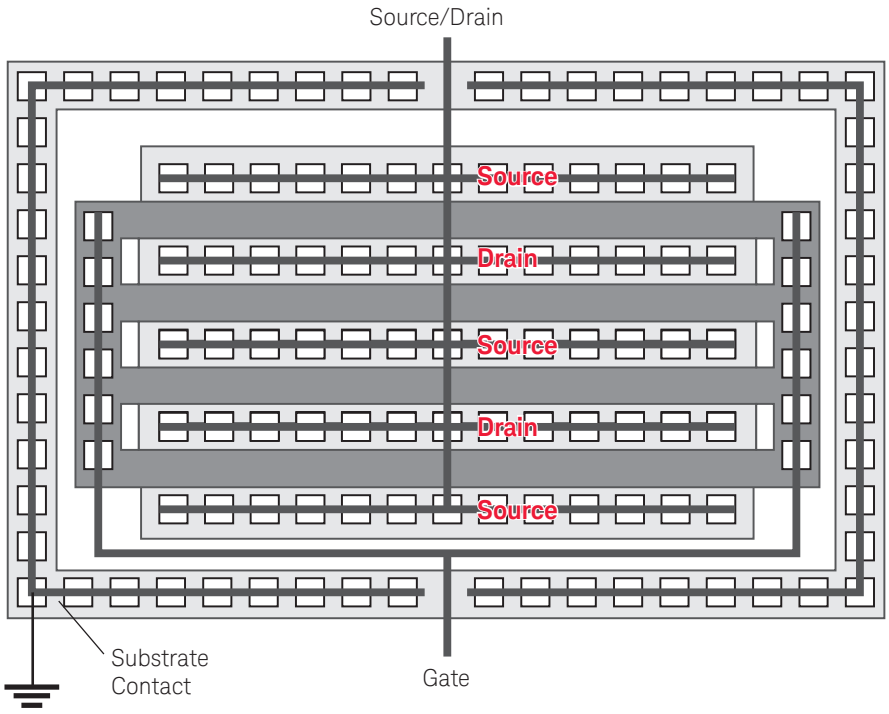


Figure 8.27. A capacitance measurement structure designed to minimize R_s and eliminate the negative capacitance effect.

The details of this structure are not important (and indeed the details are not shown). The important points to remember are:

1. Design a symmetrical structure
2. Include lots of contact vias to minimize contact and substrate resistance
3. Use a ground-signal (GS) or ground-signal-ground (GSG) layout to connect to the pads

Note: The use of these types of test structures to minimize R_s was first proposed by J. Schmitz, et al, at ICMTS 2003: "Test structure design considerations for RF-CV measurements on leaky dielectrics".

Pad layout considerations

As shown in the previous discussion, a dedicated MOSFET capacitance structure is required to obtain good capacitance measurement results at high frequencies. In addition to the structure design, the actual layout of the pads is also important. Putting some forethought into the pad layout can save you many measurement headaches later. Although high-frequency capacitance measurements require RF probes for optimal results, this does not necessarily mean that ground-signal-ground probe tips are always necessary. The following illustration shows two basic MOSFET capacitor layouts (NMOS and PMOS) that only require a ground-signal probe tip in order to make the capacitance measurement.

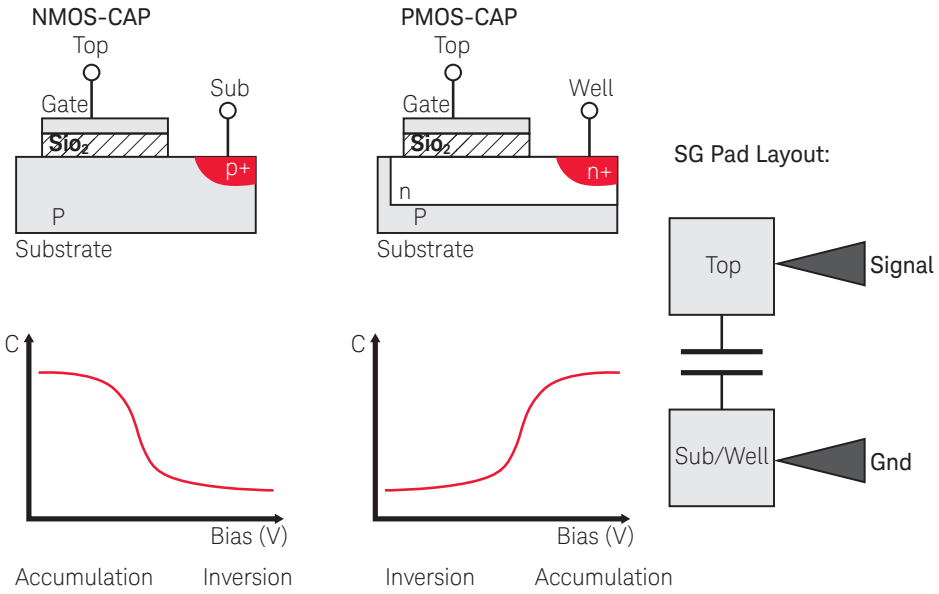


Figure 8.28. Using a signal-ground (SG) style pad layout, you can perform simple CV sweeps (gate to substrate) on both NMOS and PMOS capacitors.

Pad layout considerations *(continued)*

If you want to measure both gate to substrate and gate to source/drain capacitance, then the following (GSG) scheme meets this need and only requires the use of two probes at one time.

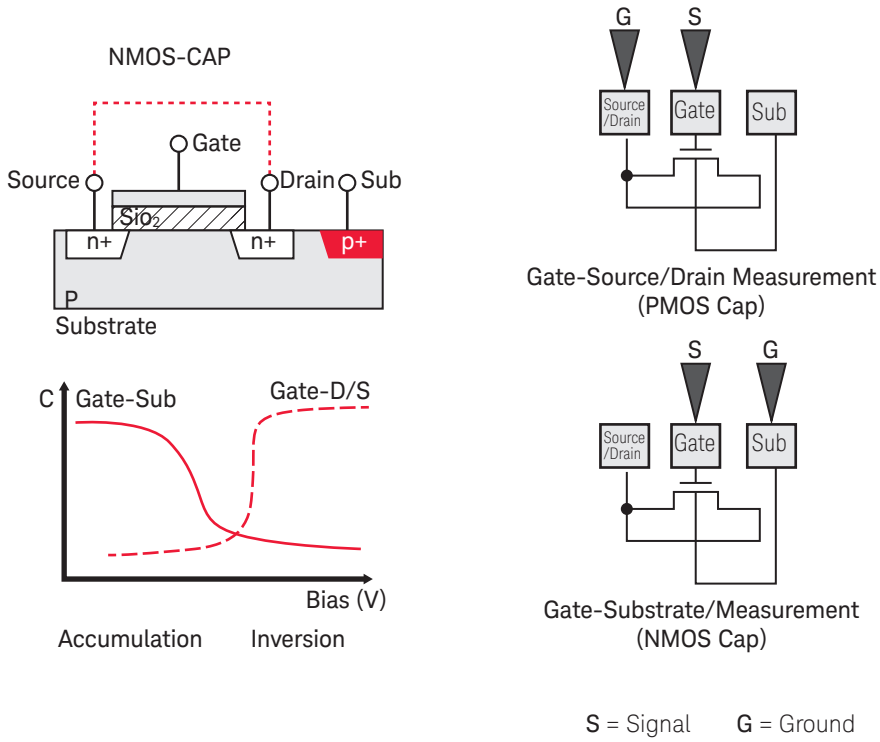


Figure 8.29. Using a carefully designed three-terminal (GSG) style pad layout, you can measure both gate to substrate and gate to drain/source capacitance.

Applying the above scheme can save you some money, since GS probe tips are less costly than GSG probe tips.

Open/Short/Load calibration

As previously mentioned, for capacitance measurements above 5 MHz, simple Open/Short calibration is no longer sufficient to guarantee measurement accuracy. The following plot compares the measurement results obtained using the 4TP cabling technique with a Keysight 4294A Impedance Analyzer (which has a frequency range up to 110 MHz) for both Open/Short and Open/Short/Load calibration.

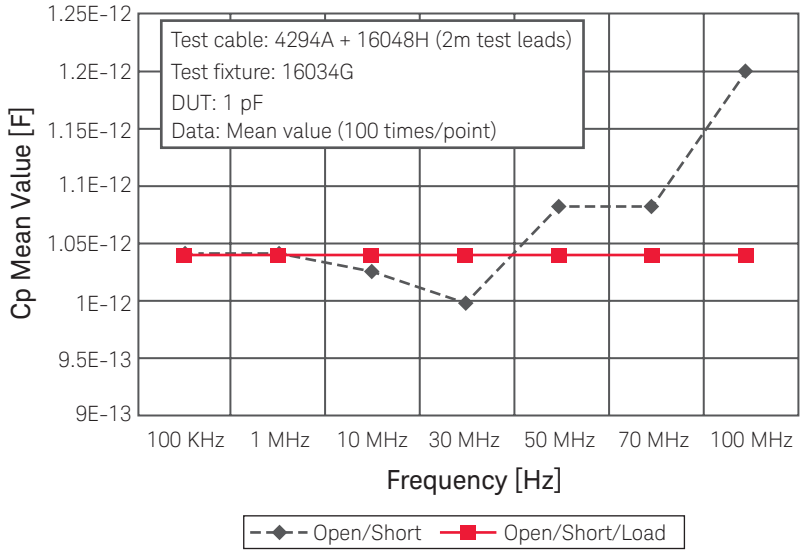
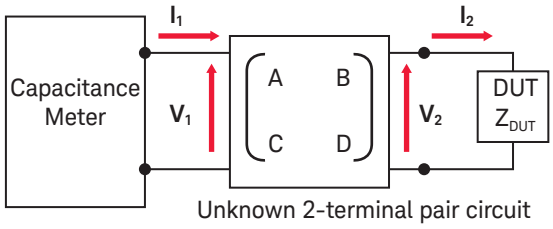


Figure 8.30. Measurement error occurs at higher frequencies unless a load calibration is performed.

To learn more about the Keysight 4294A Impedance Analyzer, you can refer to Keysight application note 4294-3 (publication number 5988-5102EN). As this measurement data shows, significant measurement error occurs above 5 MHz unless the Load compensation is performed in addition to the basic Open/Short calibration.

The Open/Short compensation scheme previously discussed was an oversimplification. The actual Open/Short/Load scheme treats the cabling as an unknown two-terminal pair circuit, and it determines the true value of the DUT impedance as shown in the diagram shown below.



$$Z_{DUT} = \frac{Z_{Std}(Z_0 - Z_{SM})(Z_{XM} - Z_S)}{(Z_{SM} - Z_S)(Z_0 - Z_{XM})}$$

Z_0 : Measured value of OPEN
Z_{SM} : Measured value of standard DUT
Z_{XM} : Measured value of the DUT
Z_S : Measured value of SHORT
Z_{Std} : True calue of the standard DUT
Z_{DUT} : Corrected value of the DUT

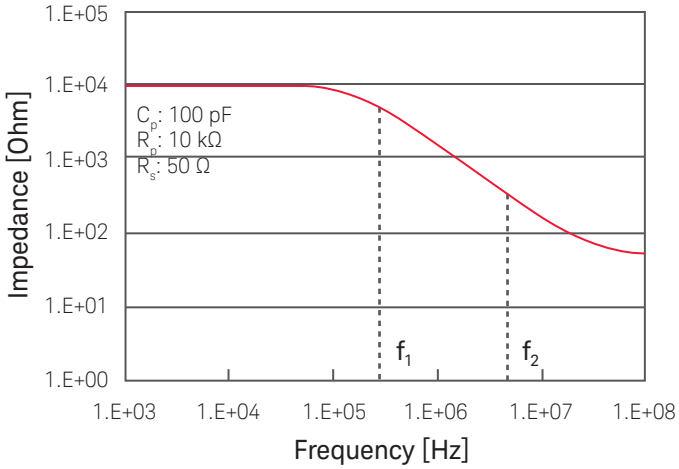
Figure 8.31. The Open/Short/Load compensation scheme actually performs an analysis on the unknown network connecting the capacitance meter with the DUT.

Open/Short/Load calibration *(continued)*

It cannot be emphasized enough that the LOAD portion of the calibration requires a well-calibrated resistance standard, and that this calibration has to occur at the ends of your probe tips. The accuracy of all of your subsequent high-frequency measurements depends upon the accuracy of the load standard, so it is very important to use a standard with known characteristics. Analytical wafer prober companies that support RF probing can supply standards for this purpose. In addition, you must make sure that the calibration plane for the Open/Short/Load calibrations is always the same or you will get inaccurate measurement data. For example, when using probe cards, one common mistake that happens is performing Open/Short compensation at the probe tips followed by performing the Load compensation using a special load standard built into the probe card. Since the calibration plane is not equal in all three cases, this results in inaccurate measurement at high frequencies.

Calculating the values of the three element capacitor model

The three-element capacitor model obviously requires the determination of three parameters: C_p, R_p and R_s. While this might seem to require three measurement points, it is actually not the case. Remember that every measurement of impedance actually returns two pieces of data: magnitude and phase. This means that two impedance measurements translates into four pieces of data, which is more than sufficient to determine the three unknown elements. This concept can be applied to the three-element capacitor model by measuring capacitance and dissipation at two different frequencies and using the equations shown below.



$$C_p = \frac{f_1^2 C_1\left(1+D_1^2\right) - f_2^2 C_2\left(1+D_2^2\right)}{f_1^2 - f_2^2}$$

$$R_p = \frac{1}{\sqrt{\omega_1^2 C_1 C_p\left(1+D_1^2\right) - \omega_1^2 C_p{}^2}} = \frac{1}{\sqrt{\omega_2^2 C_2 C_p\left(1+D_2^2\right) - \omega_2^2 C_p{}^2}}$$

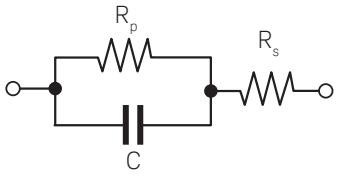
$$R_s = \frac{D_1}{\omega_1 C_1\left(1+D_1^2\right)} - \frac{R_p}{1+\omega_1^2 C_p{}^2 R_p{}^2} = \frac{D_2}{\omega_2 C_2\left(1+D_2^2\right)} - \frac{R_p}{1+\omega_2^2 C_p{}^2 R_p{}^2}$$

C_1 : Measured capacitance by parallel model at f_1
D_1 : Measured dissipation by parallel model at f_1
C_2 : Measured capacitance by parallel model at f_2
D_2 : Measured dissipation by parallel model at f_2

Figure 8.32. The two-frequency method for calculating the parameters for the three element capacitor model for thin gate oxides.

Unfortunately, the two-frequency method has been shown to be unreliable for calculating the three-element capacitor model for thin gate oxides (refer to paper by R. Clerc, et al, at INFOS 2001). Therefore, an alternate method that involves plotting the impedance and phase across frequency and then using data obtained at the point of minimum phase angle was proposed. This method is shown on the next page.

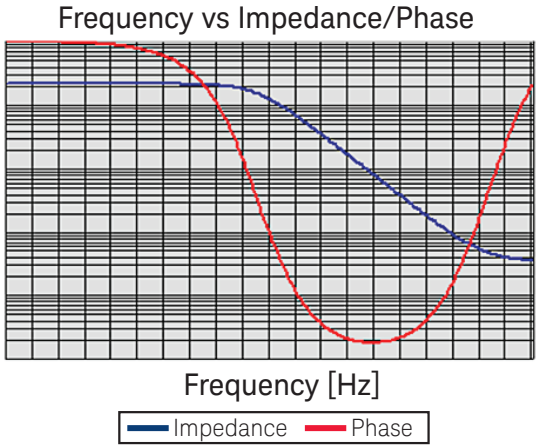
Calculating the values of the three element capacitor model *(continued)*



$$R_s = |Z_0| \frac{1 = \sin\theta_0}{\cos\theta_0}$$

$$R_p = -2|Z_0| \tan\theta_0$$

$$C = \frac{|Z_0|}{\omega_0 R_p R_s} = \frac{1}{4\pi f_0 |Z_0|}\left(\frac{\sin\theta_0 - 1}{\sin\theta_0}\right)$$



f_0: frequency at phase minimum
Z_0: Impedance at phase minimum

Figure 8.33. The minimum phase method (shown above) is the preferred method for calculating the parameters for the three-element capacitor model.

You can refer to Keysight application note 4294-3 (publication number 5988-5102EN) for more information.

An interesting fact with regards to the three-element capacitor model is that applying a higher frequency does not necessarily result in a lower value for the dissipation factor (D). The equivalent impedance for the three-element case consists of the parallel combination of R_p and C_p in series with R_s. Referring back to equations 8.11 and 8.12, we find that for the three-element model the equivalent impedance is given by:

$$Z = R_s + \frac{R_p - R_p^2 C_p j\omega}{1 + R_p^2 C_p^2 \omega^2} \qquad \text{(Equation 8.13)}$$

We can derive the real and imaginary parts of this impedance as:

$$\mathrm{Re}\{Z\} = R_s + \frac{R_p}{1 + R_p^2 C_p^2 \omega^2} \qquad \mathrm{Im}\{Z\} = \frac{-R_p^2 C_p \omega}{1 + R_p^2 C_p^2 \omega^2}$$

As discussed earlier, the dissipation factor (D) is given by the ratio of the real part to the imaginary part:

$$D = \frac{\mathrm{Re}\{Z\}}{\mathrm{Im}\{Z\}} = \frac{\dfrac{R_s\left(1 + R_p^2 C_p^2 \omega^2\right) + R_p}{1 + R_p^2 C_p^2 \omega^2}}{\dfrac{-R_p^2 C_p \omega}{1 + R_p^2 C_p^2 \omega^2}} = -\frac{R_s\left(1 + R_p^2 C_p^2 \omega^2\right) + R_p}{R_p^2 C_p \omega} \qquad \text{(Equation 8.14)}$$

Calculating the values of the three element capacitor model *(continued)*

Unlike the two-element model case, dissipation does not continuously decrease with increasing frequency but in fact has a minimum value. The following plot of dissipation versus frequency for a three-element capacitor illustrates this point.

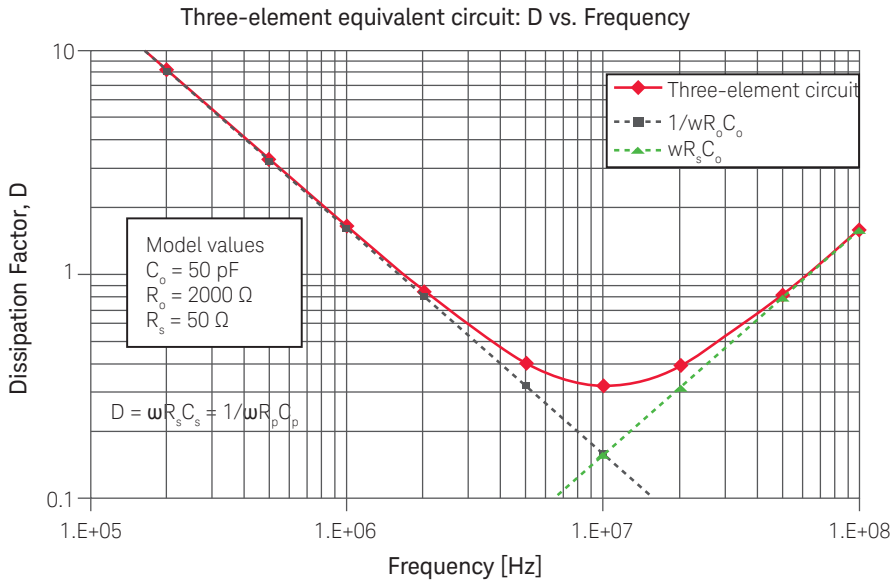


Figure 8.34. This three-element capacitor model plot shows that arbitrarily increasing the frequency does not necessarily result in a lower value for dissipation.

Therefore, in practice it may take some trial and error to determine the optimal frequency at which to perform measurements. Note: This behavior was first documented by G. Brown, at ICMTS 2005: "Capacitance Characterization in Integrated Circuit Development: The Intimate Relationship of Test Structure Design, Equivalent Circuit, and Measurement Methodology".

The advanced IV method

The 4TP cabling technique has limitations. As frequency increases, both chuck capacitance and the residual chuck inductance impact the measurement results. The best means to minimize these effects is to ground the wafer chuck, but as mentioned previously, the 4TP method requires the wafer chuck to be floating. For this reason, Keysight developed an alternative fixturing (cabling) solution for making high-frequency measurements on-wafer known as the advanced IV method. An illustration of this method is shown below.

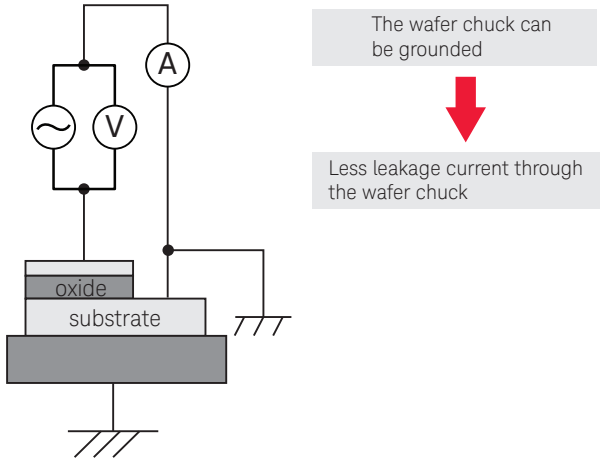


Figure 8.35. The advanced IV method allows the wafer chuck to be grounded, resulting in less leakage current through the wafer chuck due to residual capacitance and inductance.

The advanced IV method *(continued)*

The advanced IV method provides the following advantages:

1. The low side of the current meter is grounded.
2. The voltage source and the voltage meter are floating.
3. Frequency measurements up to 110 MHz can be supported.

The advanced IV method requires the Keysight 4294A or E4990A impedance analyzer with the Keysight 42941A Impedance Probe. You cannot use the 42941A probe with the 4284A or E4980A LCR meters. A picture of this instrument and the probe are shown below.

Figure 8.36. The 4294A or E4990A impedance analyzer and the 42941A impedance probe are both required to perform the advanced IV method of CV measurement.

The 4294A and E4990A have certain idiosyncratic features of which all users should be aware. One of these is a feature known as the “cable terminated auto-balancing bridge method.” Essentially, what this means is that the 4294A and E4990A insert 50 Ω termination resistors to reduce reflected waveforms when the frequency reaches 5 MHz or above whenever 1 or 2 meter cables are used with the instrument (which of course is always the case when making on-wafer measurements). The figure below illustrates this feature.

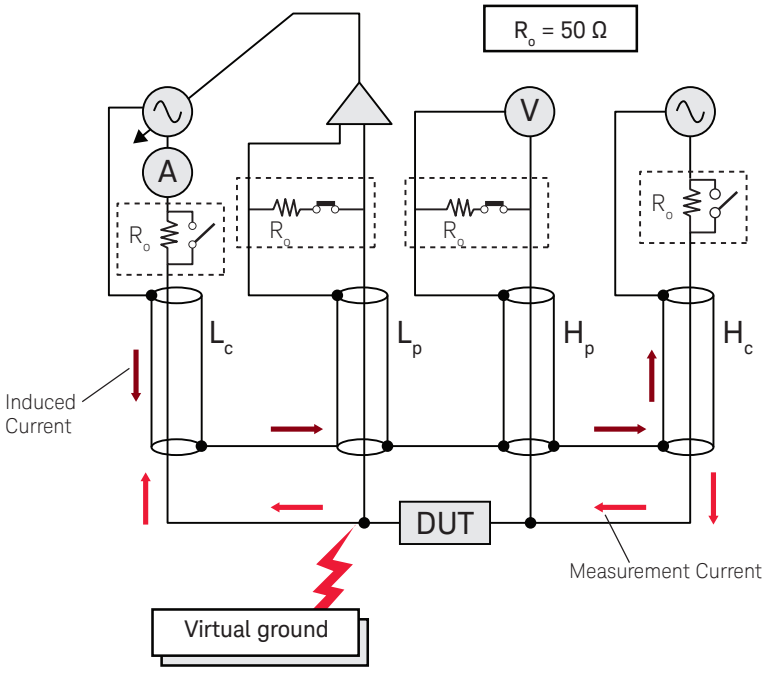


Figure 8.37. The 4294A and E4990A switch in 50 Ω resistors to reduce waveform reflections at frequencies of 5 MHz and above when 1 and 2 meter cable extensions are used.

The advanced IV method *(continued)*

Unless a load calibration is performed at 5 MHz, a discontinuity in the capacitance measurement will be observed at 5 MHz. Therefore, during the 4294A or E4990A load compensation, it is essential to make sure the 5 MHz point is included in the frequency compensation list.

Extremely thin gate oxides usually require both the advanced IV method and careful test structure layout in order to obtain satisfactory measurement results. The following graph shows a conventional test structure design with a ground-signal-ground (GSG) pad layout measured using the advanced IV method.

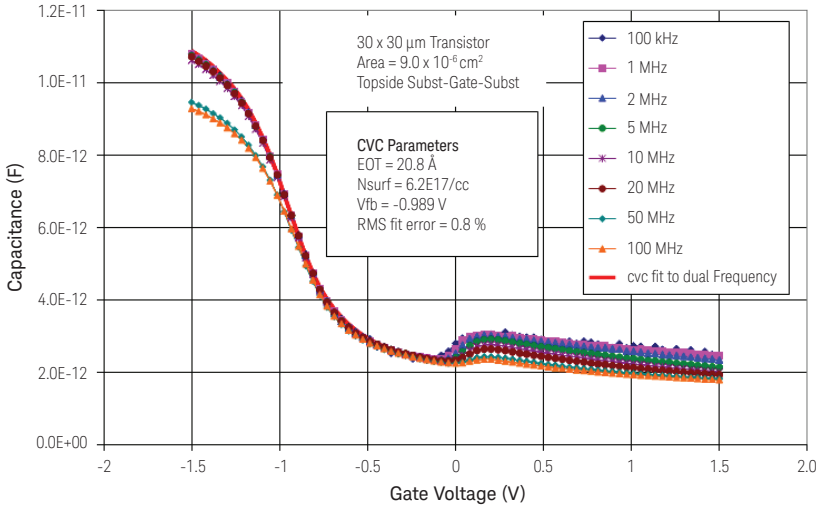


Figure 8.38. The measurement results obtained across frequency using the advanced IV method on a conventional test structure design (GSG pad layout). [Note: Data courtesy of SEMATECH]

The advanced IV method *(continued)*

As this plot shows, there are measurement variations across frequency and the device does not exhibit QSCV behavior in the inversion region. In contrast, the following graph shows a test structure with a ground-signal-ground (GSG) pad layout measured using the advanced IV method that follows the design guidelines previously discussed to minimize the series resistance (R_s).

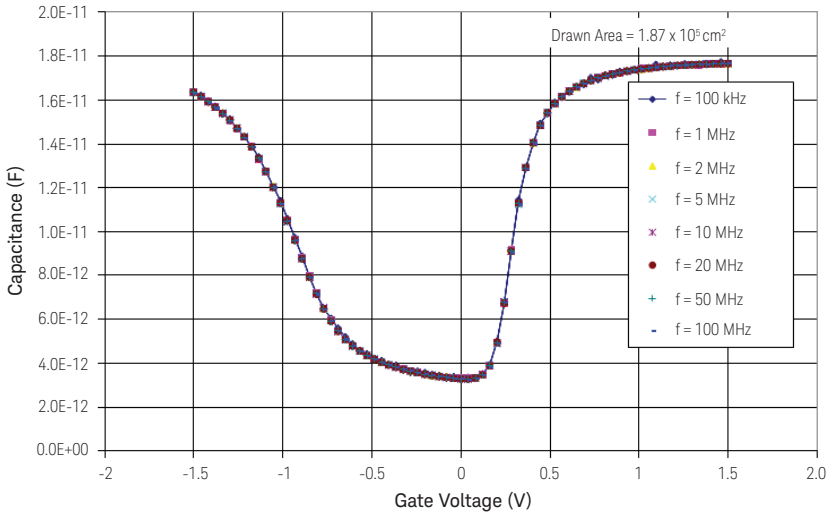


Figure 8.39. The measurement results obtained across frequency using the advanced IV method on a test structure designed to minimize the series resistance (GSG pad layout). [Note: Data courtesy of SEMATECH]

This measurement shows excellent correlation across measurement frequency as well as a QSCV response in the inversion region, and the electrical oxide thickness turns out to be 2 nanometers! Again, this result reinforces the importance of proper structure design when making capacitance measurements on extremely thin gate oxides.

Controlling the 4294A or E4990A using EasyEXPERT

The front panel user interface of the 4294A is complex with many sub-menus. In addition, the 4294A has a couple of features that sometimes cause users to make inadvertent errors. The first of these is, unlike the B1500A/B1505A/B1506A and E4980A, the 4294A does not retain cabling compensation information during power-down. One solution is to always perform Open/Short/Load compensation each time the 4294A is powered up, although this can be a bit tedious and impractical. Fortunately, the 4294A does have a built-in Flash memory that permits compensation data to be saved into a file and retrieved after power-up; however, the user must remember to do this. The second issue with the 4294A is that many of its default settings are not appropriate for semiconductor capacitance measurement. Again, it is a little tedious to change all the default settings each time the instrument is powered-up and it is also easy to miss something or make a mistake.

To help with these issues, Keysight has made available EasyEXPERT application tests that can control the 4294A. One of these application tests simply guides the user through the 4294A open, short and (optionally) load calibration procedures. After the calibration procedures are completed the application test stores the calibration data into files within the 4294A's Flash memory. The other application tests permit the user to perform capacitance versus DC voltage (CV) sweeps and capacitance versus frequency (C-f) sweeps. Note that both the CV and C-f application tests first check for the presence of the stored calibration files created by the calibration application test. If these files are not present, then the application test issues a message informing the user that they must first perform cable compensation and the application test stops execution. Of course, if the calibration information is present, then the CV and C-f application tests automatically load the calibration data into the instrument. In addition, both the CV and C-f application tests take care of changing the 4294A default settings into values appropriate for semiconductor capacitance measurement. Therefore, these application tests make it almost impossible to capture invalid data due to calibration or setup issues. Finally, these application tests make it easy for the user to store and export the test results into PC-based formats, which is not possible from the front panel of the 4294A.

A screen capture of the EasyEXPERT 4294A CV application test is shown below.

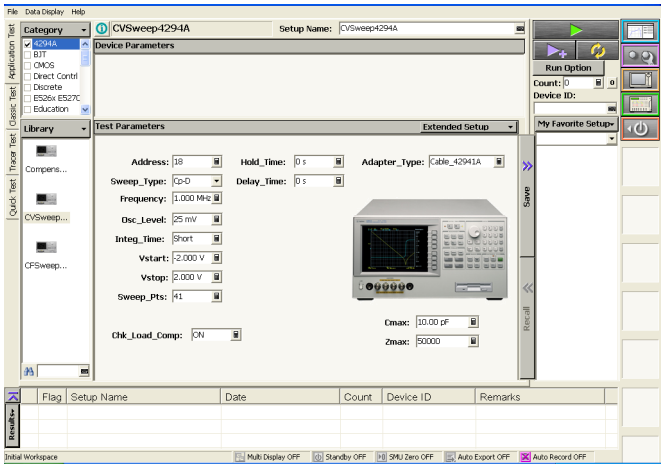


Figure 8.40. The EasyEXPERT application test to perform capacitance versus DC voltage sweeps using the 4294A.

Controlling the 4294A or E4990A using EasyEXPERT *(continued)*

This application test provides control of all of the important CV sweep parameters for the 4294A. If you cannot perform or do not want to perform Load compensation, then you can change the "Chk_Load_Comp" input to "OFF", and the application test will not check the stored calibration file to determine if the Load compensation was performed during calibration. Note: The "Integ_Time" parameter is actually a simplification of the 4294A's settings, as the 4294A does not actually support integration time in the same sense as the B1500A. However, for the vast majority of users, this feature is an improvement over the innate settings of the 4294A.

A screen capture of the EasyEXPERT C-f application test is shown below.

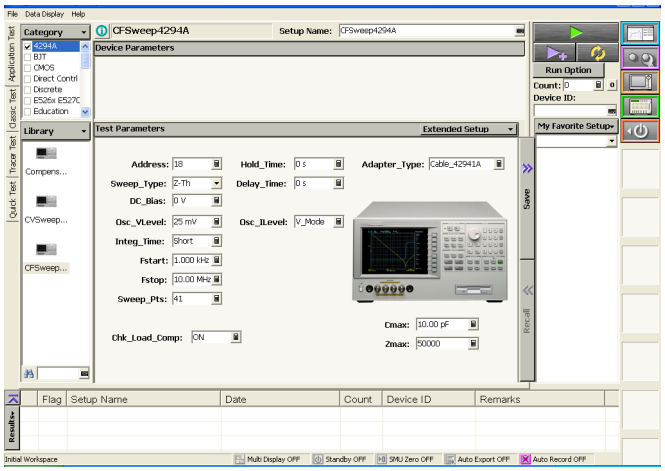


Figure 8.41. The EasyEXPERT application test to perform capacitance versus frequency sweeps using the 4294A.

The key features of this application test are similar to those of the CV sweep application test. One difference is that this application test permits the user to set the AC oscillator into either voltage or current mode. A numeric value of "0" should be entered into the oscillator level input that is not being used (a symbol name such as "V_Mode" or "I_Mode" will automatically appear). Another difference is, although the capacitance parameters are plotted on a linear scale, the frequency values are plotted on a logarithmic scale.

Note: At the time of this writing, EasyEXPERT application tests to control the E4990A were not available. However, they may be available at a future date. Please check the Keysight EasyEXPERT website for updates.

Summary of thin gate dielectric best practices

In summary, the following are the key factors to keep in mind when attempting to measure capacitance on thin gate dielectric transistors:

1. Perform an Open/Short/Load calibration (after first performing Phase compensation if it is supported) before making any measurements.
2. Use a frequency high enough to minimize dissipation and maintain accuracy.
3. Use the three-element device model to extract the device parameters.
4. Account appropriately for the effects of the wafer chuck through careful layout of the test structure (minimize R_s).
5. If the 4TP cabling method does not yield satisfactory results, consider using the advanced IV method.

Measurement Method	Measurement Probe Contact Point	Instrumentation	Comments
4TP	Wafer or Chuck (Chuck preferred)	E4980A, B1500A, B1505A, B1506A, 4294A, E4990A	– Complex setup – Max freq. ~30 MHz
Advanced IV	Wafer	4294A or E990A + 42941A	– Easy setup – Max freq. to 110 MHz – Requires RF probes

Figure 8.42. Table summarizing the 4TP and Advanced IV measurement methods.

Making capacitance measurements through a switching matrix

A long-running issue in parametric testing is "the CV-IV measurement dilemma." The dilemma referred to has to do with the different cabling needs of capacitance meters and SMUs. Capacitance meters require the use of BNC cables and SMUs require the use of triaxial cables, yet most parametric testing requires both CV and IV measurements. Manually switching from triaxial to BNC cables (and back again) is extremely tedious, and it also does not permit any sort of test automation. Obviously, switching matrices (discussed in chapter 4) represent one solution to this problem. However, by introducing a switching matrix into the CV measurement path to take care of the BNC to triaxial conversion, we solve one issue and create several new ones. In this section, we will discuss the proper way to make capacitance measurement through a switching matrix.

Switching Matrix Capacitance Measurement Issues

When we discussed making on-wafer measurements in chapter 4, we covered Keysight's positioner-based CV-IV switching solutions. As discussed in that chapter, Keysight's positioner-based solutions have both built-in capacitance compensation (to the outputs of the ASU and SCUU) and the means to short the guards together during capacitance measurement to stabilize the series inductance of the cables. However, it should be stressed that performing these same operations on a switching matrix is much more complex. Positioner-based solutions have the advantage of a fixed cable length; switching matrices must contend with signal path lengths and loadings that change each time an input-to-output relay connection is changed.

The figure shown below outlines the three sections that must be considered when making capacitance measurements through a switching matrix.

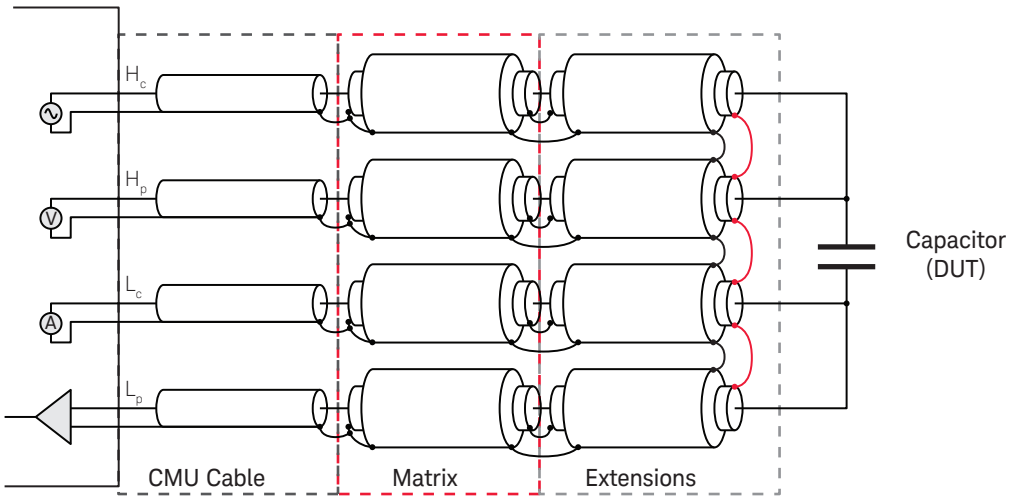


Figure 8.43. When measuring capacitance through a switching matrix, there are three factors to consider: the CMU cable, the matrix, and the extensions.

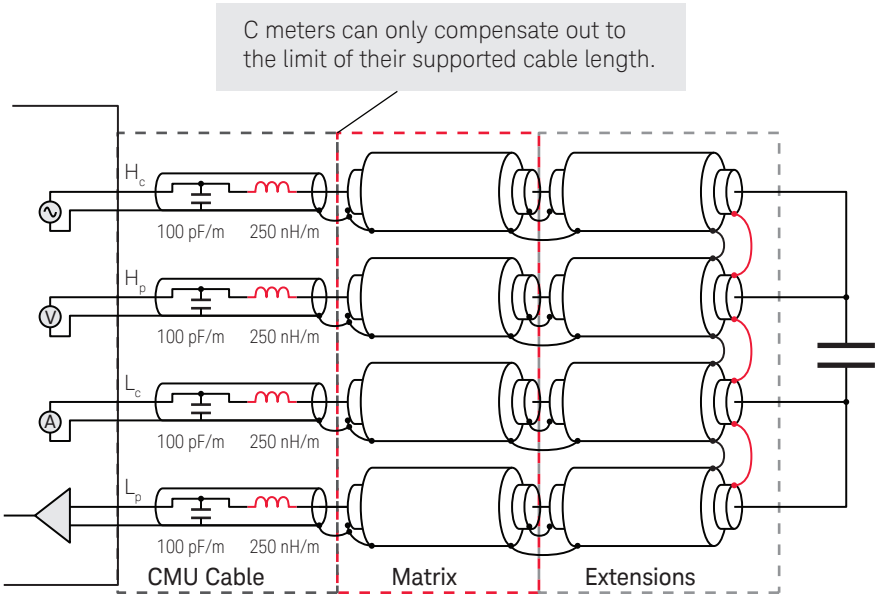
Most of the errors that occur when making capacitance measurements through a switching matrix can be attributed to one or more of the following factors:

1. Non-standard CMU cables or unsupported cable lengths are used to connect the capacitance meter to the switching matrix.
2. The guards are not shorted together close to the DUT.
3. The cable extensions create cable impedance mismatch.
4. The matrix does not support any sort of compensation scheme that takes into account changes in loading and path length due to the opening and closing of input-to-output connections.

In the following discussion, we will examine the best means to mitigate these factors.

CMU cable compensation

Although we have already discussed the issue of capacitance meter cable compensation quite extensively, it is worthwhile to repeat some of the important points again since many times these seem to be forgotten when a switching matrix is involved. One of the key points is the length of the capacitance meter cables. Proper cable length is critical to balance the bridge of the capacitance meter's measurement circuitry. If an improper cable length is used, then the capacitance meter may not be able to balance its bridge and you will get an error during the compensation process.



Figures 8.44. Using an unsupported CMU cable or an unsupported cable length are common reasons for poor measurement results when making capacitance measurements through a switching matrix.

The figure shown below describes the correct procedure to compensate the cabling going from the capacitance meter to the switching matrix.

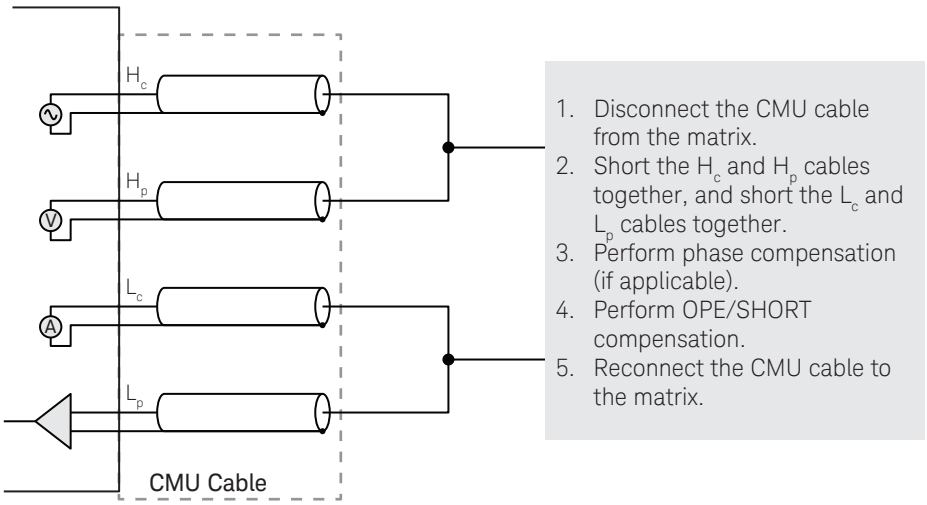


Figure 8.45. The proper procedure to compensate the cable going from the capacitance meter to the switching matrix.

CMU cable compensation *(continued)*

It is also crucial to use Keysight-supplied cables to connect to the matrix. Not only do Keysight cables come at the correct cable lengths for proper compensation, they are manufactured and tested to known quality standards and have known characteristic impedance. If you try to use off-the-shelf BNC cables, then Keysight cannot guarantee the accuracy of your measurements even if you perform compensation and everything seems to pass without any issues. The Keysight E4980A supports cable lengths of 1 m, 2 m, and 4 m. The Keysight B1500A MFCMU supports cable lengths of 1.5 m and 3 m. You should try to use as short a cable as possible to connect to your matrix.

Shorting the guards

Although the reasons for shorting the guards together during capacitance measurement were discussed in chapter 4, we will briefly review them again. The outer shield of the BNC cables coming from a capacitance meter are not at ground potential, but are "virtual grounds". The reason for shorting the driven guards together during capacitance measurement is to stabilize the series inductance of the measurement path. If the guards are not shorted, then the series inductance can change from 250 nH/m to more than 400 nH/m. Such variation in the series inductance of the measurement path makes it virtually impossible to obtain stable capacitance compensation.

1. Make the return path as close to the DUT as possible.
2. Make the return path wire length as short as possible.

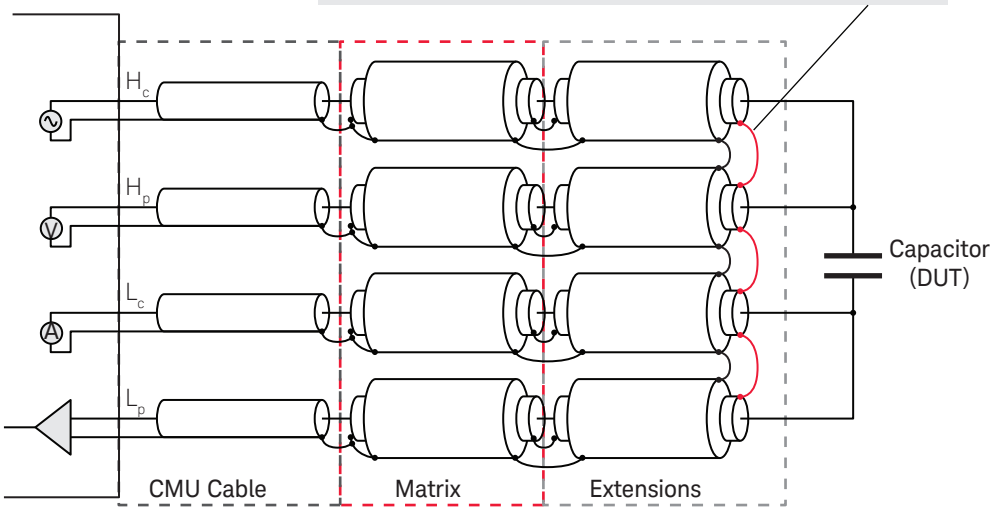


Figure 8.46. Important considerations when shorting the guards together during capacitance measurement.

Impedance mismatch

Capacitance meters require a 50 Ω cable environment when they perform Open/Short/Load compensation. Unfortunately, neither the switching matrix nor the cable extensions have characteristic impedances of 50 Ω. This affects the accuracy of the compensation.

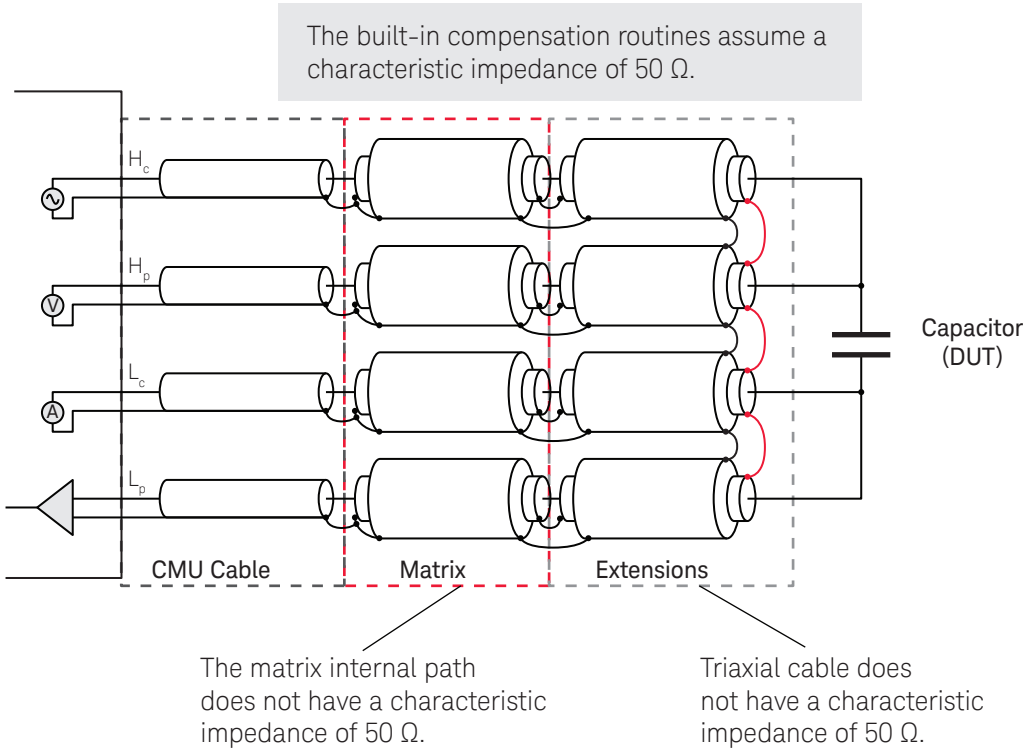


Figure 8.47. One major challenge is that the matrix path and cable extensions do not have a characteristic impedance of 50 Ω.

Ideally, the Load impedance should be in the same range as the device impedance that you want to measure. However, this is never the case for semiconductor devices. Unfortunately, the typical input impedance of a semiconductor device is very much higher than 50 Ω. This affects the accuracy of the Load compensation.

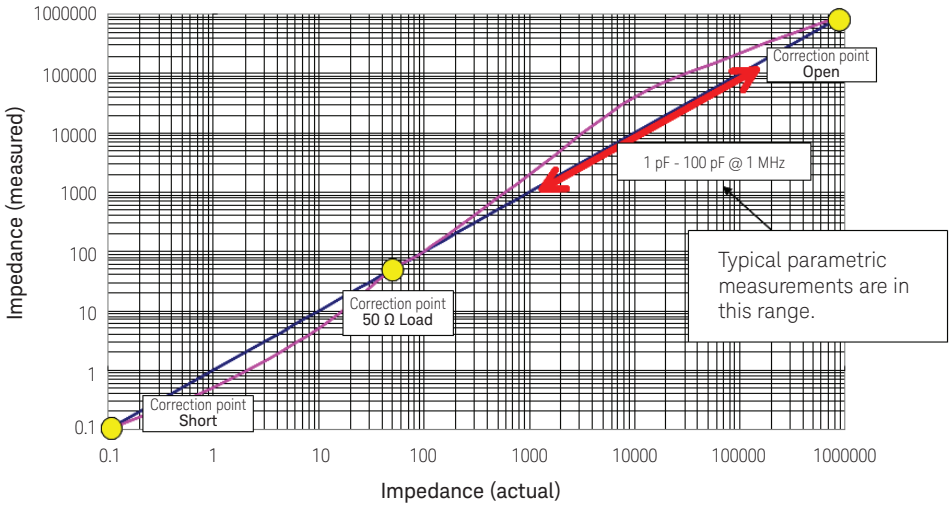


Figure 8.48. This graph illustrates why semiconductor load impedance in not ideal for making capacitance measurements.

Keysight's switching matrix solution

Keysight can supply a complete CV-IV solution that compensates for the cabling and fixturing all the way down to the probe card tips. Besides improving CV measurement, this solution also yields excellent low-current measurement performance.

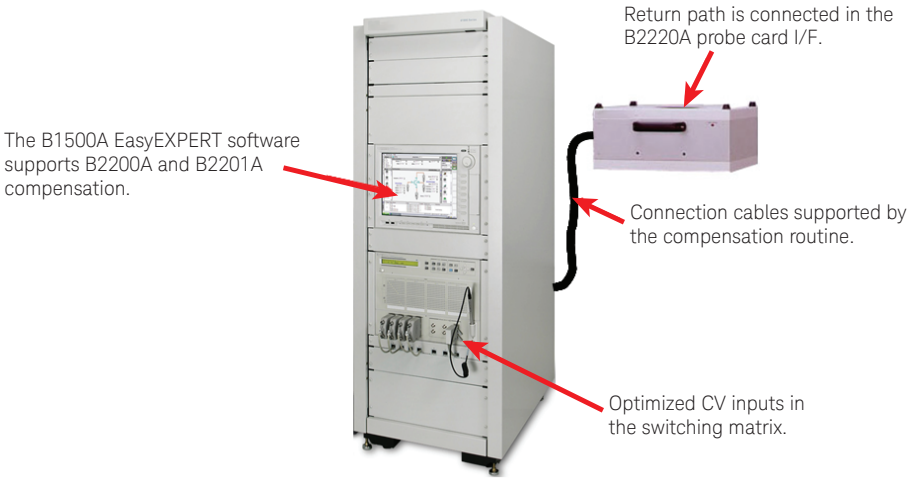


Figure 8.49. Keysight's integrated switching matrix CV-IV capacitance measurement solution.

The details of this solution are shown below.

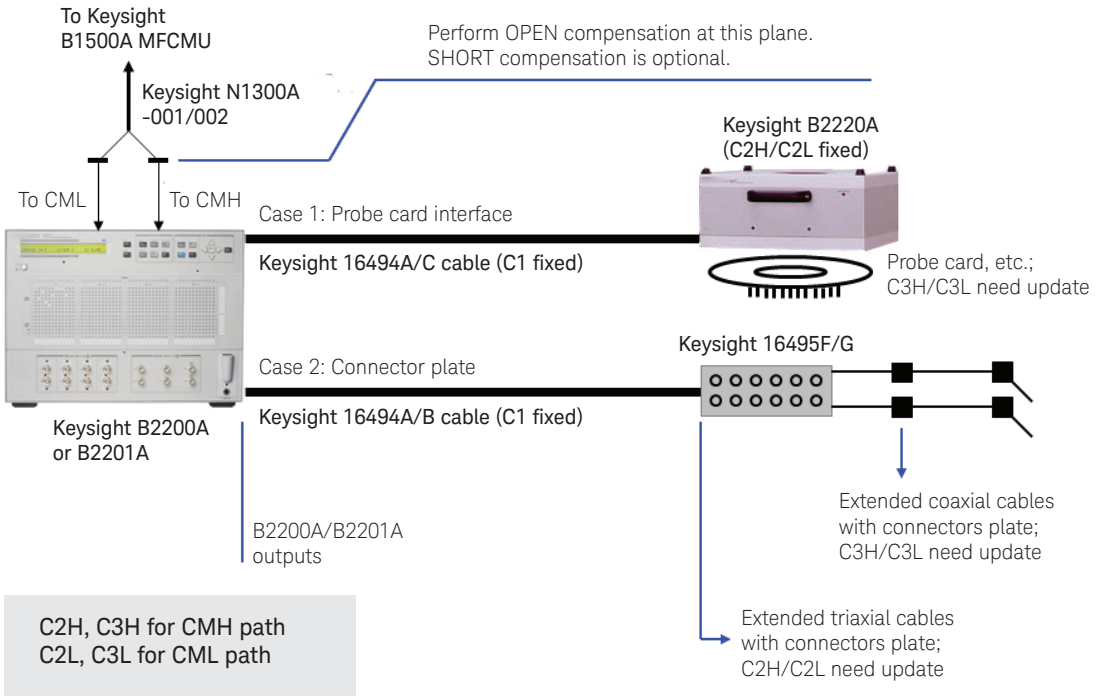


Figure 8.50. The EasyEXPERT B2200A/B2201A matrix compensation utility has built-in compensation data for both Case 1 and Case 2 shown above.

Keysight's switching matrix solution *(continued)*

The following information is provided as a reference for the figure shown above.

B2220A-024/048: Probe card interface (24 pins/48 pins)
N1300A-001/002: CMU cable for B1500A (1.5 m/3.0 m)
16494A-002/005: Triaxial cable (3.0 m/4.0 m)
16494B-001/002: Kelvin triaxial cable for 16495F/G (1.5 m/3.0 m)
16494C-002/005: Kelvin triaxial cable for B2220A (3.0 m/4.0 m)
16495F: Connector plate with 12 triaxial, interlock, and GNDU connectors
16495G: Connector plate with 24 triaxial, interlock, and GNDU connectors

EasyEXPERT also allows you to use connection options other than the B2220A Probe Card Interface or 16495F/G connector plate. However, you then have to calibrate your fixturing and create your own Compensation Data file. This process is outlined in the EasyEXPERT help menu.

The above solution works because the matrix and extension path is well-defined with known correction coefficients. EasyEXPERT will support this automated compensation up to frequencies of 1 MHz.

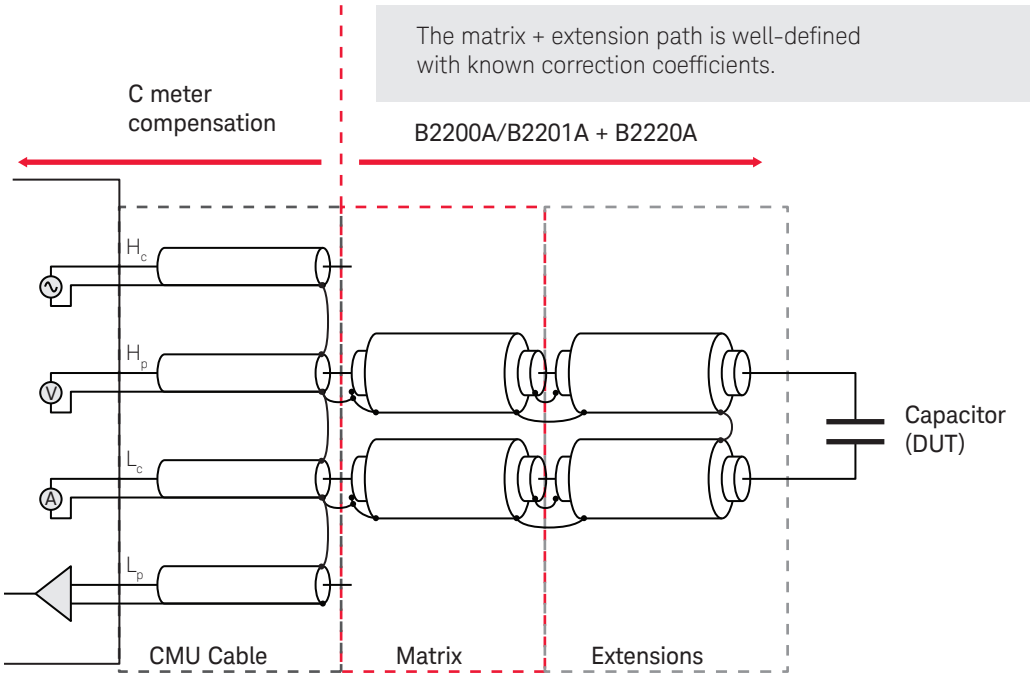


Figure 8.51. Schematic overview of Keysight's switching matrix compensation scheme.

Tips for measuring at frequencies above 1 MHz

Before proceeding, it is important to emphasize that Keysight does not offer any guarantees that you can make capacitance measurements through any of our switching matrix products at frequencies above 1 MHz. Measurements above this frequency are not supported. However, if you need to make measurements above 1 MHz, then we have a few tips for you to maximize your chances for success. If you are trying to use a switching matrix above 1 MHz, then here are the key points to keep in mind:

1. Your C meter has a maximum supported cable length. If the total length of your cabling (connection cable + matrix path + extensions) exceeds the length supported by your C meter, then your Open/Short/Load compensation will fail (you will get a "bridge unbalanced" message).
2. Every time you change your matrix settings, your path length will change. You need to perform Open/Short/Load compensation each time you make changes to the switch settings.
3. You must supply a current return path near the DUT.

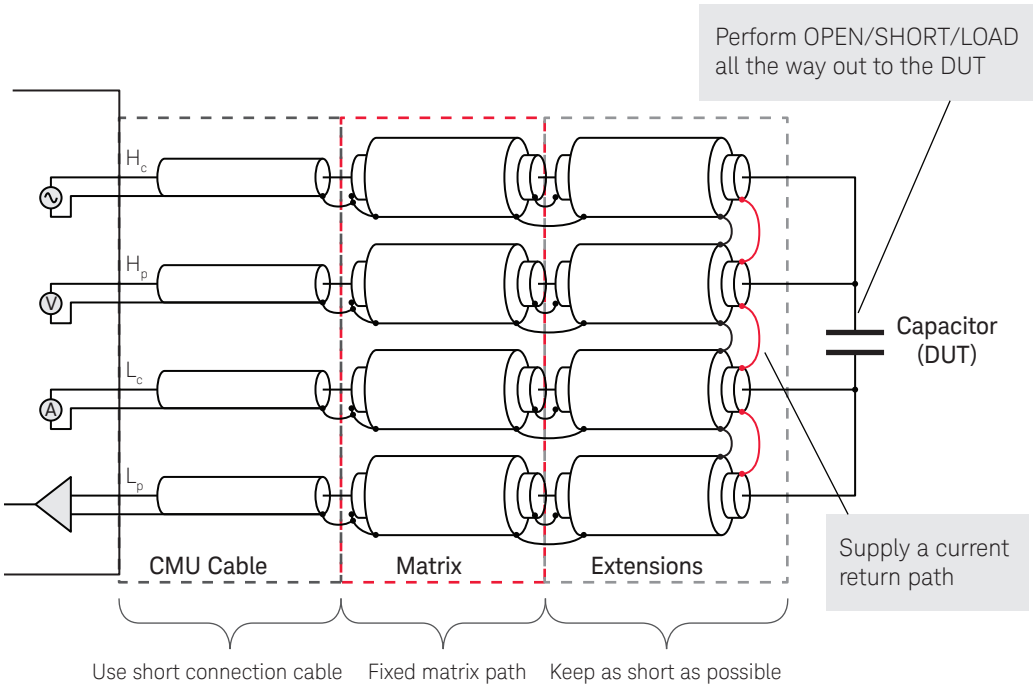


Figure 8.52. Follow this procedure if trying to go above 1 MHz when using a switching matrix.

Even though you can have success making capacitance measurements above 1 MHz through a switching matrix, it needs to be emphasized that 5 MHz represents a hard upper limit on the achievable frequency. Above 5 MHz you should use direct connections in order to obtain satisfactory capacitance measurement results.

Chapter 9 Power Device Characterization

"I learned very early the difference between knowing the name of something and knowing something."
— Richard Feynman

Introduction

Power semiconductor devices have unique measurement needs that warrant devoting a separate chapter to their characterization. Unlike small-signal devices, power diode and transistor characterization is primarily focused on power loss. Since power loss must be computed dynamically, calculating it requires characterization of device capacitances under actual operating conditions (i.e. thousands of volts of bias and/or hundreds of amps of current) as well as the measurement of other AC related parameters.

Calculating Power Loss in Power Semiconductor Devices

Semiconductor power loss consists of three components: conduction loss, driving loss and switching loss. Conduction loss is always present, and it is the loss due to the innate on-resistance (R_{on}) of a power MOSFET (or the V_{ce} saturation voltage of a bipolar device). While conduction loss itself does not depend on frequency, the average conduction power loss does vary with the duty cycle. Driving loss is probably less familiar to those not acquainted with power transistors, but it arises from the need to add or remove charge from a transistor gate in order to turn it on or off. In addition to gate charge, driving loss also depends on the applied gate voltage and the switching frequency. Switching loss is caused by device capacitance, which slows device switching transitions and causes the upper and lower totem pole transistors to both be in their linear regions simultaneously. Besides device capacitance, switching loss depends on the device gate resistance as well as (of course) the switching frequency. The following figure summarizes these three factors.

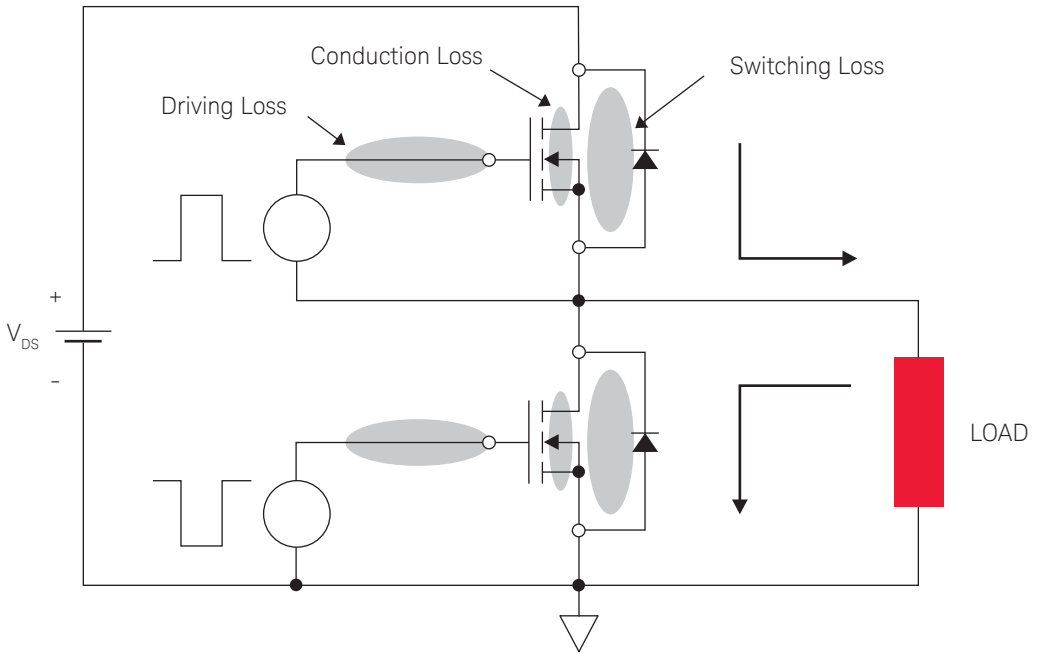


Figure 9.1. The components of power loss.

Conduction loss

As mentioned previously, conduction loss is the Ohmic (IR) power consumed while a transistor is in the on state. The following figure illustrates this concept.

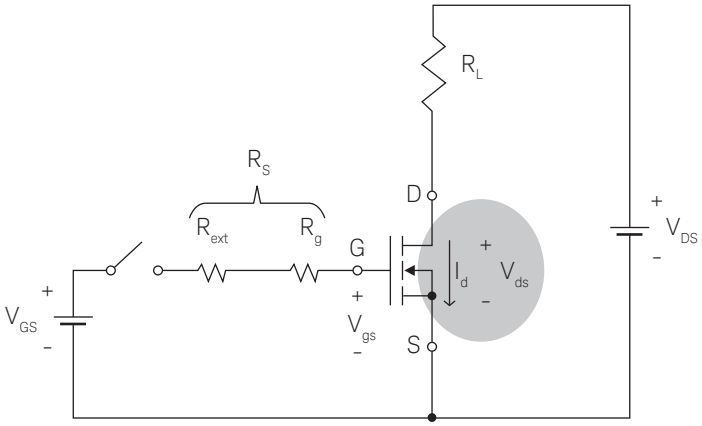


Figure 9.2. Conduction loss due to on-resistance of transistor.

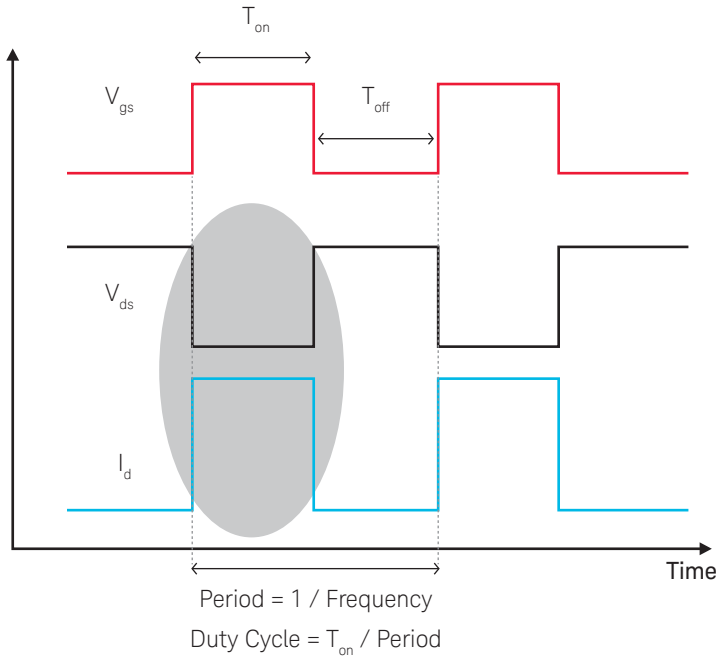


Figure 9.3. Conduction loss waveforms.

The calculation of conduction power loss is straightforward:

For MOSFETS:

(Equation 9.1)

$$P_{cond} = I_d \times V_{ds} \times \text{Duty Ratio}$$
$$= I_d^2 \times R_{ds}(on) \times \text{Duty Ratio}$$

For IGBTs & BJTs:

(Equation 9.2)

$$P_{cond} = I_c \times V_{ce}(sat) \times \text{Duty Ratio}$$

Driving loss

Driving loss is caused by current flowing through the gate resistance when the gate capacitance charges/discharges during device turn on/off. The total gate resistance consists of the sum of any externally added resistance and the device's internal gate resistance (both from the package and intrinsic to the DUT) as shown in the following figure.

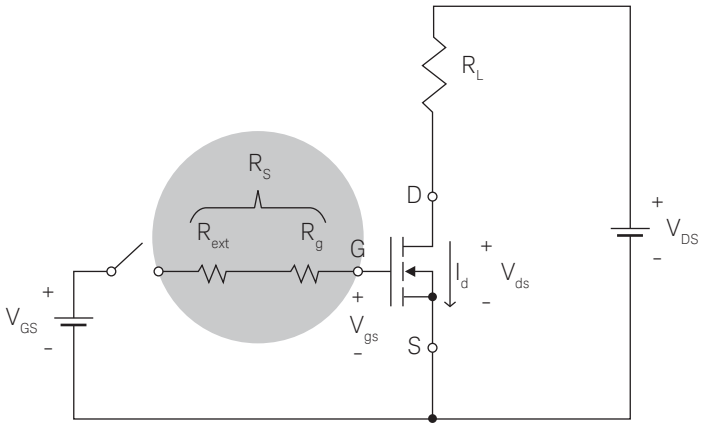


Figure 9.4. Driving power loss due to gate resistance.

As the following waveform diagram shows, driving loss occurs when the transistor is turning on as well as when it is turning off.

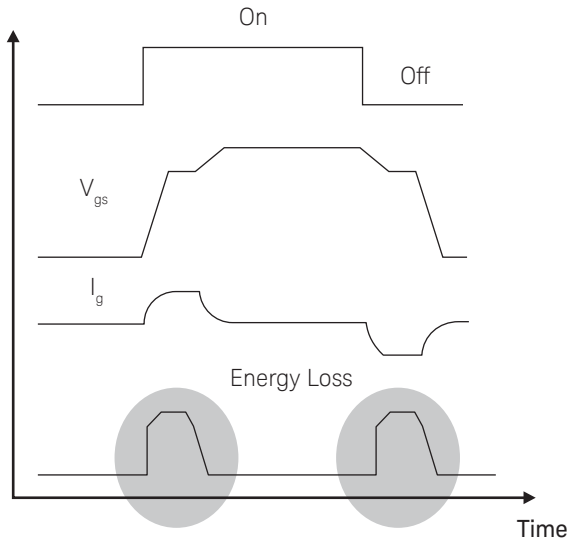


Figure 9.5. Driving loss occurs both during device turn on and turn off.

The calculation of driving loss is straightforward. We know that the energy stored in a capacitor is given by the following equation.

(Equation 9.3)

$$E = \frac{1}{2} CV^2 = \frac{1}{2} QV$$

Using this equation, we can then calculate the power loss when the device is turning on as a function of the charge necessary to turn the device on, the gate-to-source voltage (V_{gs}) and the switching frequency (f).

$$P_d(on) = \frac{1}{2} Q_g(on) \times V_{gs} \times f$$

(Equation 9.4)

We also know that (at least to first order) the charge necessary to turn the device on is equal to the charge necessary to turn the device off.

$$Q_g(on) = Q_g(off)$$ (Equation 9.5)

Thus, the power dissipated while turning the device on is the same as that generated while turning the device off. We can then write the total power dissipated due to driving loss as follows.

$$P_d(total) = P_d = Q_g \times V_{gs} \times f$$ (Equation 9.6)

It is obvious that in order to calculate driving loss you need to know the gate charge, which is the subject of the next section.

Gate charge overview

Gate charge is the charge that you have to place onto or remove from the gate (or base) of a transistor in order to turn it on or off (respectively). Gate charge is plotted as a curve with charge (Q_g) on the x-axis and the gate-to-source voltage (V_{gs}) on the y-axis. A typical gate charge curve for a silicon MOSFET is shown below.

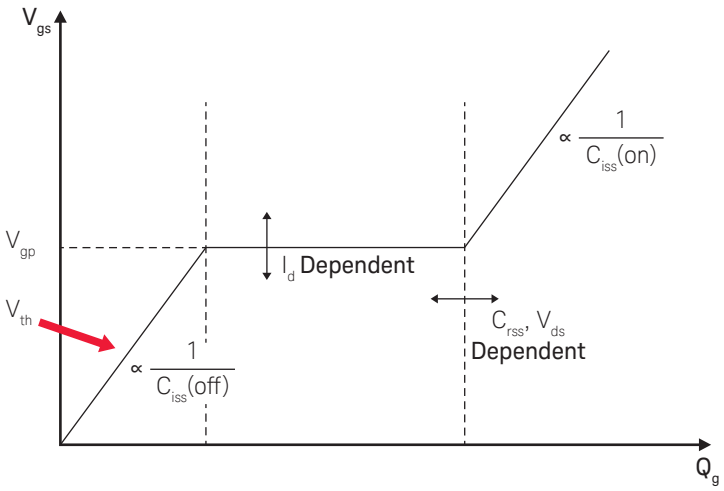


Figure 9.6. A typical silicon MOSFET gate charge curve.

In the initial portion of the curve the input capacitance (C_{iss}) is being charged and the device is still turned off. When sufficient charge is on the gate to start turning the device on then V_{gs} stops increasing temporarily. This portion of the curve is known as the plateau voltage, and it has a value designated as V_{gp}. The V_{gs} voltage will remain at V_{gp} until the device is completely turned on. Once the device is fully turned on, charge can still be placed on the transistor by continuing to charge up the input capacitance (C_{iss}). However, it should be noted that C_{iss}, which is the sum of the gate-to-drain and gate-to-source capacitances, is voltage dependent. This means that the slope of the curve for the regions before and after the plateau voltage region, while proportional to $1/C_{iss}$ in both cases, is not the same.

Gate charge equations by region

To calculate gate charge for each region, it is instructive to also plot the drain-to-source voltage (V_{ds}) on a second y-axis.

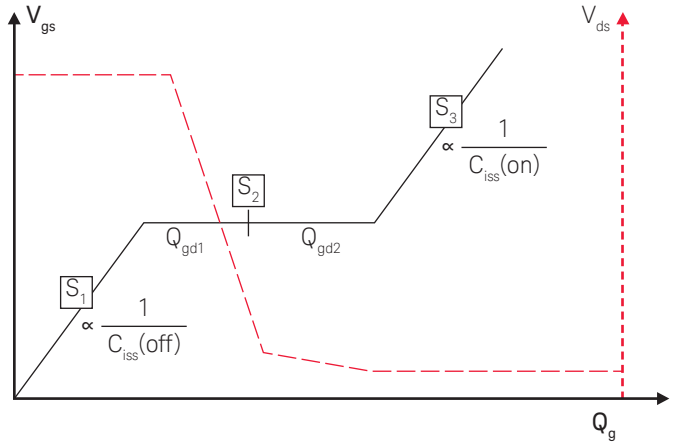


Figure 9.7. Gate charge equations for each region.

In segment one (S1) we have the following equation for V_{gs}.

$$\mathbf{S1}: V_{gs} = \frac{Q_g}{C_{gs} + C_{gd}} \approx \frac{Q_g}{C_{gs}} \qquad (\text{since } C_{gs} >> C_{gd}) \qquad \text{(Equation 9.7)}$$

Segment two (S2) has to be divided up into two regions.

$$\mathbf{S2}: Q_{gd1} = \int_0^{V_{ds}-V_{gs}} C_{gd} \cdot dV \qquad (V_{ds} > V_{gs}) \qquad \text{(Equation 9.8)}$$

$$Q_{gd2} = \int_{V_{ds}}^{V_{gs}} (C_{gs} + C_{gd}) \cdot dV \qquad (V_{gs} > V_{ds}) \qquad \text{(Equation 9.9)}$$

$$Q_{gd} = Q_{gd1} + Q_{gd2} \qquad \text{(Equation 9.10)}$$

Segment three (S3) has the same relationship between V_{gs} and gate charge (Q_g), as does segment one (S1).

$$\mathbf{S3}: V_{gs} = \frac{Q_g}{C_{gs} + C_{gd}} \qquad \text{(Equation 9.11)}$$

Gate charge switching time equations

Now that we know how to calculate the gate charge for different regions of the gate charge curve, we can proceed to calculate the switching time equations. In the following equations, we will refer to the total series resistance, R_s, which consists of the device gate resistance (R_g) and any external resistance that has been added into the circuit. If we apply an external voltage V_{gs} to the gate through the series resistance, then the time dependent voltage on the transistor gate will be given by the following equation.

$$V_{gs}(t) = V_{GS}\left\{1 - e^{-\frac{t}{C_{iss} * R_s}}\right\} \quad \text{(Equation 9.12)}$$

We can easily solve this equation for time.

$$t = (C_{iss} * R_s) * \ln\left\{\frac{V_{GS}}{V_{GS} - V_{gs}}\right\} \quad \text{(Equation 9.13)}$$

Equation 9.7 relates Q_g, V_{gs} and C_{iss}, so we can re-write the above equation as follows.

$$t = \left(\frac{Q_g}{V_{gs}} * R_s\right) * \ln\left\{\frac{V_{GS}}{V_{GS} - V_{gs}}\right\} \quad \text{(Equation 9.14)}$$

The reader should be able to convince themselves that we can then calculate the time difference between two points as a function of gate charge and gate voltage.

$$t_2 - t_1 = \left(\frac{Q_{g2} - Q_{g1}}{V_{g2} - V_{g1}}\right) * R_s * \ln\left\{\frac{V_{GS} - V_{g1}}{V_{GS} - V_{g2}}\right\} \quad \text{(Equation 9.15)}$$

The following graph is very important, as it shows both the gate-to-source (V_{gs}) and drain-to-source voltages (V_{ds}) as a function of gate charge (Q_g) for both transistor turn on and turn off. It allows us to use equation 9.15 to give a quantitative expression for all of the important switching time parameters.

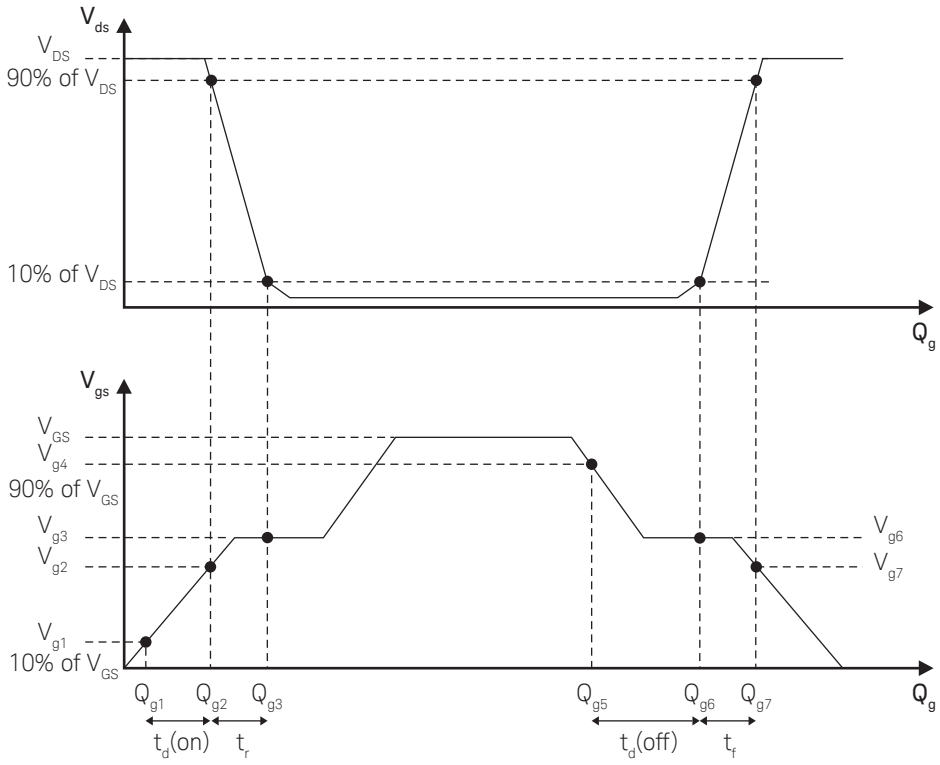


Figure 9.8. The relationships between gate voltage, gate charge and turn-on/turn-off times.

Using this graph as a reference, we can now use equation 9.15 to specify the switching time equations as shown below.

Turn On Delay Time (10% of V_{GS} to 90% of V_{DS})	$t_d(on) = \left(\frac{Q_{g2} - Q_{g1}}{V_{g2} - V_{g1}}\right) * R_S * \ln\left\{\frac{V_{GS} - V_{g1}}{V_{GS} - V_{g2}}\right\}$
Rise Time (90% of V_{DS} to 10% of V_{DS})	$t_r = \left(\frac{Q_{g3} - Q_{g2}}{V_{g3} - V_{g2}}\right) * R_S * \ln\left\{\frac{V_{GS} - V_{g2}}{V_{GS} - V_{g3}}\right\}$
Turn Off Delay Time (90% of V_{GS} to 10% of V_{DS})	$t_d(off) = \left(\frac{Q_{g6} - Q_{g5}}{V_{g6} - V_{g5}}\right) * R_S * \ln\left\{\frac{V_{g6}}{V_{g5}}\right\}$
Fall Time (10% of V_D to 90% of V_{DS})	$t_f = \left(\frac{Q_{g7} - Q_{g6}}{V_{g7} - V_{g6}}\right) * R_S * \ln\left\{\frac{V_{g7}}{V_{g6}}\right\}$

Figure 9.9. Summarized switching time equations.

Switching loss

To understand switching loss, it is useful to first consider “ideal” switching waveforms where the transition time is instantaneous as shown below.

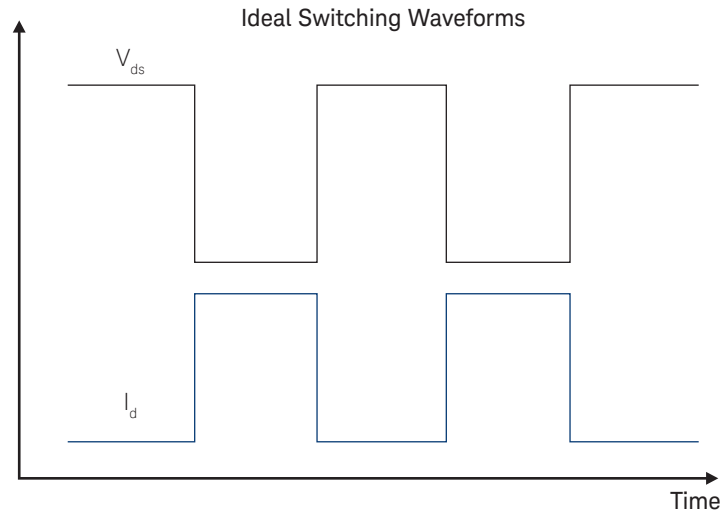


Figure 9.10. Ideal switching waveforms that do not generate any switching loss.

Since there is no period of overlap where the upper and lower totem pole transistors are both in their linear regions, the switching loss is zero. However, in any real-world situation, we have to consider the finite time required for a transistor to turn on or off, and these transition periods generate power loss as shown below.

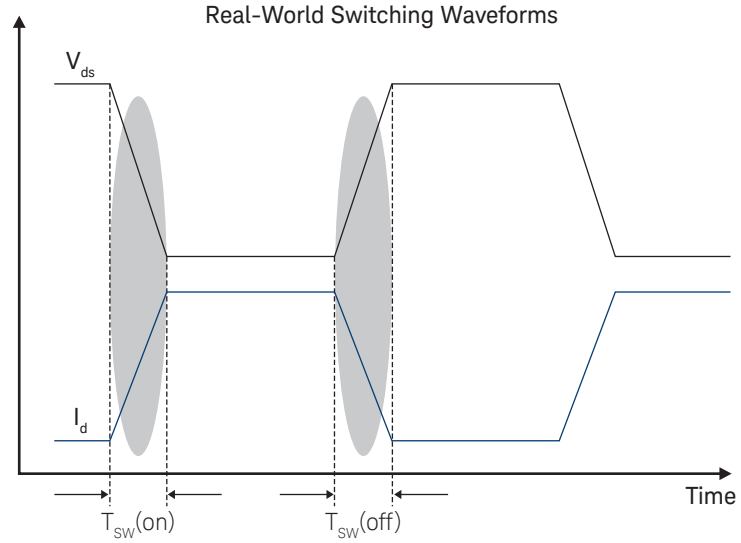


Figure 9.11. Actual switching waveforms generate power loss during their turn on and turn off transitions.

To quantitatively calculate the switching loss, you need to know what type of a load you are driving. We will look at both the resistive load and inductive load cases.

Resistive load switching loss

Consider the following switching test circuit with a resistive load.

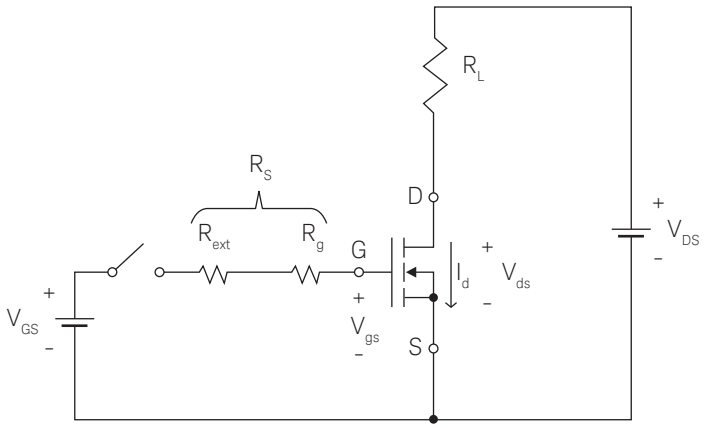


Figure 9.12. Switching test circuit with resistive load.

As the transistor turns on we expect to see the following switching waveforms.

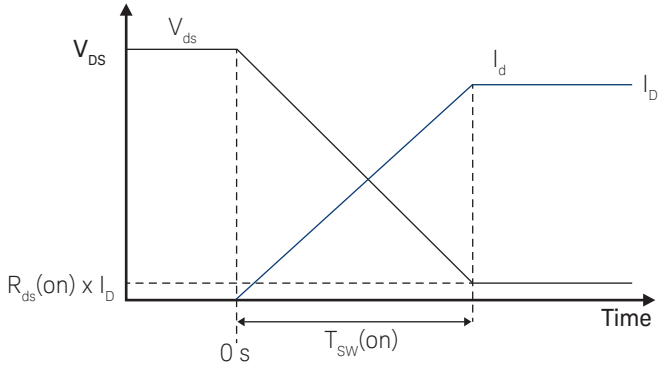


Figure 9.13. Switching waveforms for a circuit with a resistive load.

At any point in time between when the transistor starts to turn on until it is fully on, the drain to source voltage across the transistor can be described by the following equation:

$$V_{ds} = V_{DS} - \left(\frac{V_{DS} - R_{ds}(on) \times I_D}{T_{sw}(on)} \right) t \qquad \text{(Equation 9.16)}$$

Here $T_{sw}(on)$ is the time it takes the transistor to fully turn on. Since the drain current also linearly increases with increasing time we also have that:

$$I_d = \frac{I_D}{T_{sw}(on)} t \qquad \text{(Equation 9.17)}$$

To calculate the total energy loss across the transistor as it transitions from the off to on state we need to multiply these two equations together (voltage x current) and integrate them over the switching time:

(Equation 9.18)

$$E_{SW}(on) = \int_0^{T_{SW}(on)} I_d \times V_{ds}\, dt$$

$$= \int_0^{T_{SW}(on)} \left(\frac{I_D \times V_{DD}}{T_{SW}(on)} t + \frac{I_D \times V_{DS}}{\left[T_{SW}(on)\right]^2} t^2 + \frac{R_{ds}(on) \times I_D^2}{\left[T_{SW}(on)\right]^2} t^2 \right) dt$$

$$= \left[\frac{I_D \times V_{DS}}{2T_{SW}(on)} t^2 \right]_0^{T_{SW}(on)} - \left[\frac{I_D \times V_{DS}}{3\left[T_{SW}(on)\right]^2} t^3 \right]_0^{T_{SW}(on)} + \left[\frac{R_{ds}(on) \times I_D^2}{3\left[T_{SW}(on)\right]^2} t^3 \right]_0^{T_{SW}(on)}$$

$$= \frac{1}{6} I_D \times V_{DS} \times T_{SW}(on) + \frac{1}{3} R_{ds}(on) \times I_D^2 \times T_{SW}(on)$$

In most cases $R_{ds}(on)$ is small enough so that we can make the following approximation.

(Equation 9.19)

$$E_{sw}(on) \approx \frac{1}{6} I_D \times V_{DS} \times T_{sw}(on)$$

We can use similar reasoning to derive the following equation for energy loss across the transistor as it transitions from on to off:

(Equation 9.20)

$$E_{sw}(off) \approx \frac{1}{6} I_D \times V_{DS} \times T_{sw}(off)$$

We can now write an equation for the total resistive power loss:

(Equation 9.21)

$$P_{sw} = \left[E_{sw}(on) + E_{sw}(off)\right] \times f$$

$$= \frac{1}{6} I_D \times V_{DS} \times \left[T_{sw}(on) + T_{sw}(off)\right] \times f$$

Inductive load switching loss

Consider the following switching test circuit with an inductive load.

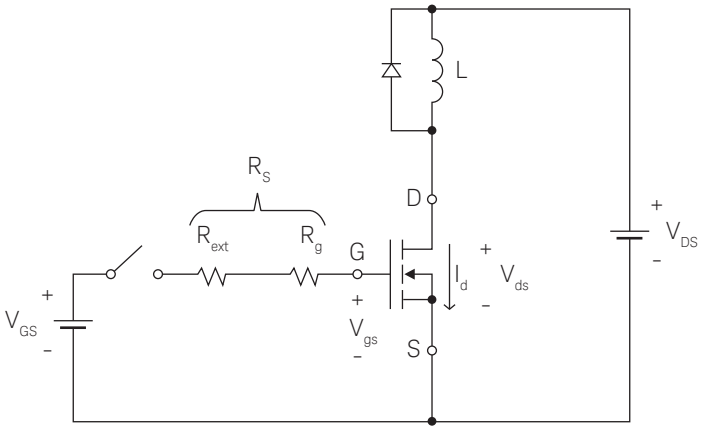


Figure 9.14. Switching test circuit with inductive load.

Note that in the inductive load case, the use of a clamping diode is necessary to protect the transistor from inductive "kick-back" as the transistor transitions from on to off. As the transistor turns on, we can relate the I_d and V_{ds} waveforms to the gate charge curve as shown in the following figure.

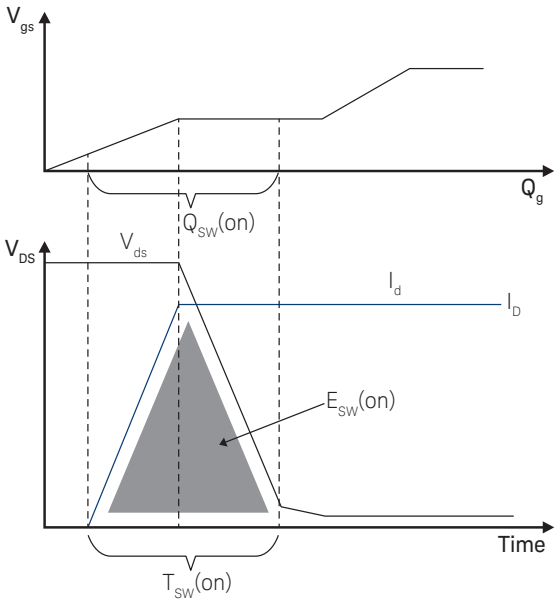


Figure 9.15. I_d and V_{ds} waveforms for an inductive load.

To calculate the energy loss during device turn on, we use a triangular approximation. We use the turn on time as the base and the final value of the drain current as the height. We then multiply this by the value of applied Vds to get the total energy lost during transistor turn on.

$$E_{sw}(on) \approx \frac{1}{2} I_D \times V_{DS} \times T_{sw}(on) \qquad \text{(Equation 9.22)}$$

As in the case of a resistive load, it is safe to assume that energy loss in switching the transistor off can be calculated using the same equation by substituting the off switching time, Tsw(off). We can therefore write the total resistive power loss as shown in the following equation.

$$P_{sw} = \left[E_{sw}(on) + E_{sw}(off)\right] \times f = \frac{1}{2} I_D \times V_{DS} \times \left[T_{sw}(on) + T_{sw}(off)\right] \times f \quad \text{(Equation 9.23)}$$

Summarized switching loss equations

The following table provides a convenient summary of the switching loss equations for the resistive and inductive load cases as well as the equations for transistor on and off switching times.

P_{SW} (resistive)	$\left(\frac{1}{6}\right) * V_{DS} * I_D * (T_{SW}(on) + T_{SW}(off)) * f$
P_{SW} (inductive)	$\left(\frac{1}{2}\right) * V_{DS} * I_D * (T_{SW}(on) + T_{SW}(off)) * f$
T_{SW} (on)	$\frac{Q_{sw}}{i_g} = R_s * \frac{Q_{sw}}{V_{GS} - V_{gp}}$
T_{SW} (off)	$\frac{Q_{sw}}{i_g} = R_s * \frac{Q_{sw}}{V_{gp}}$

Figure 9.16. Summarized switching loss equations.

As these equations show, there is considerably more power loss for an inductive load versus a resistive load. This is more easily understood when you compare the waveforms for the two cases.

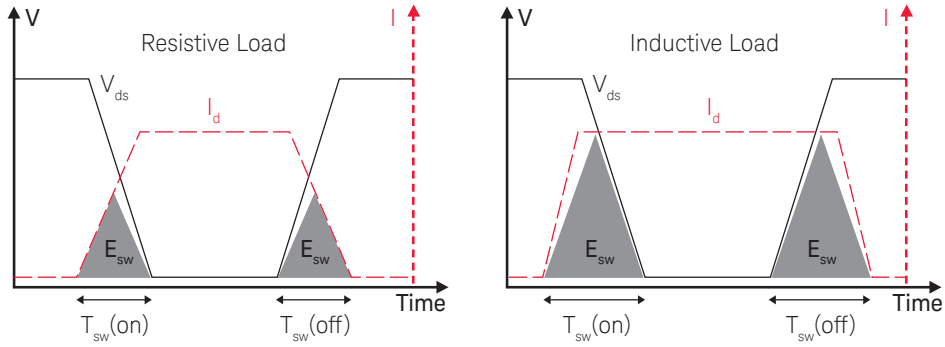


Figure 9.17. Resistive load and inductive load switching waveforms compared.

As this figure shows, the area under the current and voltage curves is more “stretched out” for the inductive load case, which results in more energy being dissipated.

Although we have derived the necessary equations to calculate switching loss, it should be clear that they are only of use if we know the device capacitance and gate charge parameters. Therefore, the next two sections will go into detail as to how to measure both of these parameters.

High DC bias capacitance measurements

Power device capacitance measurement requires us to build on the principles discussed in chapter 8. The most obvious new challenge faced when characterizing power device capacitance is that you need to measure the capacitances using a small AC signal (typically millivolts) while simultaneously applying thousands of volts of DC bias to the DUT. Standalone capacitance meters can typically only supply up to around 20 V of DC bias, and thus they cannot be used to characterize capacitance in power devices. The key to making these measurements is the high-voltage Bias-T.

High-Voltage Bias-T

The purpose of a Bias-T is to combine the small AC voltage signal from a capacitance meter or module with a large DC voltage (normally supplied by an SMU). Keysight's high-voltage Bias-T is designed to work with the MFCMU and HVSMU modules of the B1505A and B1506A. A simplified circuit schematic of the high-voltage Bias-T is shown below.

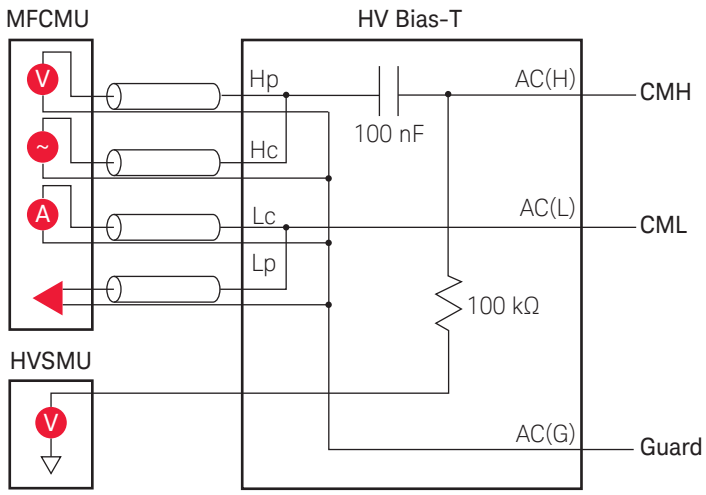


Figure 9.18. The high-voltage Bias-T takes the outputs of the MFCMU and HVSMU and converts them into the familiar CMH and CML 4TP outputs.

Note that the AC guard of the MFCMU is available, and in certain measurement situations, it is necessary to connect the AC guard to one of the DUT terminals. However, to understand the use of the AC guard, it is first necessary to explain the basics of high-power MOSFET capacitance measurement.

High-power MOSFET capacitance measurement

A conceptual diagram of a power n-channel MOSFET device is shown below.

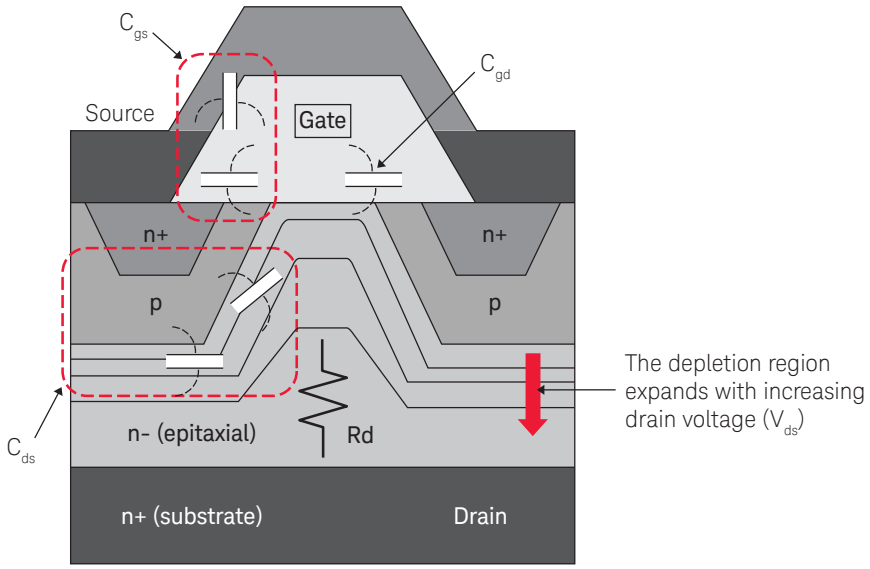


Figure 9.19. Power n-channel MOSFET diagram showing depletion regions and capacitance.

For power MOSFETs, the drain is biased to a very high voltage, and both the drain-to-source capacitance (C_{ds}) and the gate-to-source capacitance (C_{gs}) are dependent on the DC value of the drain voltage. The AC model of a MOSFET is shown below.

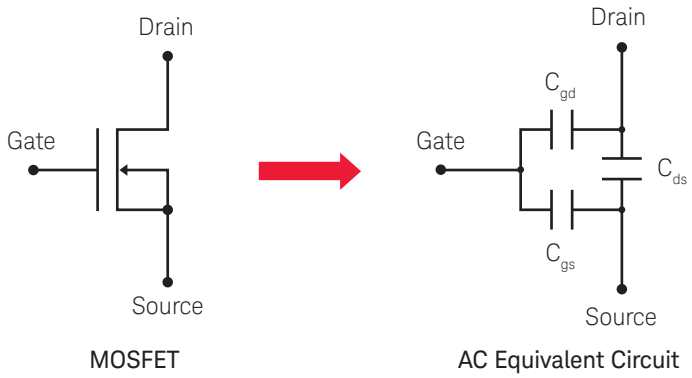


Figure 9.20. AC model of a power MOSFET.

Let us now consider what happens when we try to measure any single one of these three capacitances using a capacitance meter. The following figure shows the situation when we try to measure an unknown capacitance (C_x) on a three-terminal device without using the AC guard (i.e. unused terminal is floating).

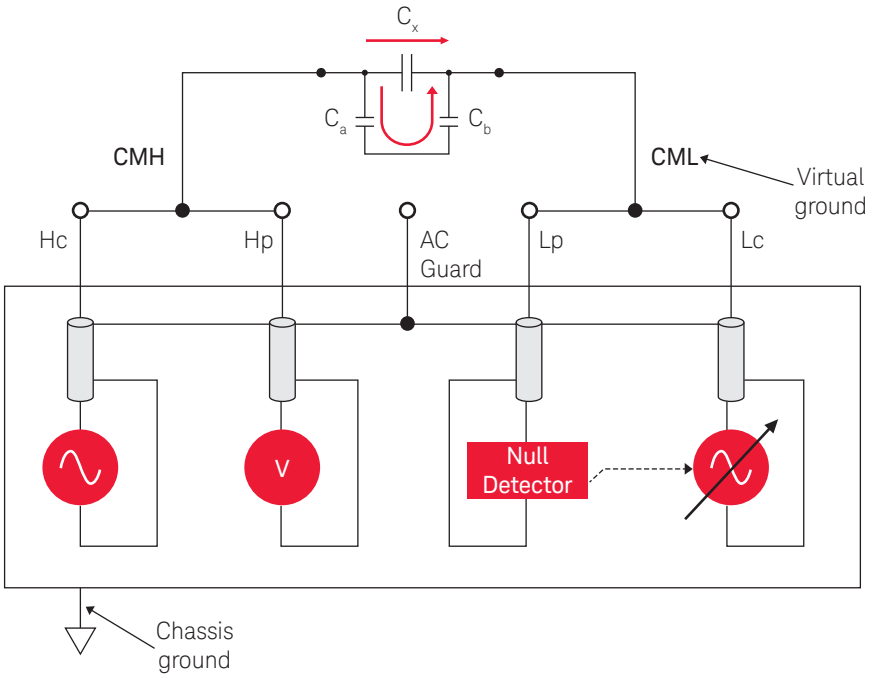


Figure 9.21. Measuring capacitance on a three-terminal device without using the AC guard.

This figure shows that when the unused terminal is floating, current can flow through the other two capacitors resulting in erroneous measurement results. The best way to prevent this from occurring is to provide an alternative current path so that the current flowing through C_a does not flow back through C_b into the CML node. We can achieve this by connecting the unused terminal to the AC guard of the capacitance meter. Note: Before proceeding, it is important to understand that the AC guard is the circuit common of the auto-balancing bridge and that it is connected to the shields of the four-terminal pair connectors. The AC guard is not the same as the ground terminal, which is connected to the chassis ground.

The next figure shows the benefit of connecting the AC guard to the unused measurement terminal.

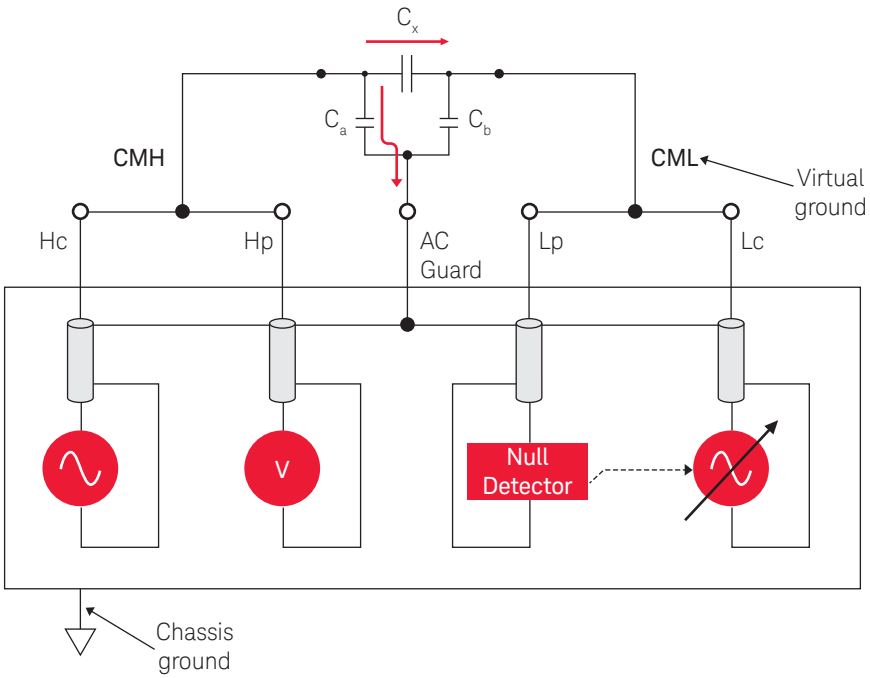


Figure 9.22. Measuring capacitance on a three-terminal device using the AC guard.

When the AC guard is connected to the third terminal, the current flowing through the parasitic path (C_a) does not affect the accuracy of the measurement of the unknown capacitance (C_x), since the capacitance measurement is done through the CML node. Of course, this scheme assumes that the impedance of the AC guard node is much less than that of the parasitic path (C_b). Although this discussion did not include the use of the HVSMU and HV Bias-T, we will explain how to incorporate them as we examine each capacitance measurement individually.

The following figure shows the correct way to measure C_{gd} using the HV Bias-T.

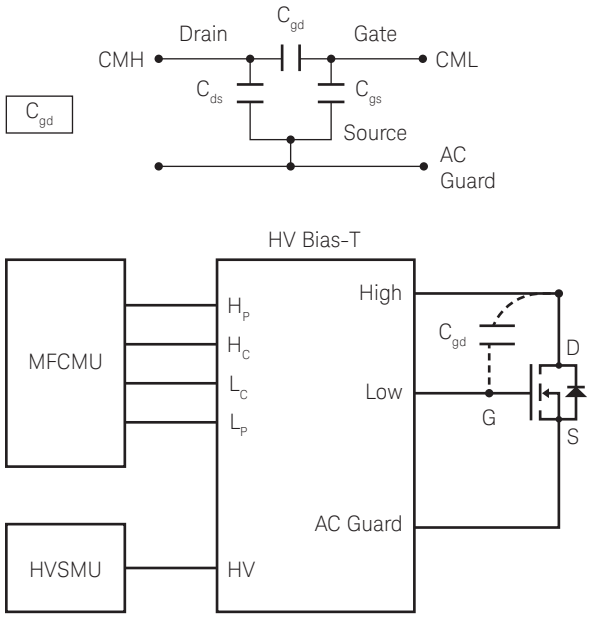


Figure 9.23. Measuring the gate-to-drain (C_{gd}) capacitance of a MOSFET using the high-voltage Bias-T and AC guard.

As can be seen, the measurement of C_{gd} using the HV Bias-T is straightforward. Similarly, the measurement of C_{ds} is also easy using the HV Bias-T as shown in the following figure.

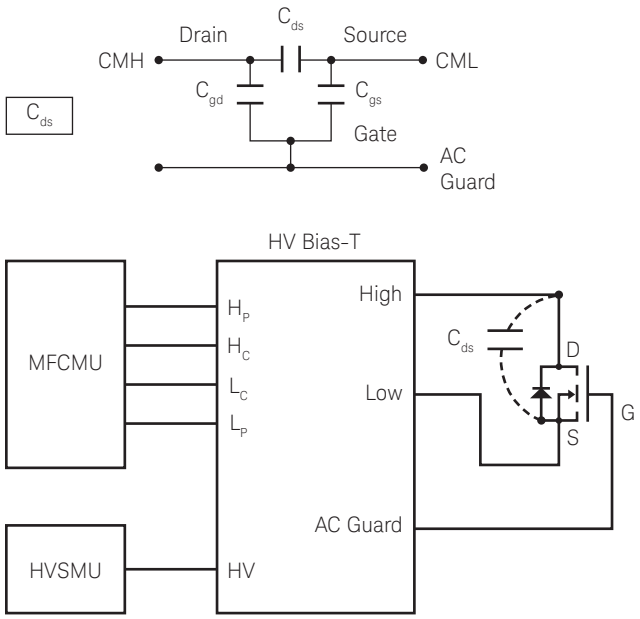


Figure 9.24. Measuring the drain-to-source (C_{ds}) capacitance of a MOSFET using the high-voltage Bias-T and AC guard.

Measuring the gate-to-source capacitance (C_{gs}) presents some challenges that are not present for the other two measurements. Although we need to short the AC guard to the drain, we also need to bias the drain to high-voltage. The solution for this is to connect the drain and AC guard via a large capacitor (much larger than C_{gd} or C_{ds}) such that the impedance seen by the drain with respect to the AC guard is much smaller than the impedance that it sees to either the source or to the gate. Conversely, we need to connect the HVSMU up to the drain through a relatively large resistor to prevent the HVSMU from interfering with the AC signal coming from the MFCMU. The following figure illustrates the proper way to make this measurement.

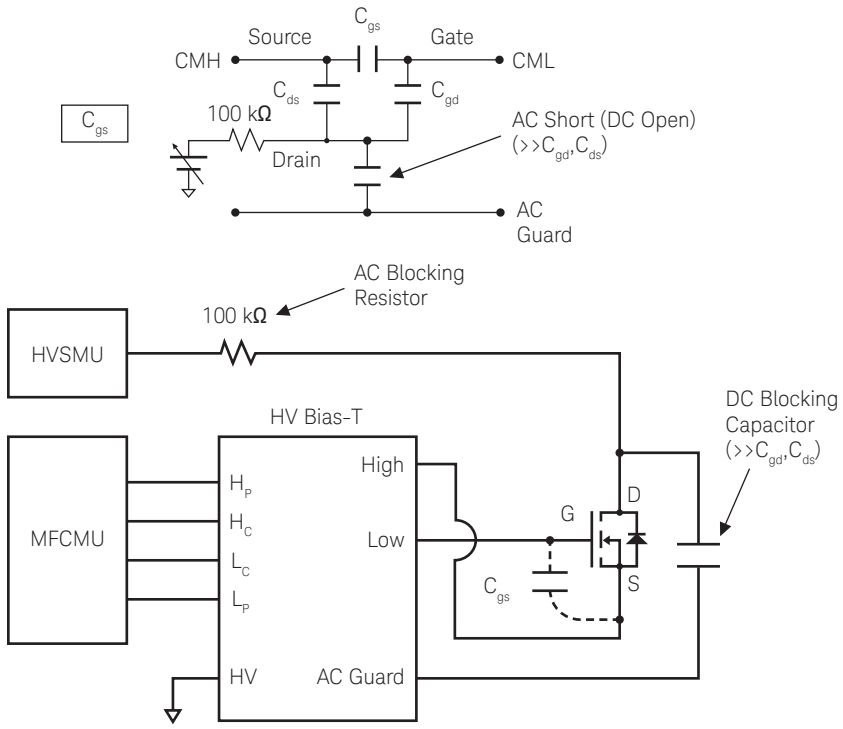


Figure 9.25. Measuring the gate-to-source (C_{gs}) capacitance of a MOSFET using the high-voltage Bias-T and AC guard.

Note: Although in this case, the high-voltage Bias-T is not actually required (and in fact is being by-passed), the connections are shown assuming that it is being used. This is simply due to the fact that the C_{gd} and C_{gs} measurements require the high-voltage Bias-T, so it is assumed that the user already has it in place.

Power MOSFET data sheets specify MOSFET capacitance differently. The typical parameters specified are output capacitance (C_{oss}), input capacitance (C_{iss}), and reverse transfer capacitance (C_{rss}). These parameters can be calculated from C_{ds}, C_{gs}, and C_{gd} using the equations shown below.

$$C_{oss} = C_{gd} + C_{ds} \quad \text{(Equation 9.24)}$$

$$C_{iss} = C_{gs} + C_{gd} \quad \text{(Equation 9.25)}$$

$$C_{rss} = C_{gd} \quad \text{(Equation 9.26)}$$

We have already discussed how to measure C_{rss} (C_{gd}) previously. C_{iss} and C_{oss} can also be measured directly on the B1505A and B1506A using the HVSMU, MFCMU and high-voltage Bias-T; however, the connections are different than for the cases of C_{gd}, C_{ds} and C_{gs}. Since C_{oss} is a simpler measurement to make, we will discuss it first. To measure C_{oss} we simply need to short the gate and source terminals using a wire as shown in the following figure.

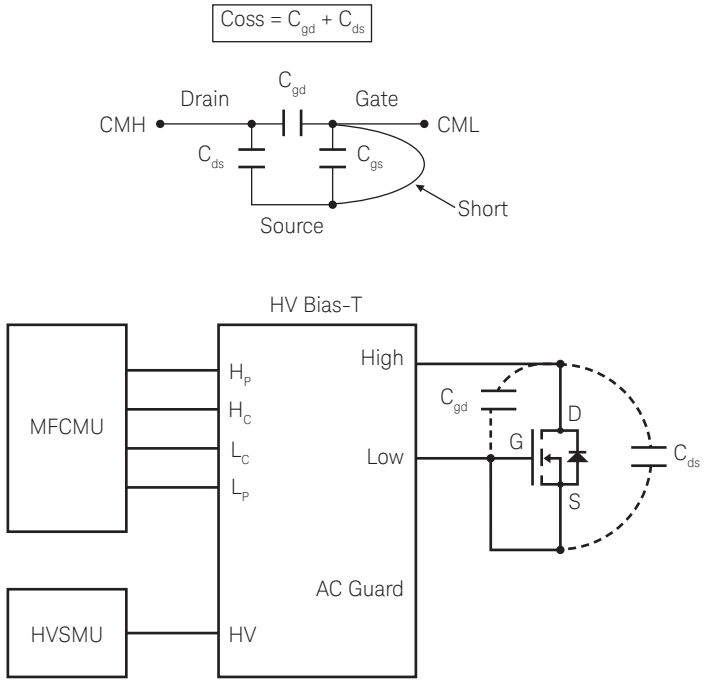


Figure 9.26. Measuring C_{oss} using the B1505A's HVSMU, MFCMU and high-voltage Bias-T.

Note that when making this measurement, we do not need to use the AC guard since we want to measure the current flowing through both C_{gd} and C_{ds}.

The measurement of C_{iss} is more challenging because we need to connect the CMH terminal to the drain and also bias the drain to high-voltage. The solution is similar to the case of C_{gs} measurement in that we need to create an AC connection to the CMH terminal of the MFCMU and a DC connection to the HVSMU. The following figure illustrates this technique.

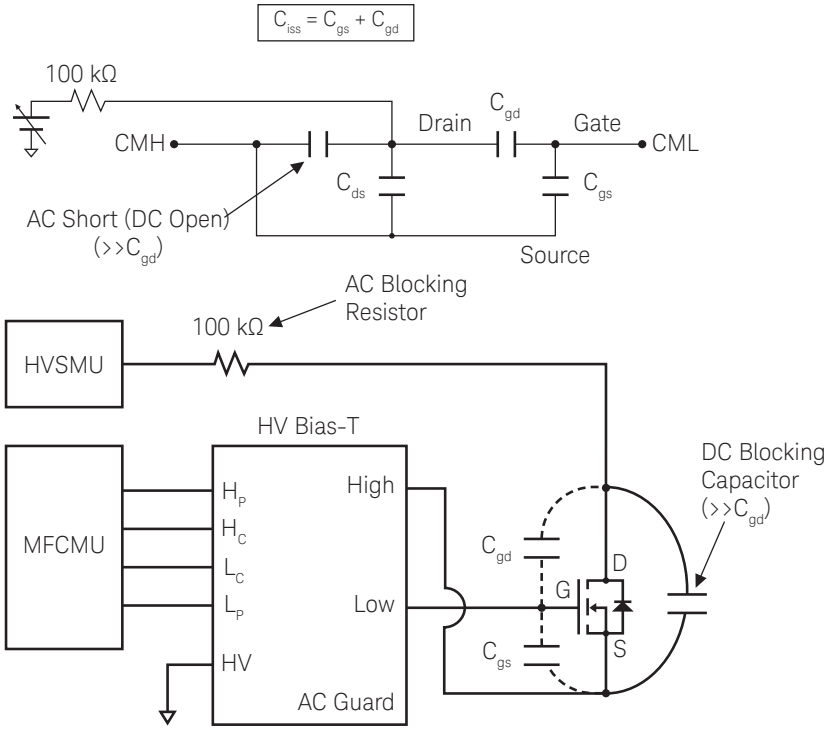
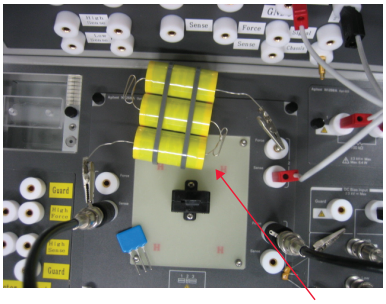


Figure 9.27. Measuring C_{iss} using the B1505A's HVSMU, MFCMU and high-voltage Bias-T.

A large capacitor is placed between the CMH and drain terminals to short out the drain-to-source capacitance (C_{ds}) to AC signals. This DC blocking capacitor has to be much larger than the gate-to-drain capacitance (C_{gd}) so that the effective impedance between the CMH and CML terminals consists only of the parallel combination of C_{gs} and C_{gd}. As in the case of the C_{gs} measurement previously discussed, the HVSMU must be connected to the drain node through a relatively large resistor to prevent it from interfering with the AC capacitance measurement. The following is an illustration showing these connections for a packaged MOSFET using the N1259A test fixture.



The capacitors necessary to measure Ciss can be quite large physically.

Figure 9.28. Measuring C_{iss} on a MOSFET using the N1259A test fixture.

Measuring capacitance of “normally on” devices

In all of the power transistor capacitance measurement examples so far, we have assumed that the transistor is not on when zero volts of DC bias is being applied to its gate (i.e. it is “normally off”). However, for some power devices, especially lateral GaN devices, this is not a valid assumption. If a power device has a negative threshold voltage, then the previous techniques need to be modified, since the HVSMU cannot supply current to a transistor in its on state. The obvious modification is that the gate needs to be biased sufficiently negative with respect to the source to make sure that the transistor never turns on during capacitance measurement. The challenge is that while we are negatively biasing the gate to source we also have to employ all of the other techniques we just covered to correctly measure the junction capacitances.

As before, the measurement of the gate to drain (C_{gd} or C_{rss}) is the simplest, so we shall cover it first. In this case, we use an additional SMU (either an MPSMU or a HPSMU) to bias the source positive with respect to the gate. However, we have to add a blocking resistor in this path to prevent the AC capacitance measurement from interfering with the MPSMU/HPSMU operation. In addition, we need to connect the source to the AC guard through a large capacitor that provides an AC short but prevents the bias SMU from interfering with the operation of the capacitance measurement. The following figure shows the details of this measurement setup.

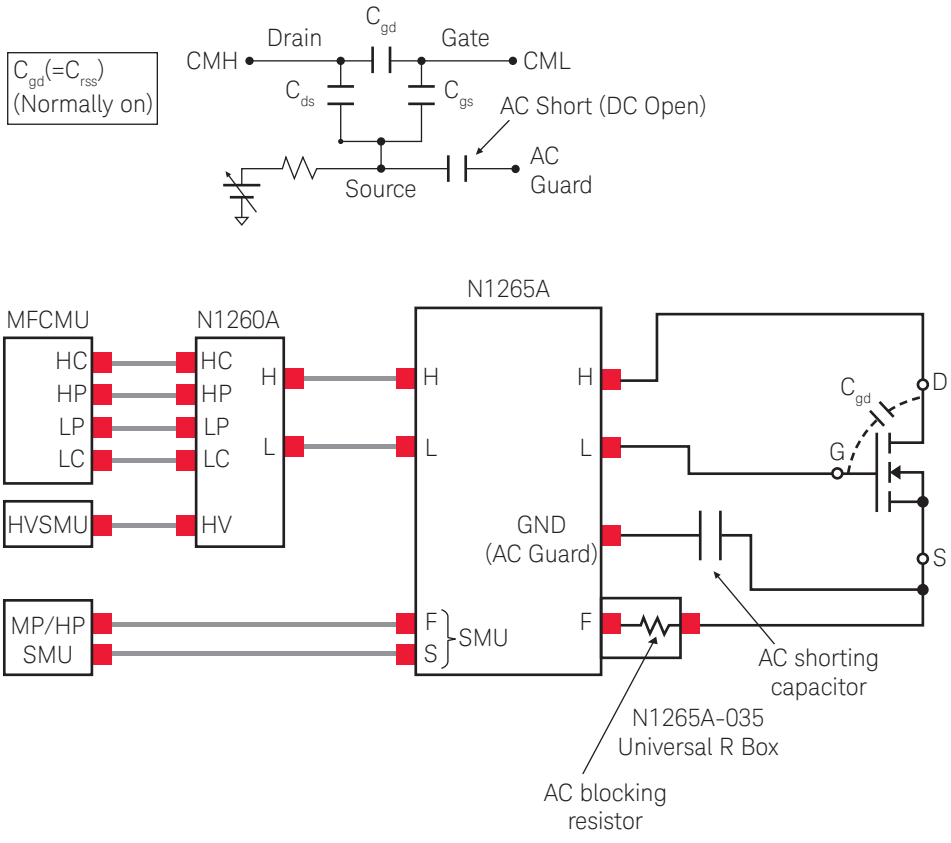


Figure 9.29. Measuring C_{gd} (=C_{rss}) on a normally on device using the N1260A Bias-T and the N1265A test fixture.

To measure the C_{oss} of a normally on device we only need to make a slight modification to the C_{gd} (or C_{rss}) measurement connections. Instead of connecting an AC shorting capacitor to the AC guard, we instead connect it directly to the gate (which is also connected to CML). This provides an AC short for the gate to source capacitance and places the gate to drain and drain to source capacitances in parallel. We can then use the additional MPSMU/HPSMU to bias the source positive with respect to the gate (through an AC blocking resistor) just as before. The following figure shows the details of this measurement setup.

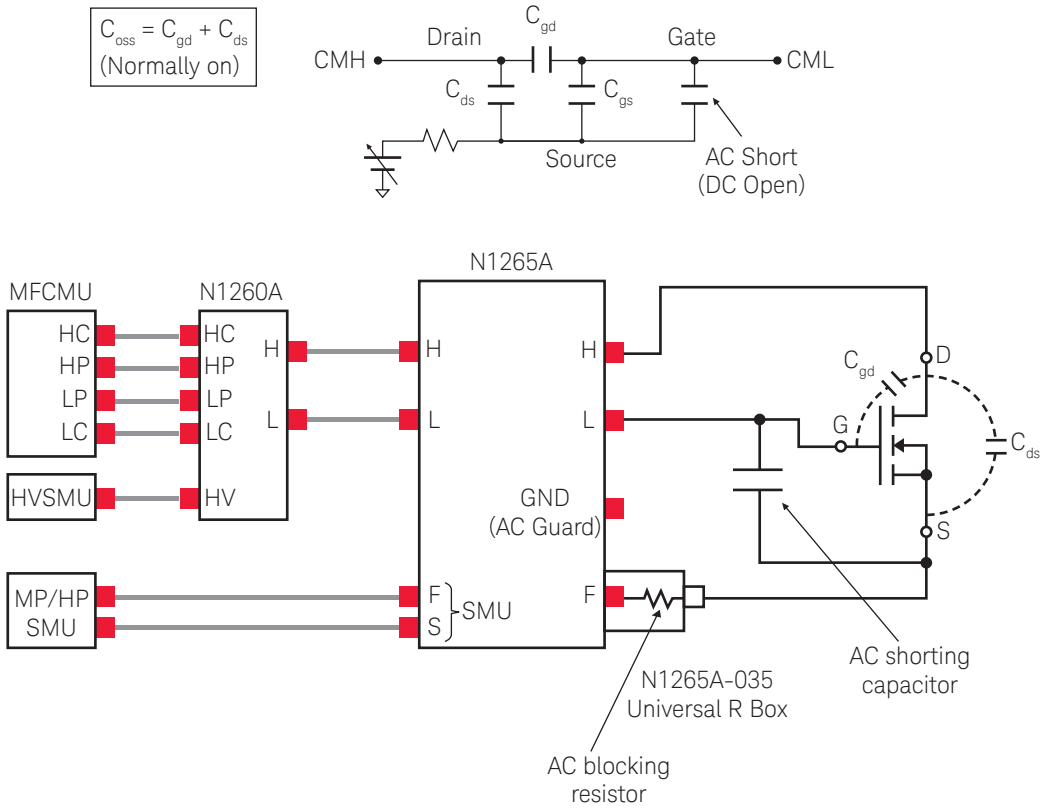


Figure 9.30. Measuring C_{oss} on a normally on device using the N1260A Bias-T and the N1265A test fixture.

As in the case of a normally off device, C_{iss} is the most challenging measurement. One major change is that the HVSMU needs to be directly connected to the drain through an AC blocking resistor (which can be the 100 kΩ series resistor that is included in the N1265A's module selector unit, if you are using it). In addition, the gate to source bias SMU is connected to the capacitance module using the Bias-T (an adapter is required). This allows the negative gate to source bias to be applied while high voltage is applied to the drain. Of course, an AC shorting capacitor also has to be connected between the drain and source. The following figure shows the details of this measurement setup.

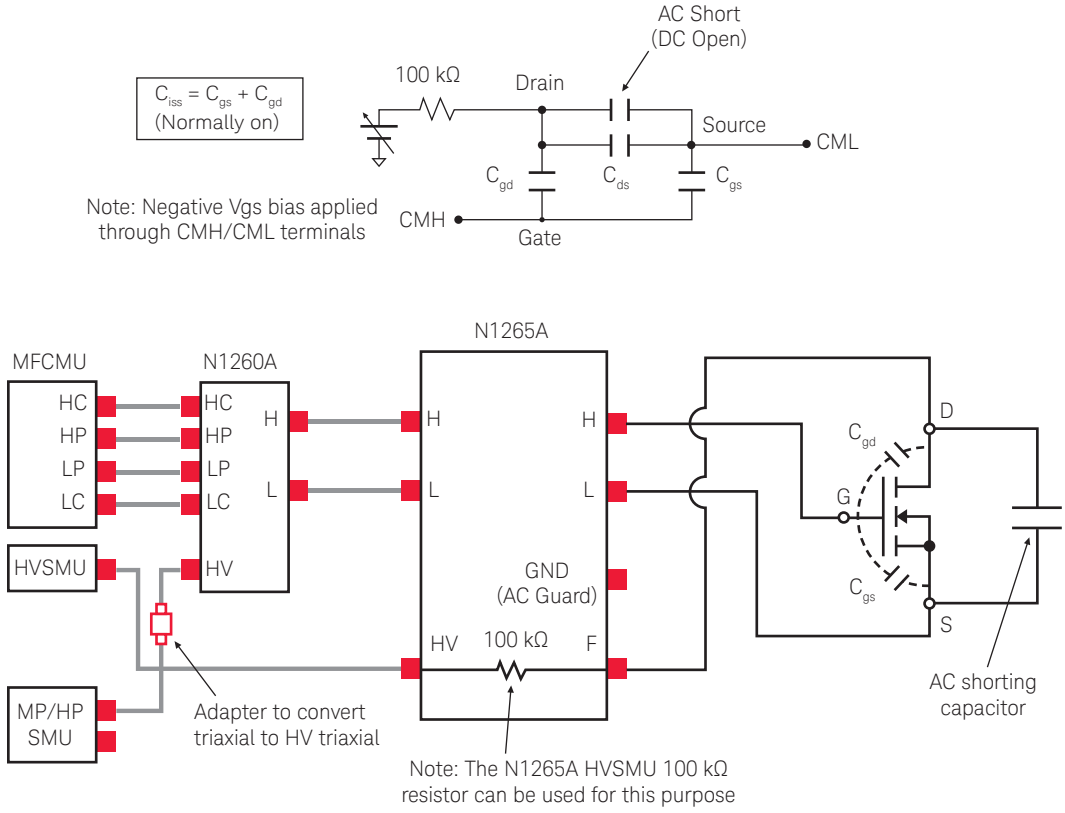


Figure 9.31. Measuring C_{iss} on a normally on device using the N1260A Bias-T and the N1265A test fixture.

From these discussions, it should be clear that performing capacitance measurements on power devices is not easy. It requires a good understanding of measurement theory, and the connections for each measurement can be quite complicated and require external components. Also, test automation is not possible as each different capacitance measurement requires connection changes. To address these challenges, Keysight created a solution that combines the high-voltage Bias-T with relays and other components to produce a complete automated capacitance measurement solution. This solution supports over twenty different configurations to enable capacitance measurement on all types of power devices (MOSFETs, IGBTs, BJTs, etc.). A block diagram of this solution is shown below.

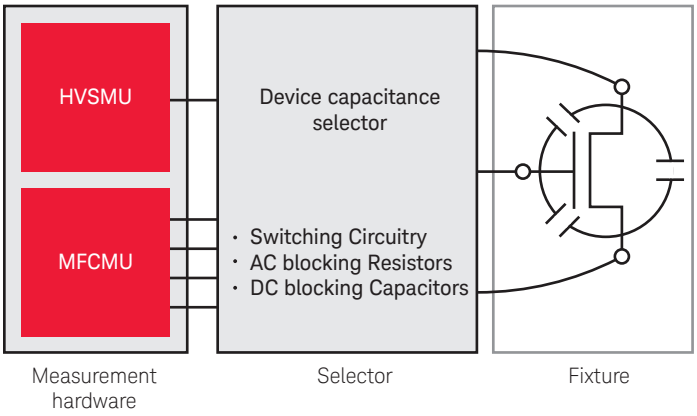


Figure 9.32. Keysight's automated power device capacitance measurement solution.

For B1506A units with the options that support capacitance and gate charge measurement, the device capacitance selector is integrated into the B1506A's test fixture. For the B1505A, you can obtain this functionality using the N1272A Device Capacitance Selector and the N1273A Capacitance Test Fixture. A picture of these B1505A accessories is shown below.

Figure 9.33. The N1272A device capacitance selector and N1273A capacitance test fixture for the B1505A.

Using this solution, both the B1505A and B1506A can perform automated capacitance measurements at up to 3 kV of DC bias. An example of these measurements for a power MOSFET device is shown below.

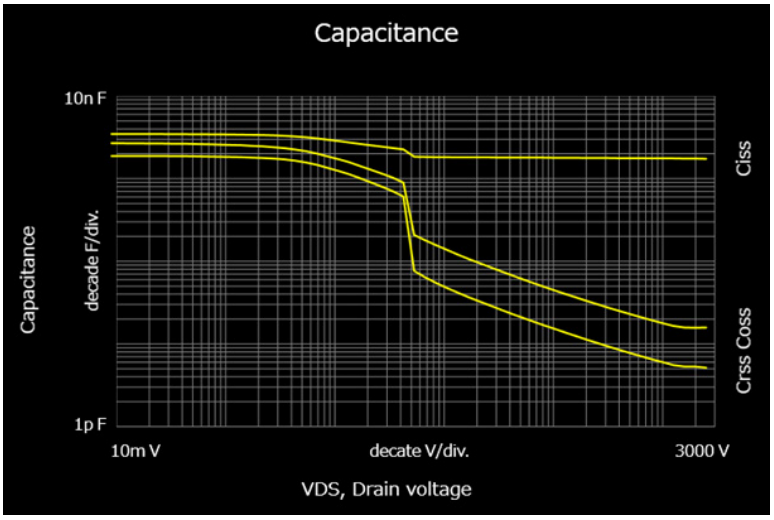


Figure 9.34. High-power MOSFET capacitance measurement example (C_{rss}, C_{iss} and C_{oss} measured up to 3 kV).

High-power capacitance measurement frequency issues

For low-power capacitance measurements, load compensation generally only needs to be performed for measurements above 5 MHz. However, when using the high-voltage Bias-T and making high-power capacitance measurements, load compensation needs to be performed at much lower frequencies to insure accurate measurement results. Most data sheets specify the C_{iss}, C_{oss} and C_{rss} at a frequency of 1 MHz. However, the following data shows a plot of gate to source (C_{gs}) capacitance (C_p-G) versus frequency using the high-voltage Bias-T after performing open/short capacitance compensation. Note that the measured conductance becomes negative as the frequency increases beyond 100 kHz.

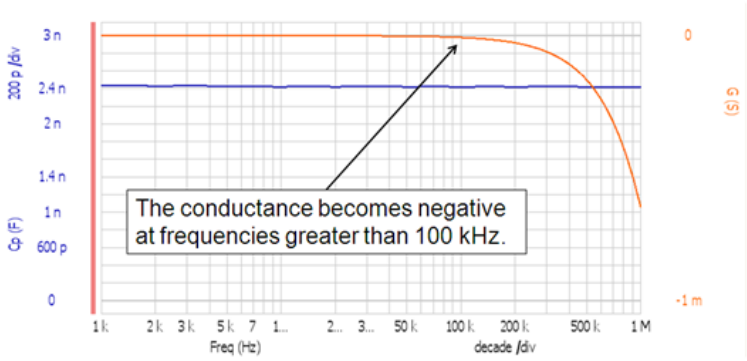


Figure 9.35. Plot of C_p-G versus frequency using the high-voltage Bias-T.

While this effect can be eliminated by performing a load capacitance compensation, it is not very practical to do a load compensation in most high-power device wafer probing environments. Therefore, if the load compensation cannot be performed, it is best to measure the MOSFET capacitance parameters at frequencies no greater than 100 kHz. The following table summarizes the relative measurement accuracy of a C_p-G measurement on a power MOSFET as a function of compensation performed for the both 100 kHz and 1 MHz (when using the high-voltage Bias-T).

	100 kHz		1 MHz	
	C_p	**G**	**C_p**	**G**
Open	Small error (1%)	OK	Large error	Large error
Open/Short	OK	OK	OK	Large error
Open/Short/Load	OK	OK	OK	OK

Figure 9.36. Table comparing relative accuracy of C_p-G measurements at 100 kHz and 1 MHz for different levels of capacitance compensation when using the high-voltage Bias-T.

Note: This table is based on the relatively large junction capacitances (on the order of nanofarads) that are typical of power MOSFETs.

Using the techniques just explained, it is possible to directly measure power MOSFET capacitances at voltages up to 3 kV. This represents considerable improvement over the conventional methods to measure these same parameters, which typically consist of homemade test setups that involve measuring a device's step response and extracting the value of capacitance from the resulting RC time constant.

Measuring gate resistance

Gate resistance is measured using an LCR meter, which is why it falls under the umbrella of capacitance measurement. To understand this, it is helpful to look at a trench MOSFET example:

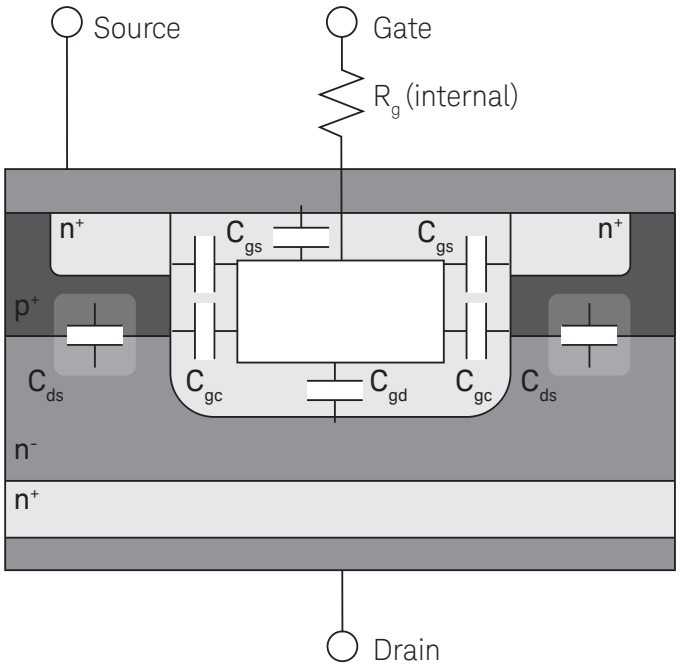


Figure 9.37. Trench MOSFET cross-section showing the various capacitances and their relation to the gate resistance.

The gate resistance is in series with three internal junction capacitances:

1. C_{gs} (gate to source capacitance)
2. C_{gd} (gate to drain capacitance)
3. C_{gc} (gate to channel capacitance)

Under the correct bias conditions, by measuring the series capacitance-resistance (C_s-R_s) between the gate and source terminals we can determine R_g.

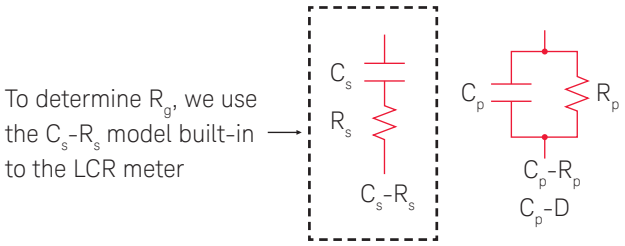


Figure 9.38. Measuring gate resistance using the LCR meter C_s-R_s model.

The measured value of R_g depends on the gate bias, and in general it is best to measure it using a value of V_{gs} that keeps the device completely off. When the device is off, the gate to source capacitance (C_{gs}) becomes the dominant factor in the series capacitance as shown in the following figure.

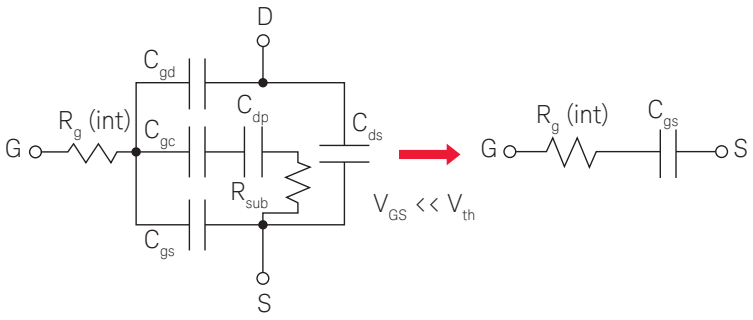


Figure 9.39. Equivalent transistor circuit when $V_{gs} < V_{th}$.

The exception to this rule is when the device on-resistance (R_{ds}) is much less than the gate resistance. In this case, it is OK to measure the gate resistance using a value of V_{gs} that completely turns on the device as shown in the following figure.

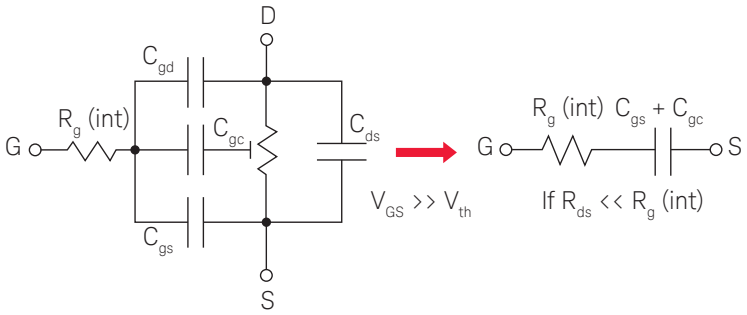


Figure 9.40. Equivalent transistor circuit when $V_{gs} > V_{th}$.

Gate charge measurement

The concept of gate charge measurement is relatively straightforward. The transistor is switched on or off while driving a specified load and during this time the current going into the gate is monitored. Provided that you can measure gate current and time accurately, you can determine the charge going into (or out of) the gate. Traditionally, gate charge is performed using three different types of loading:

1. Constant current load
2. Resistive load
3. Inductive load

The key challenges to making this measurement are as follows:

1. Obtaining a VDD supply capable of supplying high voltage in the off state and high current in the on state.
2. Obtaining a gate drive supply stable enough to provide accurate time-dependent output voltage and current.
3. Designing a gate drive circuit that can accurately measure time-dependent current and voltage.

The two-pass (HV/HC) gate charge measurement technique

The two-pass gate charge measurement technique eliminates the need to simultaneously provide both high voltage and high current by breaking the gate charge measurement down into two parts. The assumptions made for the two-pass gate charge method are as follows:

1. The values of C_{gs}(off) and C_{iss}(off) are almost the same at high-current and high-voltage.
2. The value of V_{ds}(on) is virtually the same for high-current and high-voltage operation.

These assumptions are both reasonable for devices typically used in inverter/converter circuits.

The high-voltage and high-current curves are combined together as follows:

1. The derived gate charge curve tracks the initial slope of the HV curve until it reaches the level of the plateau voltage measured by the HC curve.
2. The derived gate charge curve then follows the HC plateau voltage line until it intercepts the HV curve.
3. The derived gate charge curve then extends upward from this point using the slope of the HC curve.

The following figure illustrates this procedure.

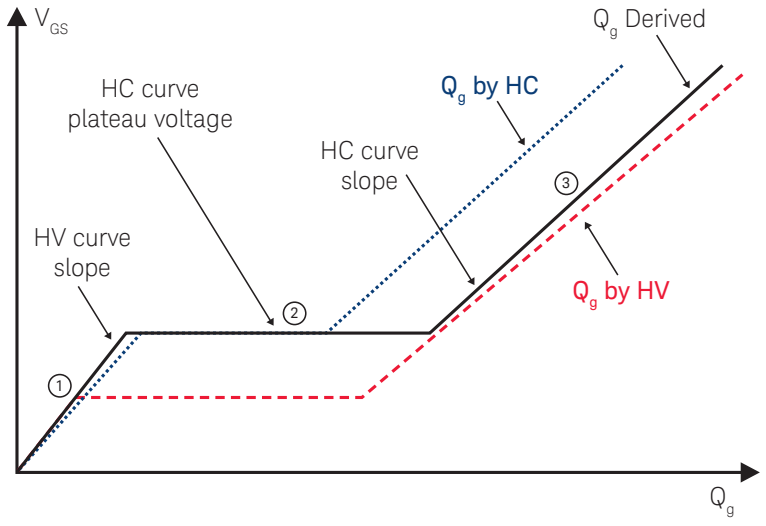


Figure 9.41. Details of the two-pass gate charge measurement technique.

The Keysight B1505A and B1506A can both measure gate charge, but in general, a gate charge adapter is required for this measurement. This arises from the requirement of gate charge to be measured while driving a specified load, which necessitates a special test fixture to accommodate the load. In addition, when making a constant current load measurement, the B1505A requires a manual connection change when switching from high-voltage to high-current modes (Note: The B1506A does not require a manual connection change when switching modes for gate charge measurements as this switching capability is built into its test fixture). The reason for this becomes obvious when you look at the connections necessary for each measurement. Let us first consider the connections for the high voltage gate charge measurement.

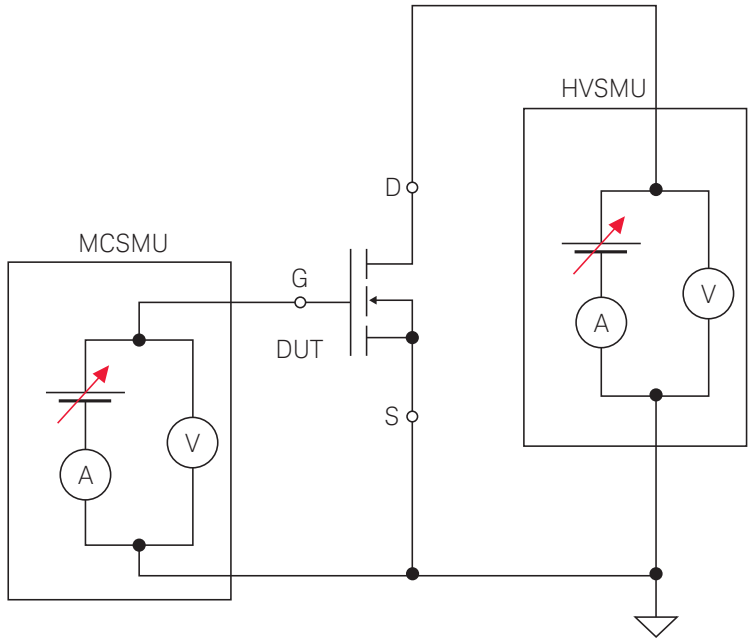


Figure 9.42. High-voltage (HV) gate charge measurement connection scheme.

When we are measuring the HV gate charge curve there is no load. However, when making the high-current gate charge curve there are two supported load options.

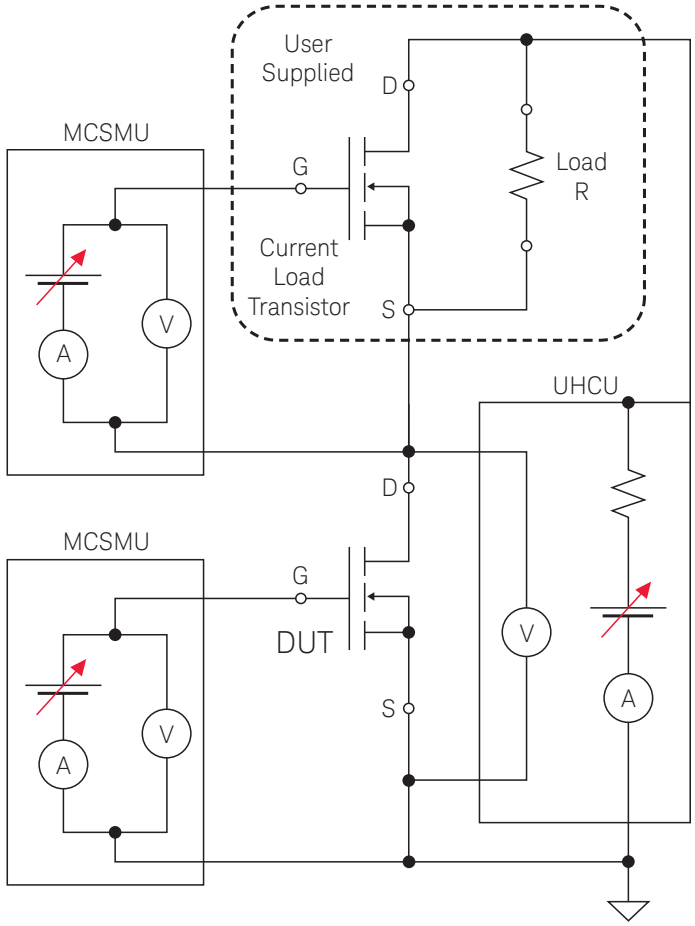


Figure 9.43. High-current (HC) gate charge measurement connection scheme showing both supported load options.

Note: It is important to understand that the two load options shown for the HC gate charge measurement are mutually exclusive. The user must supply and physically connect either a current load transistor or a load resistor into the gate charge adapter. The reasons a physical connection change is required when using a current load and going from HC gate charge measurement to HV charge measurement should now be clear. You cannot place high-voltage across the load transistor without risking damage to it. A load resistor does not present the same issue since it cannot be damaged, and the current coming from the HVSMU is very small (a few milliamps), and does not have any appreciable effect on the voltage seen by the DUT.

Turning the theoretical concepts just discussed into actual gate charge measurements requires a combination of knowledge, skill and occasionally, a little bit of luck. The easiest way to explain the methodologies involved is via concrete measurement examples. Although we will use the B1505A and EasyEXPERT software in these examples, the measurements work equally well on the B1506A using Easy Test Navigator software.

Gate charge test fixture overview

The gate charge test fixtures for the B1505A and B1506A are very similar. Both test fixtures have a socket for a load device to support a constant current load, as well as terminals to support a resistor for resistive loading. Please refer to the figure shown below.

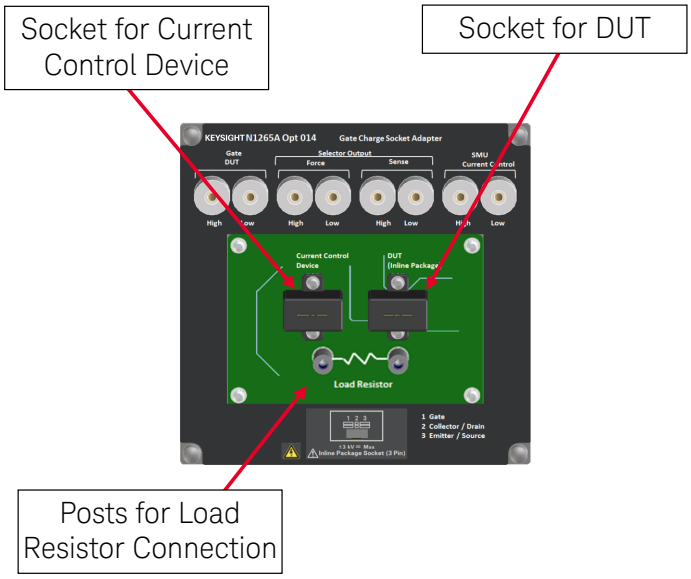


Figure 9.44. The gate charge test fixture for the B1505A supports both constant current loads and resistive loads.

The B1506A gate charge adapter has banana plugs in its front to support the testing of modules. The B1506A comes with cables and an isolation pad to support module gate charge measurement. Note: The B1505A can perform on-wafer gate charge measurements, but requires adapters that will be discussed later.

Gate charge test fixture calibration and setup

Gate charge measurement is not a simple IV measurement, so to get accurate measurement data, the test fixture requires calibration. It is important to remember that if you change adapters or test leads, then you need to recalibrate the gate charge test fixture. In addition, if you add a gate to source capacitor to your test fixture to suppress oscillations, then you also need to recalibrate the gate charge test fixture. It is best to perform both open and short calibration on the test fixture, although either one of these can be skipped if desired. There are three calibration options:

1. Start a new calibration
2. Recall existing calibration data (saved in a file)
3. Recall factory default calibration data

If you are using the B1505A and N1265A test fixture with either the 500 A or 1500 A UHC option, then make sure that the gate resistor option is set to zero Ohms. This allows you to specify the gate resistance through the EasyEXPERT gate charge application tests. Also, if you are using the B1505A with the N1272A device capacitance selector and you are using the same SMU for biasing during capacitance measurements as you are using for your gate load current drive, then you need to deselect it in the N1272A setup window.

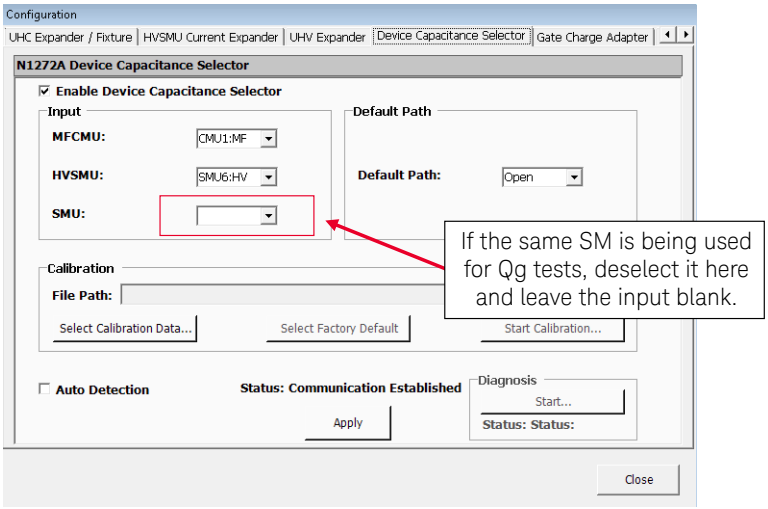


Figure 9.45. Leave the SMU input blank in the N1272A setup window if using the same SMU for your gate charge measurement.

The final point to remember is that you need to configure the gate charge adapter as shown in the following figure. Note that you need to specify both the SMU going to the gate of your DUT as well as the SMU going to the gate of your current load device (if one is being used). It is crucial not to confuse these.

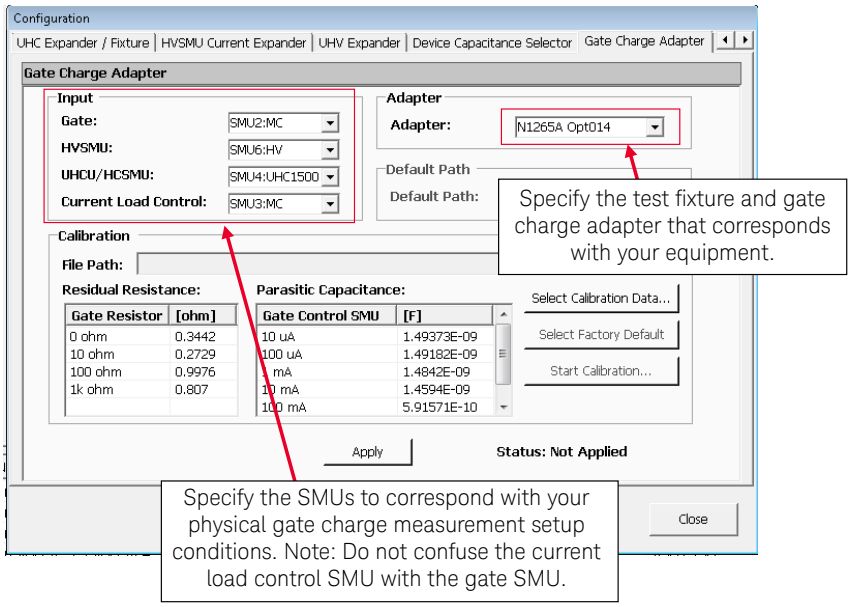


Figure 9.46. EasyEXPERT gate charge adapter setup example.

Gate Charge Measurement on a Silicon IGBT

We will now show an example of gate charge measurement on a silicon IGBT that has the following basic characteristics:

- VCES: 330 V
- Maximum IC (DC): 180 A
- Maximum IC (pulsed): 450 A

- C_{ies}: 3880 pF
- C_{oes}: 305 pF
- C_{res}: 180 pF
  (VGE = 0 V, VCE = 30 V, f = 1 MHz)

In application test mode, we select the “IGBT” library and type “Q” in the search window to display the gate charge application tests for IGBTs (which all start with “Q_g”) as shown in the following figure.

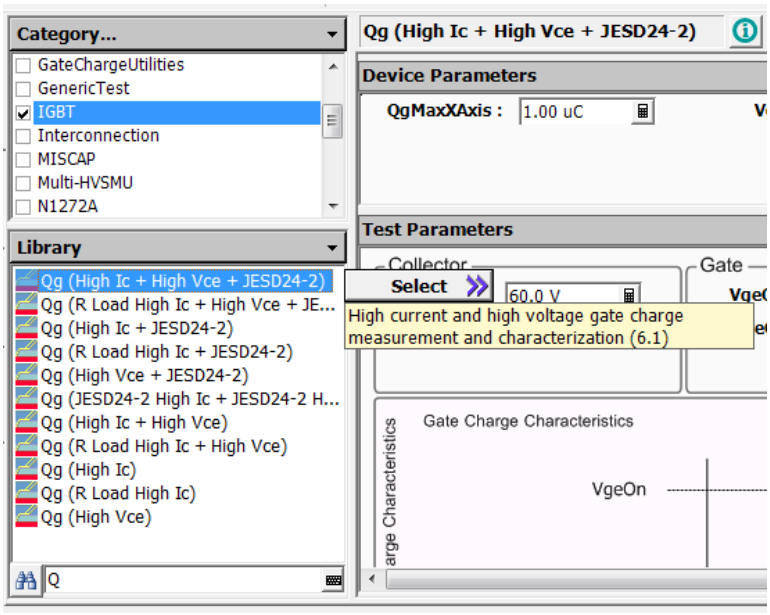


Figure 9.47. Selecting the IGBT gate charge application tests.

We have a datasheet for this device that we can use to help select the parameters for the application test. The data sheet lists the total gate charge as 169 nc, so we can safely set the “QgMaxXAxis” parameter to 200 nc. The typical gate to emitter threshold voltage is listed as 4.0 V, so this provides us with a value for the “VgeTh” parameter. The data sheet lists the gate to emitter gate charge test conditions as Vce = 200 V and Ic = 40 A, so we can use these for the “VceOff” and “IcOn” parameters respectively. Finally, the data sheet shows gate to collector gate charge test condition as Vge = 15 V, so this provides us with a value for “VgeOn”. Determining the gate current (Ig) requires a little calculation and some knowledge of the equipment. From the data sheet’s Qg plot, we can assume that the maximum expected value of Q_g is 180 nC. However, we also need to account for the SMU’s stray output capacitance, which is approximately 1.6 nF per volt. Although we set the” VgeOn” condition at 15 V, there is an additional 3.5 V of swing allowed under the conditions set in the “High Voltage Setup”. This makes the potential swing 18.5 V, and we then use this value to calculate the capacitance due to the SMU.

Stray Output Capacitance Due to SMU = 1.6 nF x 18.5 V = 29.6 nC

This then allows us to calculate the total charge.

Total Charge = 180 nC + 30 nC = 210 nC

The B1505A and B1506A measurement hardware require that the gate charge measurement takes place over a 400 microsecond interval. Therefore, we can use the above capacitance to calculate a rough value for the maximum Ig required.

$$I_g = \frac{210\ nC}{400\ \mu s} \approx 525\ \mu A$$

An I_g of 1 mA is therefore a reasonable starting value. The following figure shows the key inputs to the IGBT gate charge application test.

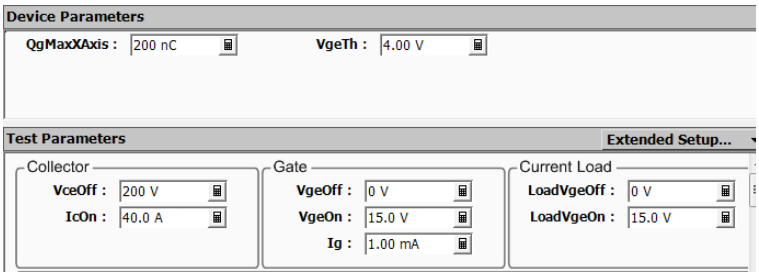


Figure 9.48. The key inputs for the IGBT gate charge measurement application test filled in using data sheet parameters and the calculated value for I_g.

Although it may be necessary to add in some external gate resistance to prevent oscillation, it is best to start with these parameters set to zero and modify them only if necessary.

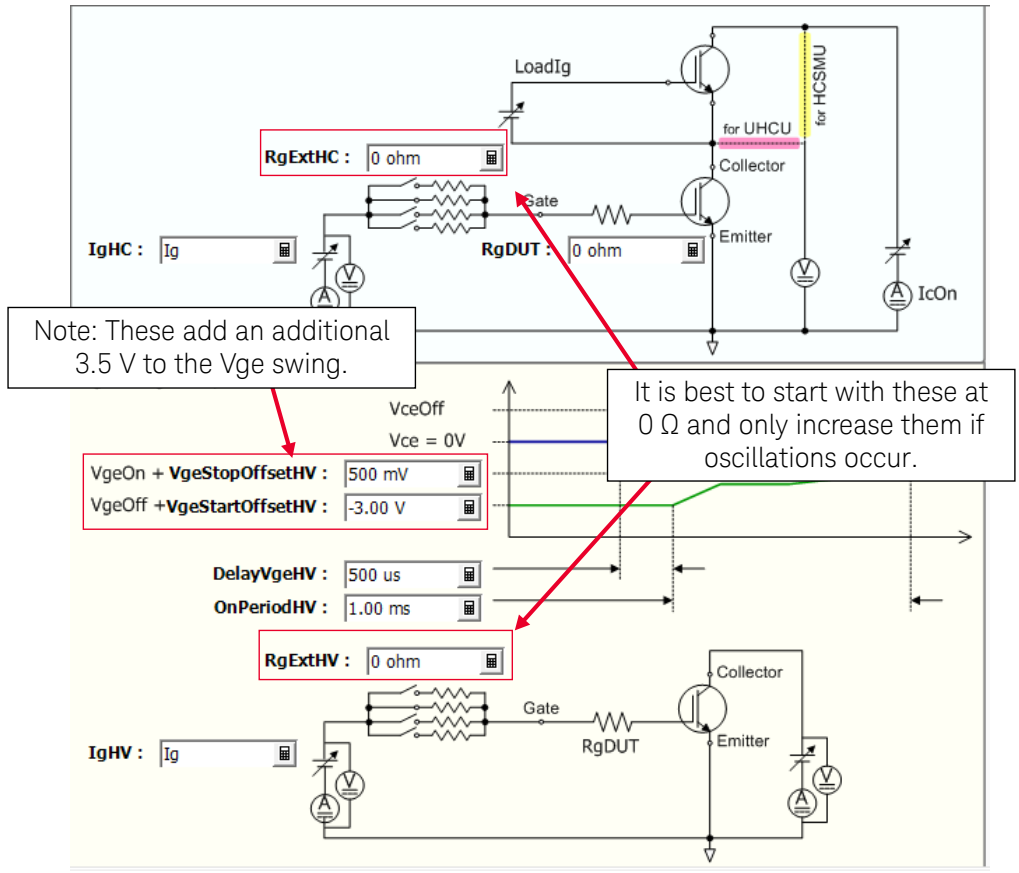


Figure 9.49. The external gate resistance parameters for the gate charge application test can be increased from zero to prevent oscillation if necessary.

When making a gate charge measurement in EasyEXPERT, it is best to have “Multi Display” and “Auto Record” modes set to “ON” (see very bottom of EasyEXPERT screen). Also, in the “Results” window at the bottom of the EasyEXPERT screen you should have the filter set to “Expand Application Test Results” as shown in the following screen capture.

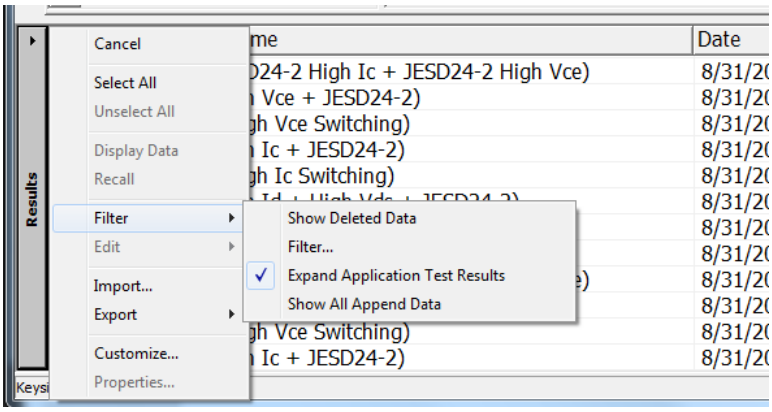


Figure 9.50. Configure the results window to expand the application test results.

Doing this allows you to record and view the high voltage and high current switching waveforms, which is important for debug and verification. The next step is to make the physical cable connections necessary to perform a high current measurement using a current load device. When you click on the “Measure” button to perform the application test, a window will appear to remind you to set up the cabling for a high current measurement. Once the high current measurement completes, then a second window will appear, prompting you to modify the physical cable connections to perform a high voltage measurement. After doing this and clicking on “OK” then the gate charge curve will appear (if the measurement completes without errors).

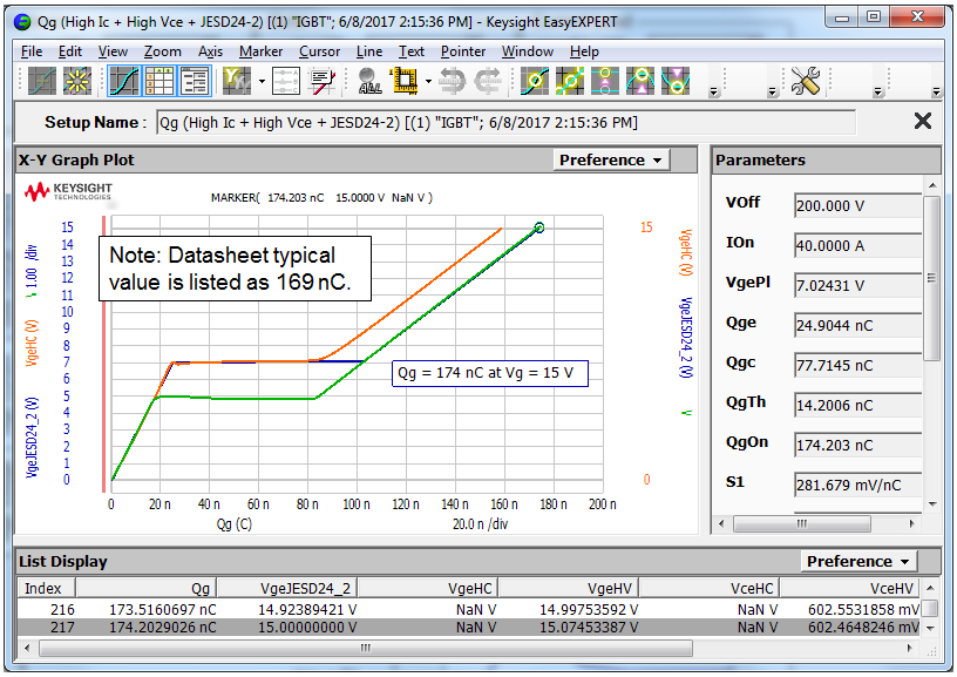


Figure 9.51. Gate charge measurement result for IGBT device.

The measured value of 174 nC for the gate charge is very close to the data sheet value of 169 nC. Note: The gate charge application tests show the HC and HV curves in addition to the derived gate total gate charge curve, but the HC and HV curves can be removed from the display if desired.

It is always a good idea to review the switching waveforms after completing a gate charge measurement. Since we previously set the filter to expand application tests, we can select the two saved switching waveforms and do a “Display Data”.

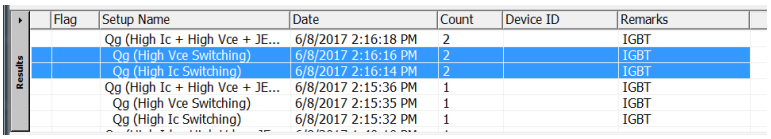


Figure 9.52. Select the two saved switching waveforms and display the data to view the switching waveform behavior.

We will first look at the high-current switching waveforms.

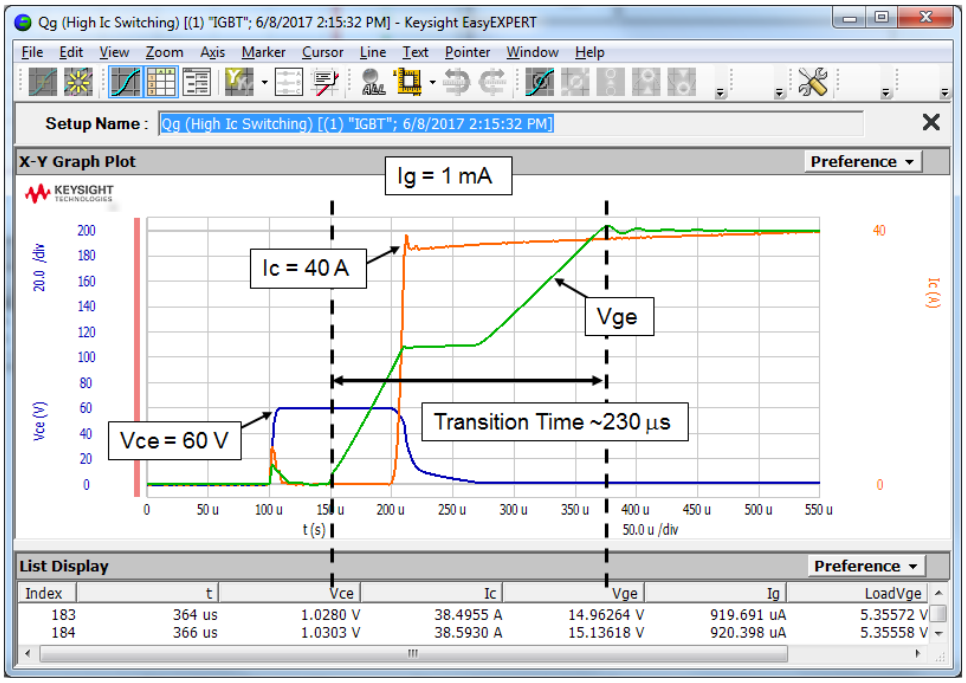


Figure 9.53. The high current switching waveforms for the IGBT gate charge measurement with 1 milliamp of gate current.

Although the waveforms look good, the transition time is only about 230 microseconds. You typically want the transition time to be on the order of 300 to 350 microseconds, so the gate drive current should be decreased. Looking at the high-voltage switching waveforms, we can also see that the transition is a little too fast.

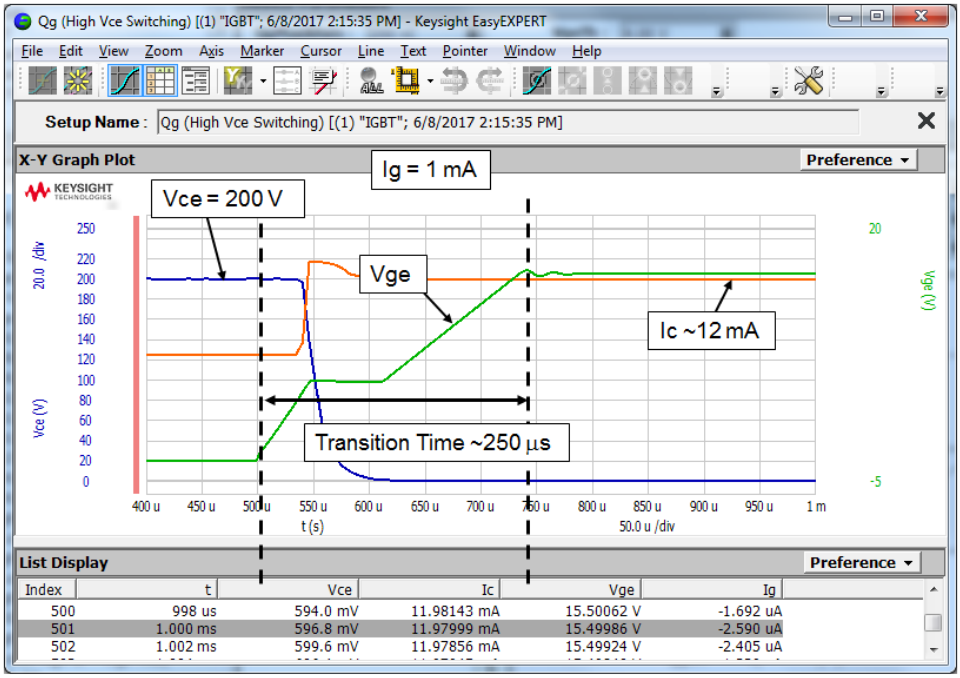


Figure 9.54. The high voltage current switching waveforms for the IGBT gate charge measurement with 1 milliamp of gate current.

Decreasing the I_g from 1 mA to 750 microamps will improve the appearance of these waveforms. The following figures show the high current switching waveform for this case.

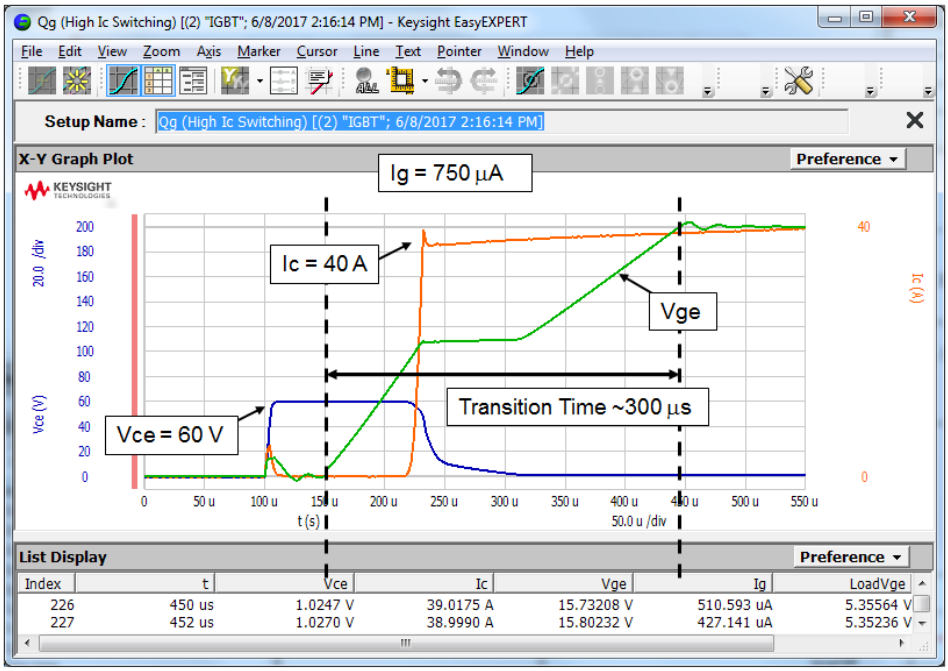


Figure 9.55. The high current switching waveforms for the IGBT gate charge measurement with 750 microamps of gate current.

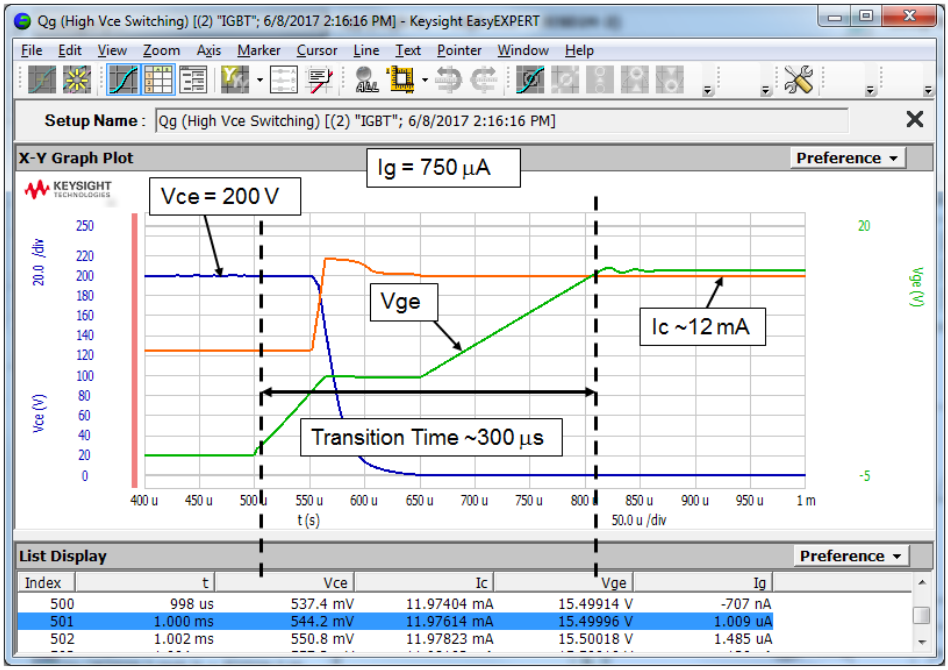


Figure 9.56. The high voltage current switching waveforms for the IGBT gate charge measurement with 750 microamps of gate current.

The following figure shows the updated derived gate charge curve with the HC and HV curves removed.

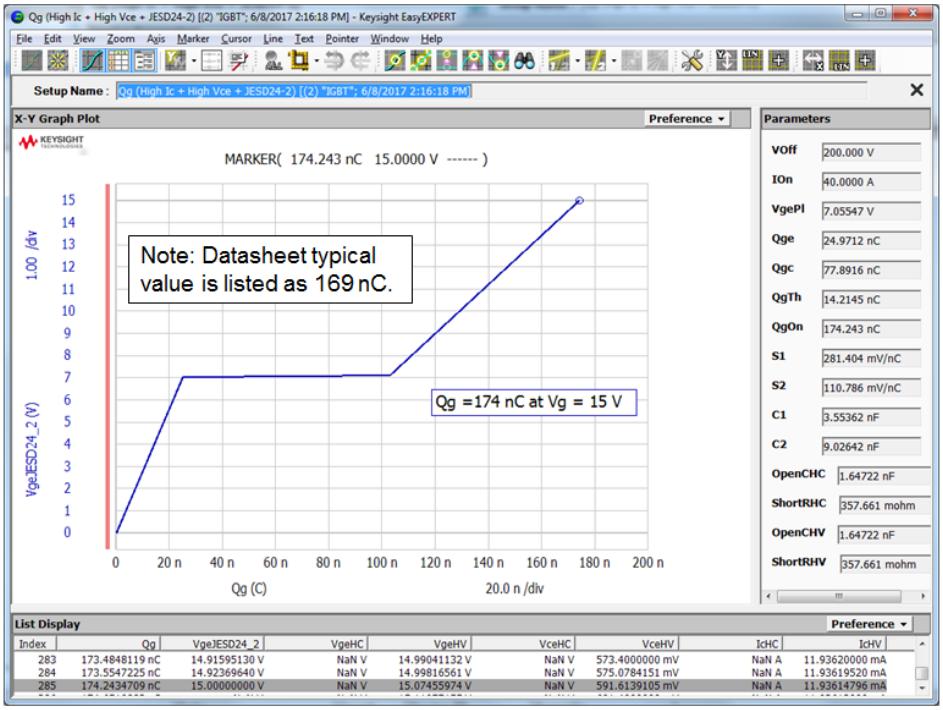


Figure 9.57. The updated gate charge curve with the HC and HV curves removed.

As before, the measured value of 174 nC is very close to the data sheet typical value of 169 nC.

Gate charge measurement using UHCU with R-load

Since it is difficult or impossible to find load devices capable of handling hundreds of amps of current, high current gate charge measurements using the ultra-high current unit (UHCU) are usually done with a resistive load. In this example, we will measure the gate charge of a power MOSFET at Id = 195 A. The characteristics of this MOSFET are shown below.

- VDSS: 40 V
- ID(DC): 350 A
- ID(pulse): 1,390 A

- C_{iss}: 8,920 pF
- C_{oss}: 2360 pF
- C_{rss}: 930 pF
  (V_{gs} = 0 V, VDS = 25 V, f = 1 MHz)

When picking the load resistor value, you need to take into consideration the output resistance of the UHCU in the regime in which you intend to operate. For the UHCU we have:

- UHCU Rout = 120 mΩ, I ≤ 500 A
- UHCU Rout = 40 mΩ, I > 500 A

We want to measure the gate charge with a drain current of 195 A, and apply a drain-to-source voltage of 32 V. We can then calculate the required resistance to be:

$$R = \frac{V}{I} = \frac{(32\ V)}{(195\ A)} \approx 160\ m\Omega$$

Since the output resistance of UHCU is 120 mΩ, we need to add an external resistance of approximately 40 mΩ.

In application test mode, we select the “PowerMOSFET” library and type “Q” in the search window to display the gate charge application tests for power MOSFETs as shown in the following figure.

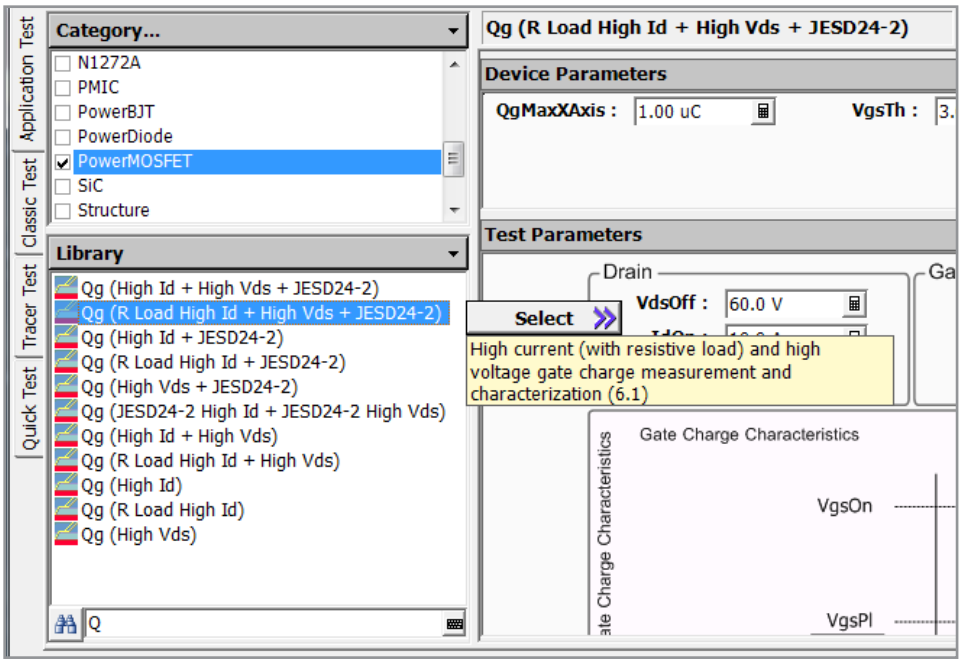


Figure 9.58 - Selecting the power MOSFET gate charge application tests.

As was the previous case for the IGBT, we have a datasheet for this MOSFET device and we can use it to help select the parameters for the application test. The data sheet lists the total gate charge as 220 nc, therefore we set the "QgMaxXAxis" parameter to 250 nc. The gate to source threshold voltage is given as a range from 2.0 V to 4.0 V, and we take the midpoint of these two values and set the "VgsTh" parameter to 3.0 V. The datasheet also shows the gate to source gate charge test condition as V_{gs} = 10 V, which provides us with a value for "VgsOn". The datasheet also specifies that gate charge is measured with Id = 195 A, and we have calculated a load resistance using a value of Vds = 32 V. This then allows us to fill in values for "VdsOff" and "IdOn" as shown in the following figure.

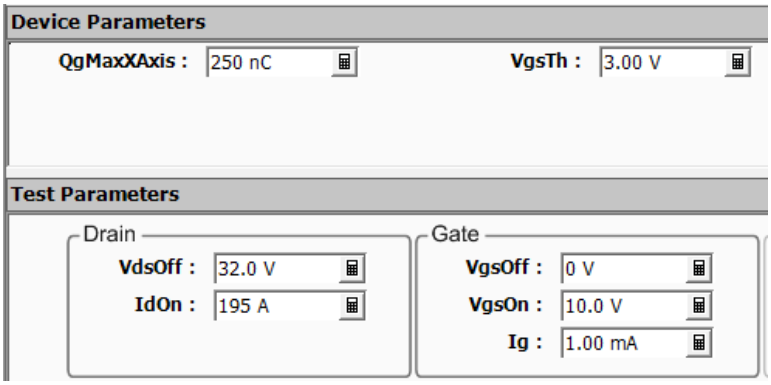


Figure 9.59. The key inputs for the power MOSFET gate charge measurement application test filled in using data sheet parameters and the calculated value for I_g.

Note: The application test will automatically adjust the value of VdsOff to obtain the specified value of IdOn; the specified value of VdsOff is only used as a starting point. In addition to the parameters shown above, as for the case of the IGBT gate charge measurement, we set the external gate resistor values to zero Ohms as a starting point.

We can now proceed to run the gate charge application test. However, before proceeding, it is worthwhile to note that in this case we do not have to make any connection changes when switching from high current to high voltage testing. Because the value of the external resistance is so small (40 milliohms) and the current during high-voltage test is also so small (milliamps), the inclusion of the resistor has no effect on the high voltage test results. After running the test, we see the following gate charge curve.

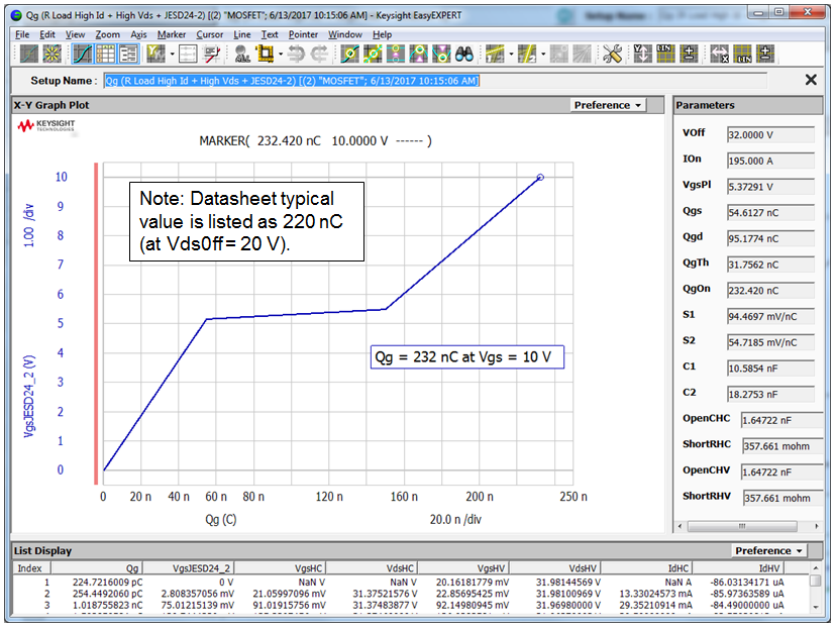


Figure 9.60. Gate charge measurement for power MOSFET at VdsOff = 32 V.

The measured gate charge is 232 nC, which is higher than the typical listed datasheet parameter of 220 nC. The datasheet uses a VdsOff value of 20 V, whereas we applied a value of 32 V. Measuring gate charge at Id = 195 A and Vds = 20 V is technically not possible as the following calculation using the output resistance of the UHCU shows.

$$I = \frac{V}{R} = \frac{20\ V}{120\ m\Omega} \approx 167\ A$$

Even with no external resistor, the UHCU cannot output 200 A at 20 V. However, the expected difference in the slope of the gate charge curve with VdsOff = 20 V and VdsOff = 24 V should not be very much. Therefore, we can try re-running the test with a smaller value of VdsOff = 24 V, which yields the following gate charge curve.

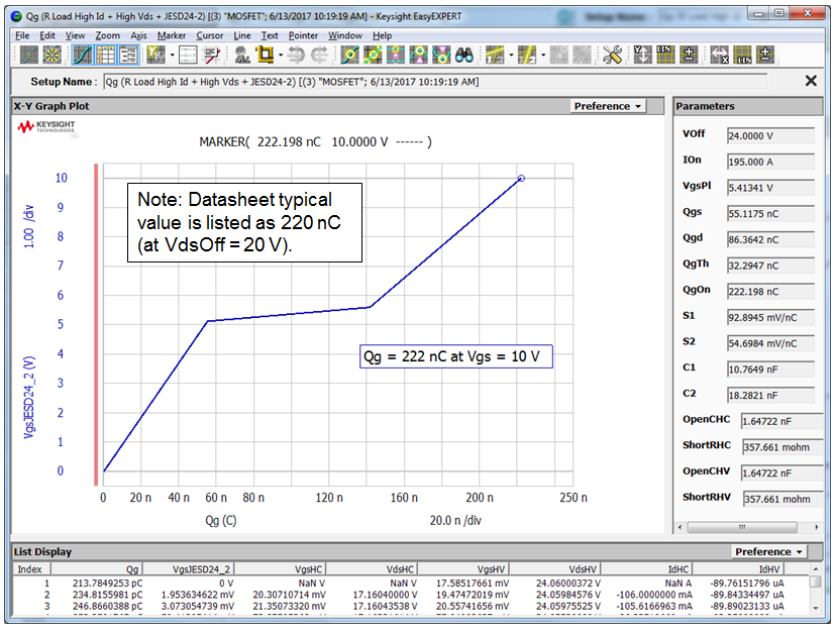


Figure 9.61. Gate charge measurement for power MOSFET at VdsOff = 24 V.

This gives very good agreement with the datasheet value of 220 nC.

As before it is a good idea to examine the high current and high voltage switching waveforms to make sure that the gate charge measurement executed correctly. The following two figures show the waveforms for this measurement.

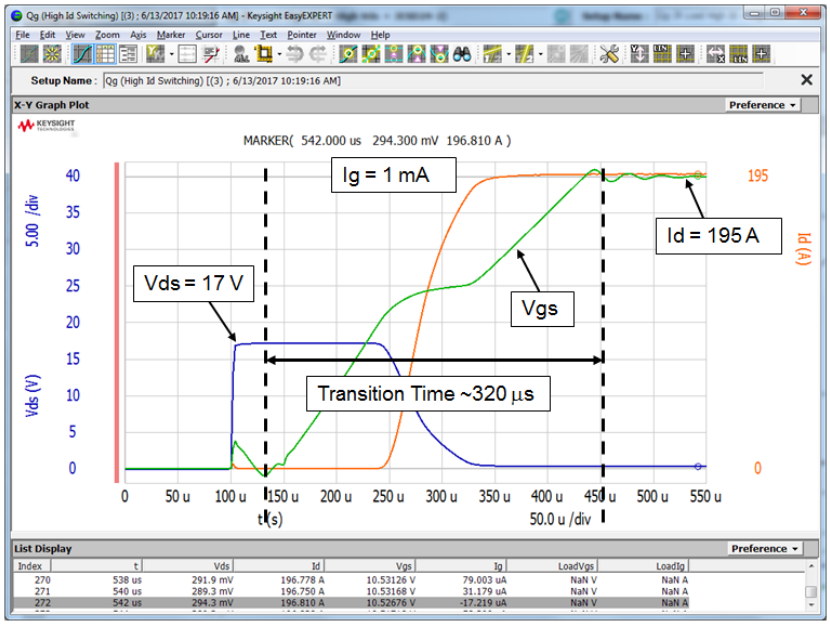


Figure 9.62. The high current switching waveforms for the MOSFET gate charge measurement with VdsOff = 24 V.

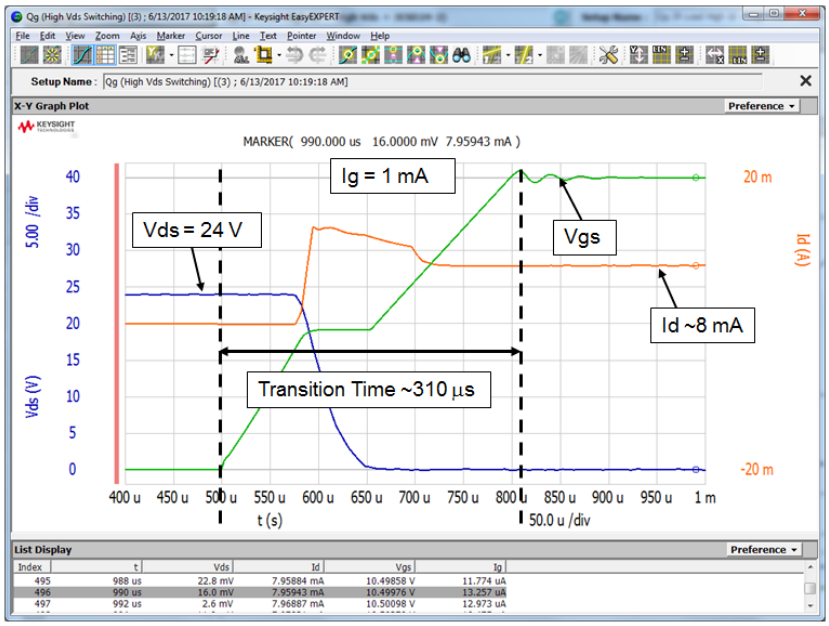


Figure 9.63. The high voltage switching waveforms for the MOSFET gate charge measurement with VdsOff = 24 V.

Both of the waveforms looks reasonable, so there is no need to modify any of the test parameters.

On-wafer gate charge measurement

In most cases, performing on-wafer measurement of power devices does not yield data relevant to the performance of the packaged devices, since so much of the device performance is ultimately tied in with the thermal performance of the package. However, gate charge measurement is one exception to this, as the on-wafer gate charge measurement typically does not vary greatly from that of a packaged device. The only instrument that supports on-wafer gate charge measurement is the B1505A, and there are two cases to consider:

1. The B1505A HCSMU is used to measure gate charge.
2. The B1505A UHCU is used to measure gate charge.

Both of these solutions support the module selector unit, making it possible to switch between high current, high voltage and gate charge measurement without having to do any recabling. Although very similar in approach, we will discuss each case separately.

On-Wafer Gate Charge Measurement using the HCSMU (B1512A)

When making gate charge measurements using the high current SMU (B1512A), the N1274A gate charge adapter is required. The following figure shows an illustration of the N1274A.

Figure 9.64. The N1274A (20 A/3 kV) gate charge adapter for wafer probing.

Note: The N1274A has a small “compartment” with a removable lid into which you can place a TO packaged device to use as a current load. The connection scheme for measuring a vertical power device on-wafer is shown in the following figure.

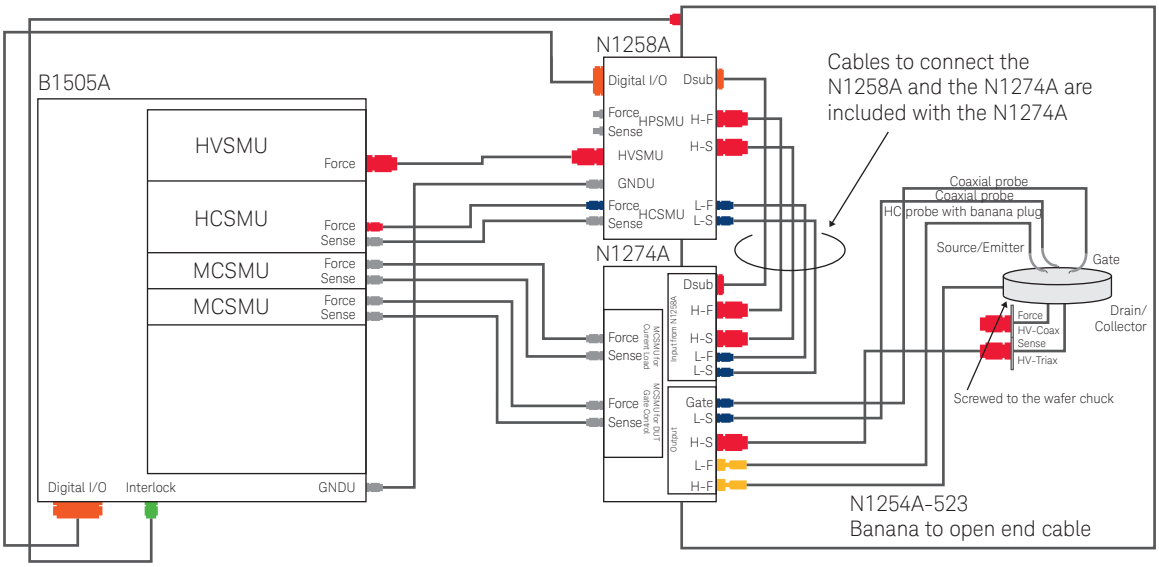


Figure 9.65. Connection scheme to measure a vertical power transistor using the N1274A gate charge adapter and the N1258A module selector.

The connection scheme for measuring a lateral power device on-wafer is shown in the following figure.

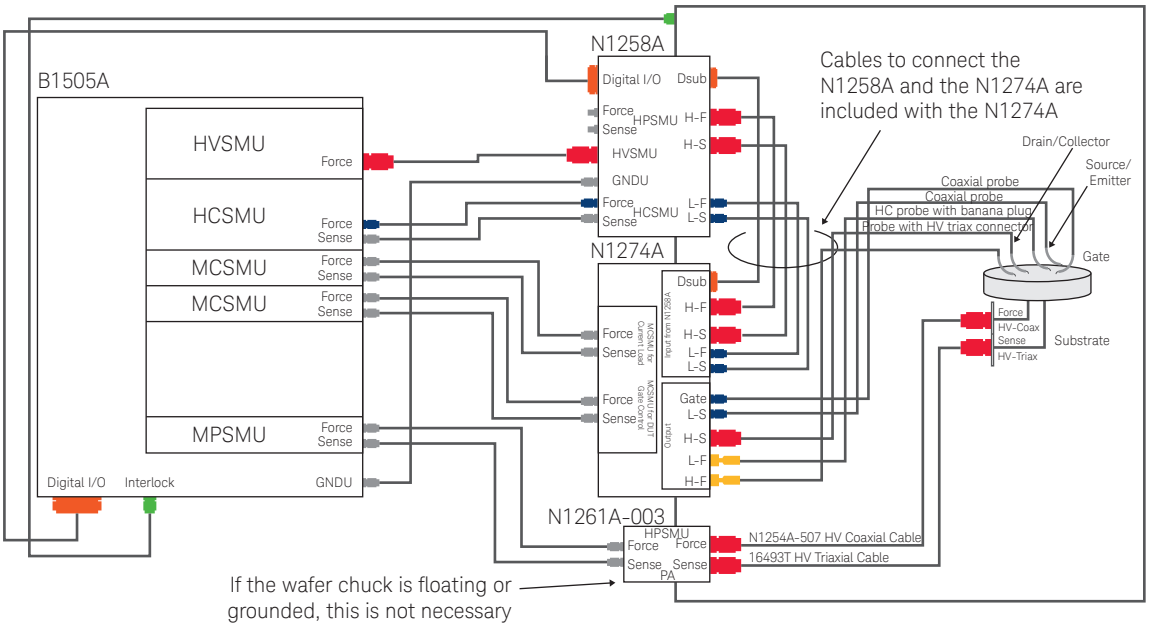


Figure 9.66. Connection scheme to measure a lateral power transistor using the N1274A gate charge adapter and the N1258A module selector.

Of course, in addition to correctly cabling everything together, you also need to correctly configure the modules within the EasyEXPERT “Configuration” menu. Configure the “N1258A/N1259A – 300 Module Selector” tab first, and make sure that the default output is set to whichever SMU is the high-voltage SMU. Next configure the “Gate Charge Adapter” by selecting “N1274A” as the adapter and selecting “IV” as the default path.

The cables supplied with the N1274A to connect to the wafer prober are not long (approximately 30-35 cm), so the N1258A and the N1274A need to be located physically close to the wafer chuck. The following picture shows this arrangement.

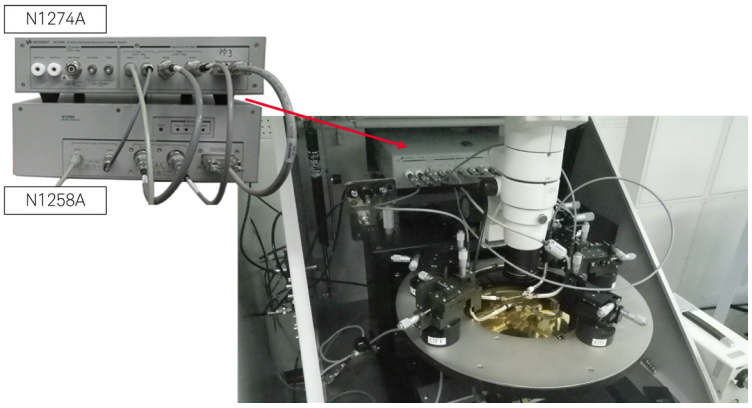


Figure 9.67. Picture of an on-wafer gate charge measurement setup using the N1258A and N1274A.

Since gate charge is a dynamic measurement, it is prone to oscillation. The following tips can help to prevent or eliminate oscillation:

1. Attach ferrite cores around the gate probe needle as close to the DUT as possible.
2. Attach a capacitor between the gate and the source as close to the DUT as possible.
3. Twist the high force and low force cables together.
4. If using a current load, select a FET of appropriate size (not too much larger than the DUT).
5. Make sure to perform Q_g adapter calibration with the capacitor (if used – see tip #2 above) in-place.

The following figure summarizes these best practices.

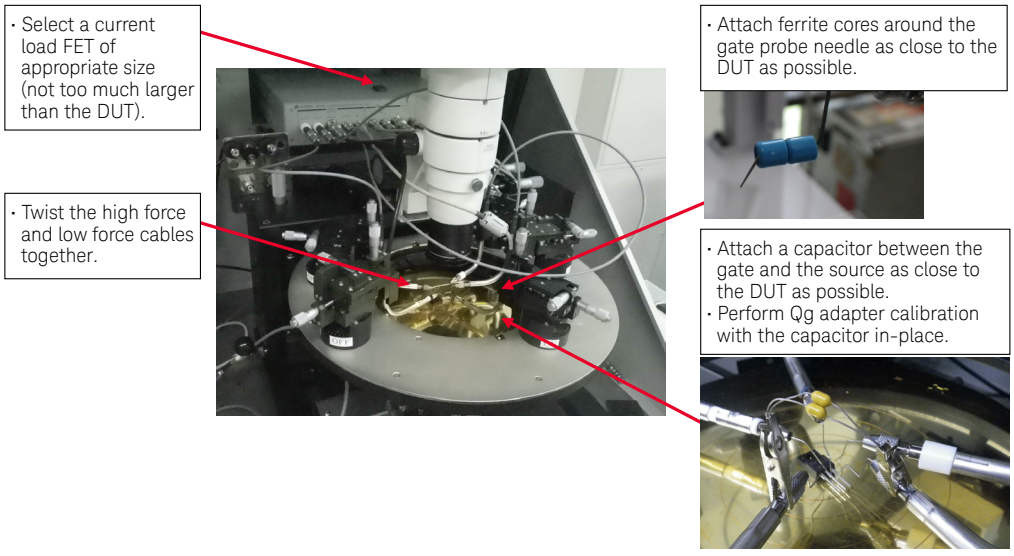


Figure 9.68. Tips to help prevent oscillation when making on-wafer gate charge measurements.

On-Wafer Gate Charge Measurement using the UHCU (N1265A)

When making gate charge measurements using the ultra-high current unit (N1265A), the N1275A gate charge adapter is required. The following figure shows an illustration of the N1275A.

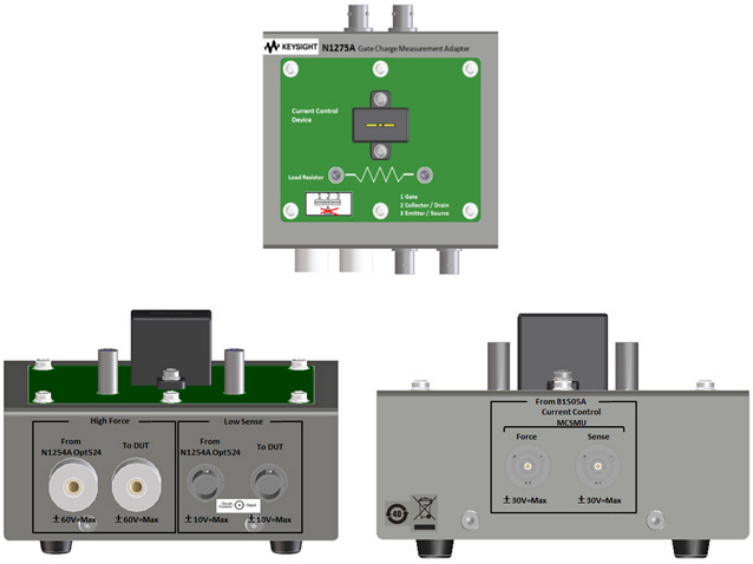


Figure 9.69. The N1275A (500 A/3 kV) gate charge adapter for wafer probing.

When using the N1265A, the N1254A-524 ultra high current prober system cable is also required. The connection scheme for measuring a vertical power device on-wafer is shown in the following figure.

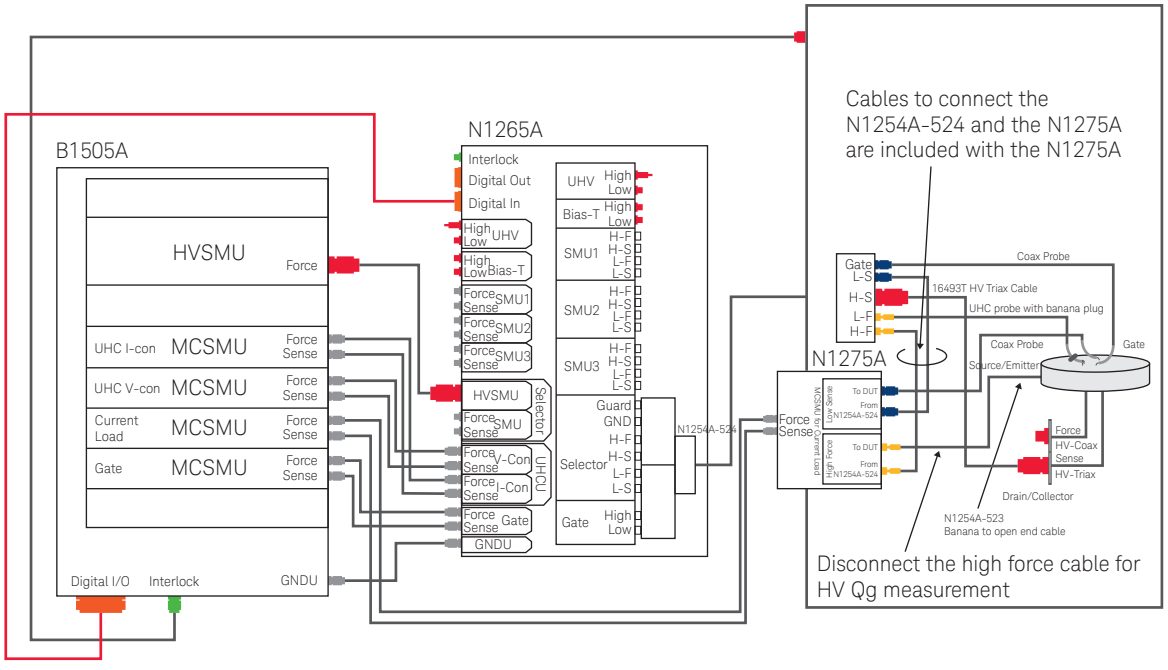


Figure 9.70. Connection scheme to measure a vertical power transistor using the N1275A gate charge adapter and the N1265A test fixture/current expander.

The connection scheme for measuring a lateral power device on-wafer is shown in the following figure.

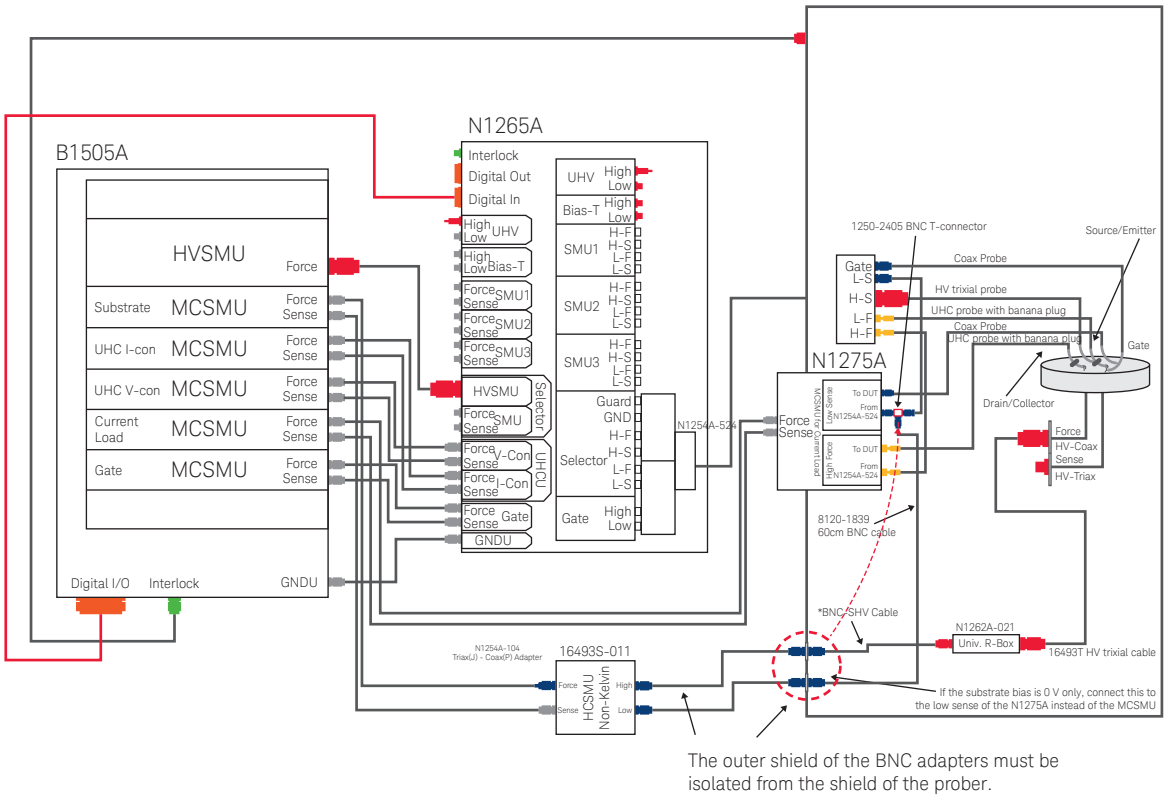


Figure 9.71. Connection scheme to measure a lateral power transistor using the N1275A gate charge adapter and the N1265A test fixture/ current expander.

Following the same procedure as for the N1274A, in the EasyEXPERT “Configuration” window, you need to configure the “UHC Expander / Fixture” and select the N1275A in the “Gate Charge Adapter” tab.

As in the case of the N1274A, the cables supplied with the N1275A to connect to the wafer prober are not long (approximately 30 cm). Therefore, the end of the N1254A-524 cable and the N1275A need to be located physically close to the wafer chuck. The following picture shows this arrangement.

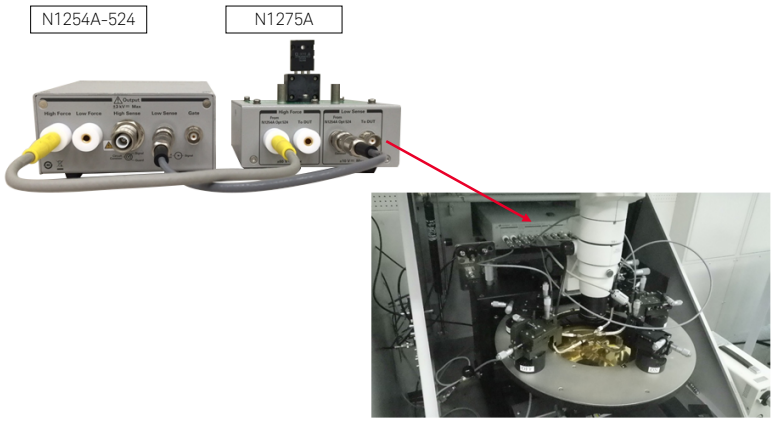


Figure 9.72. Picture of an on-wafer gate charge measurement setup using the N1254A-524 cable and N1275A.

Follow the same procedures to eliminate oscillation as for the case of the N1274A should they occur.

Automating On-Wafer Gate Charge Measurement

Although gate charge measurement can require a physical connection change when switching from high-current to high-voltage test mode, EasyEXPERT has solutions for automating this test on-wafer. There are gate charge application tests specifically designed to work with semiautomatic wafer probers. These application tests will first step across the wafer and perform all of the HC measurements. They will then prompt the user to change the connections for HV measurement, and after the user clicks on "OK, they will again step across the wafer and perform all of the HV measurements. When the process is completed, the on-wafer gate charge application tests will combine the HC and HV tests for each of the die locations tested and automatically save the data into separate test records. An illustration of this process is shown below.

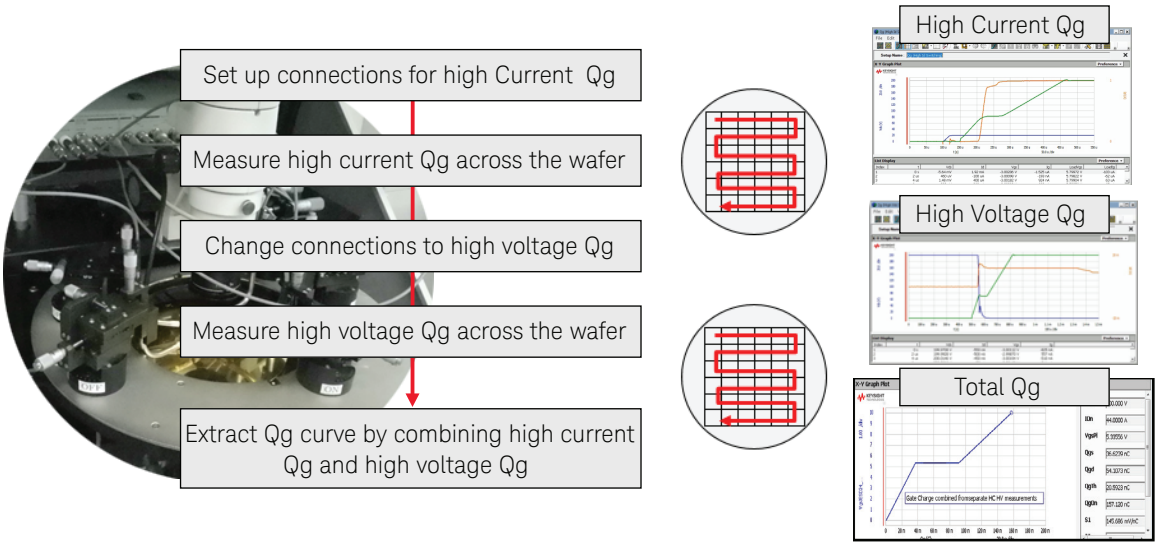


Figure 9.73 - Illustration of the procedure to automatically measure gate charge across an entire wafer using EasyEXPERT and the B1505A.

Once the measurements are complete, it is easy to select all of the individual test records and export them into a single file.

The B1505A User's Guide describes the on-wafer gate charge measurement procedure in detail, but the basics are not complicated. If you are using the N1274A gate charge adapter, then there are two special application tests that you need to use:

1. The "Q_g (High Id + High Vds+ JESD24-2)" application test, if you are using a current load transistor.
2. The "Q_g (R Load High Id + High Vds+ JESD24-2)" application test, if you are using a resistive load.

The procedure to perform either of these tests on-wafer is basically the same. First, fill in the appropriate device measurement parameters for the application test that you are using. Next, perform the following steps:

1. Click on the "Repeat" button to open the Repeat Measurement Setup dialogue box.
2. Select the correct start, iteration and final procedure prober drivers for the wafer prober that you are using. Also, check the box to have the device ID filled in automatically and make sure that the "Counter Reaching to" box in the "Repeat Stop Condition" is not selected.
3. Click on the "Run" button to start the wafer testing.

If you are using the N1275A gate charge adapter, then you will need to create a "My Favorites" group containing the following gate charge application tests for your particular device:

- Q_g (High Id + JESD24-2) or Q_g (R Load High Id + JESD24-2)
- Q_g (High Vds + JESD24-2)
- Q_g (JESD24-2 High Id + JESD24-2 High Vds)

You then need to fill in all of the appropriate device measurement parameters for these instances of the application tests and save the changes to the “My Favorites” group that you created. Next, you need to switch to EasyEXPERT “Quick Test” mode and make sure that the tests are set up to run as just shown (high current first, high voltage next and the combining application test last). All application tests in the “My Favorite group” should initially have their green execution flags deselected.

To perform automated on-wafer gate charge measurement, follow these procedures:

1. Select the green execution flag for the high Id test that you are using.
2. Click on the “Repeat” button to open the Repeat Measurement Setup dialogue box.
3. Select the correct start, iteration and final procedure prober drivers for the wafer prober that you are using. Also, check the box to have the device ID filled in automatically and make sure that the “Counter Reaching to” box in the “Repeat Stop Condition” is NOT selected.
4. Click on the “Run” button to start the wafer testing.

After the high current measurements completes, change the connections to perform high voltage measurement and perform the following procedures:

1. Deselect the green execution flag for the high Id test and select the green execution flag for both the high voltage and the combining application tests.
2. Click on the “Repeat” button to open the Repeat Measurement Setup dialogue box and follow the same procedures as for the high Id test.
3. Click on the “Run” button to start the wafer testing.

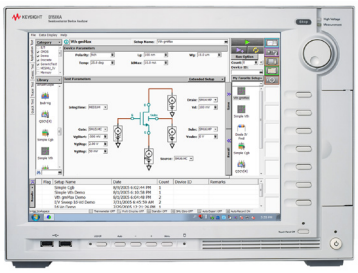
After the high voltage testing is completed, the combined results (derived gate charge curves) will be available in the EasyEXPERT test results window. The X-Y coordinate data (die index) for each location measured on the wafer will be automatically saved into the Device ID. The application tests will support on-wafer gate charge measurement for up to 1,000 die locations.

Appendix A Keysight Technologies' Parametric Measurement Solutions

The following parametric measurement instruments are available from Keysight Technologies. Only some of the key measurement capabilities are listed. For more information, please refer to the technical data sheet or web page for each instrument.

Semiconductor Parameter/Device Analyzers

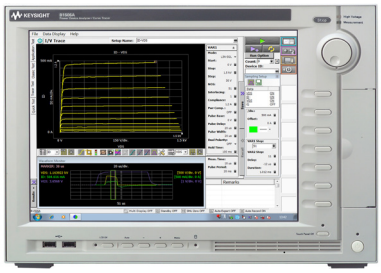
B1500A Semiconductor Device Analyzer



- Modular configuration: 10 slots
- MS Windows-based instrument with EasyEXPERT group+ software
- Furnished with > 300 application tests
- Current measurement resolution: 0.1 fA
- Voltage measurement resolution: 0.5 µV
- ±200 V and ±1 A HPSMU available
- 1 kHz to 5 MHz capacitance module available
- ±40 V PGU module available (with ALWG capability)
- WGFMU module has ALWG and supports fast and fast pulsed IV with 5 ns sampling rate

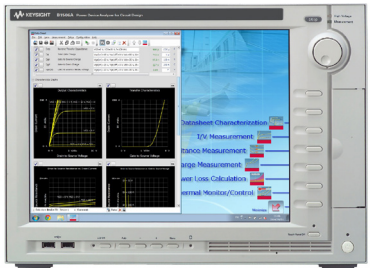
Power Device Analyzers/Curve Tracers

B1505A Power Device Analyzer / Curve Tracer



- Modular configuration: 10 slots
- MS Windows-based instrument with EasyEXPERT group+ software
- Curve tracer mode with knob sweep support
- Current measurement resolution: 10 fA
- Voltage measurement resolution: 0.2 µV
- ±1500 A current expander available (N1265A)
- ±3 kV high voltage SMU available
- ±10 kV ultra-high voltage unit available
- 1 kHz to 5 MHz CV measurements at 3 kV DC bias
- Gate charge measurement up to 3 kV and 1100 A

B1506A Power Device Analyzer for Circuit Design



- MS Windows-based instrument with Easy Test Navigator and EasyEXPERT group+ software
- Current measurement resolution: 10 fA
- Voltage measurement resolution: 0.2 µV
- ±20 A, ±500 A and ±1500 A options
- ±3 kV high voltage measurements
- 1 kHz to 5 MHz CV measurements at 3 kV DC bias
- Gate charge measurement up to 3 kV and 1100 A

Modular Source/Monitor Units

E5260A 8-Slot High Speed Measurement Mainframe

- Modular configuration: 8 slots
- Basic user interface for spot measurements and program debug
- Code compatible with the 4142B
- SMUs measure several times faster than 4142B SMUs
- Output up to 4 A (4 x HPSMUs) at once
- 8 programmable trigger in/out lines in addition to BNC trigger in/out connectors
- 4.0 A ground unit

E5262A 2-Channel (Medium Power, Medium Power) Source Monitor Unit

- Fixed configuration: 2 x MPSMUs
- Basic user interface for spot measurements and program debug
- Code compatible with the 4142B
- SMUs measure several times faster than 4142B SMUs
- 8 programmable trigger in/out lines in addition to BNC trigger in/out connectors
- 2.2 A ground unit

E5263A 2-Channel (Medium Power, High Power) Source Monitor Unit

- Fixed configuration: 1 x MPSMU, 1 x HPSMU
- Basic user interface for spot measurements and program debug
- Code compatible with the 4142B
- SMUs measure several times faster than 4142B SMUs
- 8 programmable trigger in/out lines in addition to BNC trigger in/out connectors
- 2.2 A ground unit

E5270B 8-Slot Precision Measurement Mainframe

- Modular configuration: 8 slots
- Basic user interface for spot measurements and program debug
- Code compatible with the 4142B
- Supports HRSMU with 1 fA and 0.5 µV measurement resolution
- HRSMU supports ASU for 0.1 fA measurement resolution
- Output up to 4 A (4 x HPSMUs) at once
- 8 programmable trigger in/out lines in addition to BNC trigger in/out connectors
- 4.0 A ground unit
- TIS commands supported for both Basic and C so that algorithms can be transported to the 4070 or 4080

Benchtop Source/Measure Units

B2901A Precision Source Measure Unit - 100 fA 1ch
B2902A Precision Source Measure Unit – 100 fA 2ch

- Single and dual SMU channel instruments
- 4.3" LCD-based GUI display
- Set/measure range: ±210 V, ±3 A (DC), ±10.5 A (Pulsed)
- Setting resolution: 1 pA & 1 µV
- Measurement resolution: 100 fA & 100 nV
- Supports 4-wire (Kelvin) measurements
- USB port for storing data, graphs and measurement setups via Flash drive
- Supports web-browser control (fully compliant with LXI class C specifications)
- Fully controllable from a PC via GPIB, LAN or USB ports

B2911A Precision Source Measure Unit – 10 fA 1ch
B2912A Precision Source Measure Unit – 10 fA 2ch

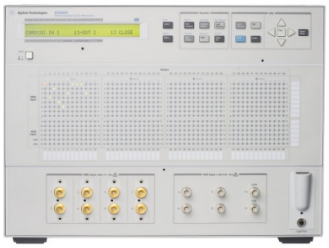
- Single and dual SMU channel instruments
- 4.3" LCD-based GUI display
- Set/measure range: ±210 V, ±3 A (DC), ±10.5 A (Pulsed)
- Setting resolution: 10 fA & 100 nV
- Measurement resolution: 10 fA & 100 nV
- Supports 4-wire (Kelvin) measurements
- USB port for storing data, graphs and measurement setups via Flash drive
- Supports web-browser control (fully compliant with LXI class C specifications)
- Fully controllable from a PC via GPIB, LAN or USB ports

Switching Matrices

E5250A Low Leakage Switch

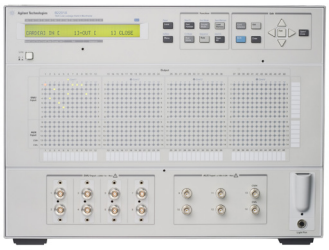
- Triaxial inputs: 6
- Coaxial inputs: 4
- Internal paths: 6
- Low-current paths: 2
- Maximum output ports: 48
- Effective current measurement resolution: 20 fA
- Bandwidth: 10 MHz

B2200A Femtoamp Low Leakage Switch



- Triaxial inputs: 8
- Coaxial inputs: 6
- Internal paths: 14
- Low-current paths: 8
- Maximum output ports: 48
- Effective current measurement resolution: 1 fA
- Bandwidth: 30 MHz

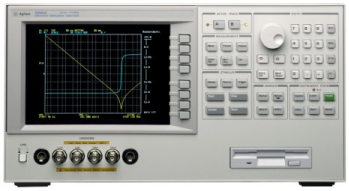
B2201A 14-channel Low Leakage Switch



- Triaxial inputs: 8
- Coaxial inputs: 6
- Internal paths: 14
- Low-current paths: 8
- Maximum output ports: 48
- Effective current measurement resolution: 10 fA
- Bandwidth: 30 MHz

Impedance Analyzers

4294A Precision Impedance Analyzer (Note: Discontinued product)



- Frequency range: 40 Hz to 110 MHz
- Test voltage signal range: 5 mV to 1 V (rms)
- Test current signal range: 200 µA to 20 mA (rms)
- DC voltage bias range: 0 V to ±40 V
- DC current bias range: 0 mA to ±100 mA

E4990A Impedance Analyzer

- Frequency range: 20 Hz to 120 MHz
- Test voltage signal range: 5 mV to 1 V (rms)
- Test current signal range: 200 µA to 20 mA (rms)
- DC voltage bias range: 0 V to ±40 V
- DC current bias range: 0 mA to ±100 mA

Glossary

Accuracy – The degree of conformity of a measured or calculated quantity to its actual (true) value.

ALWG – The abbreviation of Arbitrary Linear Waveform Generation. ALWG is a waveform generation technique in which a waveform is synthesized from a concatenated set of user-defined line segments.

Analog-to-Digital Converter (ADC) – A circuit used to convert an analog signal into the digital equivalent of that signal.

Atto-sense and Switch Unit (ASU) – A special module that works with a high-resolution SMU to obtain current measurement resolution of 0.1 fA. The ASU also has a switching capability that allows it to switch between its SMU connection and other instrumentation connected to it via coaxial inputs.

Auto Ranging – An instrument measurement feature whereby the measurement circuitry will automatically switch between measurement ranges until the optimal measurement range for the quantity under test is reached.

Bipolar – An alternative transistor technology to MOS, bipolar transistors differ from MOS transistors in that they are current-controlled devices and they rely upon minority carriers for their operation. Bipolar transistors are most often used in RF and high-power applications.

BNC Cable – See coaxial cable.

Capacitance – The amount of electric charge that can be stored for a given electric potential. The most common capacitor type is the two-plate capacitor, where the formula for capacitance is given by C = Q/V.

CDA – The abbreviation of Clean Dry Air. This is air that has been filtered and dehumidified to remove particulates and moisture so that it can be safely used in pneumatic devices and systems.

CMU – The abbreviation of Capacitance Measurement Unit. This is a very generic term covering any sort of hardware that can perform capacitance measurement. It could mean a benchtop instrument such as the E4980A or 4294A, or a plug-in module such as the B1500A's MFCMU.

Coaxial Cable – Also known as a "BNC" cable. A coaxial cable consists of an inner conductor surrounded by a tubular insulating layer typically made from a flexible material with a high dielectric constant, all of which is then surrounded by another conductive layer, and then finally covered again with a thin insulating layer on the outside.

Common Mode Rejection Ratio (CMRR) – The ability of a circuit to reject input signals common to both input leads. CMRR is measured in positive decibels, and is defined as 20 times the log in base 10 of the ratio of the differential gain divided by the common-mode gain.

Common Mode Voltage – The voltage difference between the internal reference of an instrument and true earth ground.

Debye Length – In semiconductor physics the solution of Poisson's equation in the presence of free charge always leads to a characteristic Debye length, which provides a qualitative measure of the extent of the space charge region.

Depletion Approximation – In semiconductor physics, the assumption that a semiconductor junction can be divided into distinct regions which are either completely neutral or completely depleted of mobile carriers

Dew Point – The dew point is the temperature at which a given volume of air must be cooled (at constant barometric pressure) for water vapor to condense into water.

Drift – A relatively slow change in the measurement reading of an instrument over time that occurs even when the measurement environment and the quantity being measured are stable.

DUT – The abbreviation of Device Under Test. DUT is a very general term that can be used for simple components such as resistors, capacitors and transistors up to and including entire integrated circuits (ICs).

Faraday Cage – An enclosure formed of conductive material (or a mesh of such material) to shield the inside of the enclosure from the effects of the surrounding electromagnetic fields.

Fixed Ranging – An instrument measurement feature in which the measurement range (as the name implies) never changes from the specified value. Note: If the quantity being measured exceeds the maximum value supported by the selected measurement range then the measurement will fail and an error message will be generated.

Flat Band Capacitance – The capacitance of a MOS capacitor under flat band voltage conditions.

Flat Band Voltage – A non-thermal equilibrium condition in a MOS capacitor in which the voltage applied at the gate is set to a value that exactly compensates the difference in the work functions of the gate material and the semiconductor. In this condition, the energy bands in the silicon are flat in the surface regions as well as in the bulk, hence the name flat band voltage.

Ground Loop – An undesirable measurement condition in which two or more conductive planes or surfaces are connected to ground through multiple points, thereby allowing current to flow between them due to common mode voltage differences.

Ground Unit – A special type of SMU that has no measurement capabilities and is always set at zero volts.

Guarding – The technique of surrounding a signal line with an actively driven conductor maintained at the same voltage potential as that of the signal to eliminate leakage currents.

Hot Carrier Injection (HCI) – A phenomenon (mostly impacting MOSFETs) where an electron or hole in the conducting channel gains sufficient kinetic energy (i.e. becomes “hot” enough) to overcome potential barriers and inject itself into the gate dielectric. Since this can impact the transistor threshold and switching characteristic, it is an important reliability concern.

Impedance – The total opposition a device or circuit offers to the flow of an alternating current (AC) at a given frequency. It is represented as a complex quantity that can be graphically represented on a vector plane where the x-axis represents the real part (resistance, R) and the y-axis represents the imaginary part (reactance, X).

Input Impedance – The effective impedance seen at the inputs to an electronic circuit.

Kelvin Measurement – Also known as a “4-wire” measurement, the Kelvin measurement technique uses four terminals to make a resistor measurement: two to force current and two to measure voltage. By using this technique, the effects of cable resistance on the measurement can be eliminated.

Limited Auto Ranging – An instrument measurement feature similar to auto ranging, except that the measurement circuitry will stop going into any lower measurement ranges once the specified range limit has been reached.

Measurement Range – The instrument measurement state (usually specified in decade increments) in which the measurement circuitry is internally configured to produce the optimum measurement result for a given range of values.

MOS – The abbreviation of Metal Oxide Semiconductor. The most common type of transistor structure, MOS transistors consist of some sort of gate material (metal or polysilicon) separated from a conducting (semiconductor) channel by an insulating material (usually silicon dioxide). MOS transistors are either n-channel type (NMOS) or p-channel type (PMOS), and circuits containing both NMOS and PMOS transistors are known as complimentary MOS (CMOS) technologies.

Multi-Frequency Capacitance Measurement Unit (MFCMU) – A module for the B1500A and B1505A that can measure capacitance and that has a frequency range from 1 kHz to 5 MHz.

National Institute of Standards and Technology (NIST) – A non-regulatory agency of the United States Department of Commerce, NIST supplies standard reference materials of the highest quality and metrological value. All Keysight instruments are calibrated to NIST traceable standards.

Noise (Electronic) – The random variations in current and voltage that occur in an electronic circuit. There are many different types of noise that can occur in electronic circuits.

PLC – The abbreviation of Power Line Cycle. As the name implies, this is the length of time it takes the AC power supply to complete one cycle (20 ms for 50 Hz systems and 16.7 ms for 60 Hz systems).

Pulse Generator Unit (PGU) – An instrument or instrument module capable of generating extremely fast voltage and/or current pulses (in the nanosecond range or below).

Output Impedance – The effective impedance seen at the outputs of an electronic circuit.

Repeatability – The degree to which repeated measurements or calculations show the same or similar results.

Resolution – The lowest resolvable quantity of data that an instrument can accurately measure.

SCFM – The abbreviation of Standard Cubic Feet per Minute. In the English system, this is the volumetric flow rate of a gas corrected to standardized conditions of temperature, pressure and relative humidity. Note: Be careful as the “standardized” conditions can vary between definitions. In the SI metric system, the term Normal Cubic Meter (Nm3) is used.

SHV – The abbreviation of Safe High Voltage. The SHV is a type of connector used with BNC cables for safe connection to high voltage sources. The connector uses a bayonet mount similar to those of BNC connectors, but it is easily distinguished due to its very thick and protruding insulator. The insulation makes SHV connectors safer for handling high voltages by preventing accidental contact with the live conductor in an unmated connector or plug. The connector is also designed such that when it is being disconnected from a plug, the high voltage contact is broken before the ground contact, thereby preventing accidental shocks.

Single Point Grounding – The preferred technique to avoid ground loops, whereby all ground planes are each connected to a common ground through a single point.

SMU CMU Unify Unit (SCUU) – A switching module designed to work with the B1500A’s MFCMU and two of either the high-resolution or medium power SMUs. The SCUU form-factor is designed to accept the four BNC outputs of the MFCMU and the two pairs of Kelvin triaxial outputs from the two HRSMUs and/or MPSMUs. The SCUU provides the capability to switch between the MFCMU and the HRSMU/MPSMU pair automatically, and its outputs are two pairs of triaxial outputs (force and sense).

Source/Monitor Unit (SMU) – Also sometimes designated as "Source/Measure Unit." The SMU is a measurement module that can force voltage or current and simultaneously measure voltage and/or current. Although often thought of as a DC measurement unit, the SMU does have some time sampling and pulsing capability.

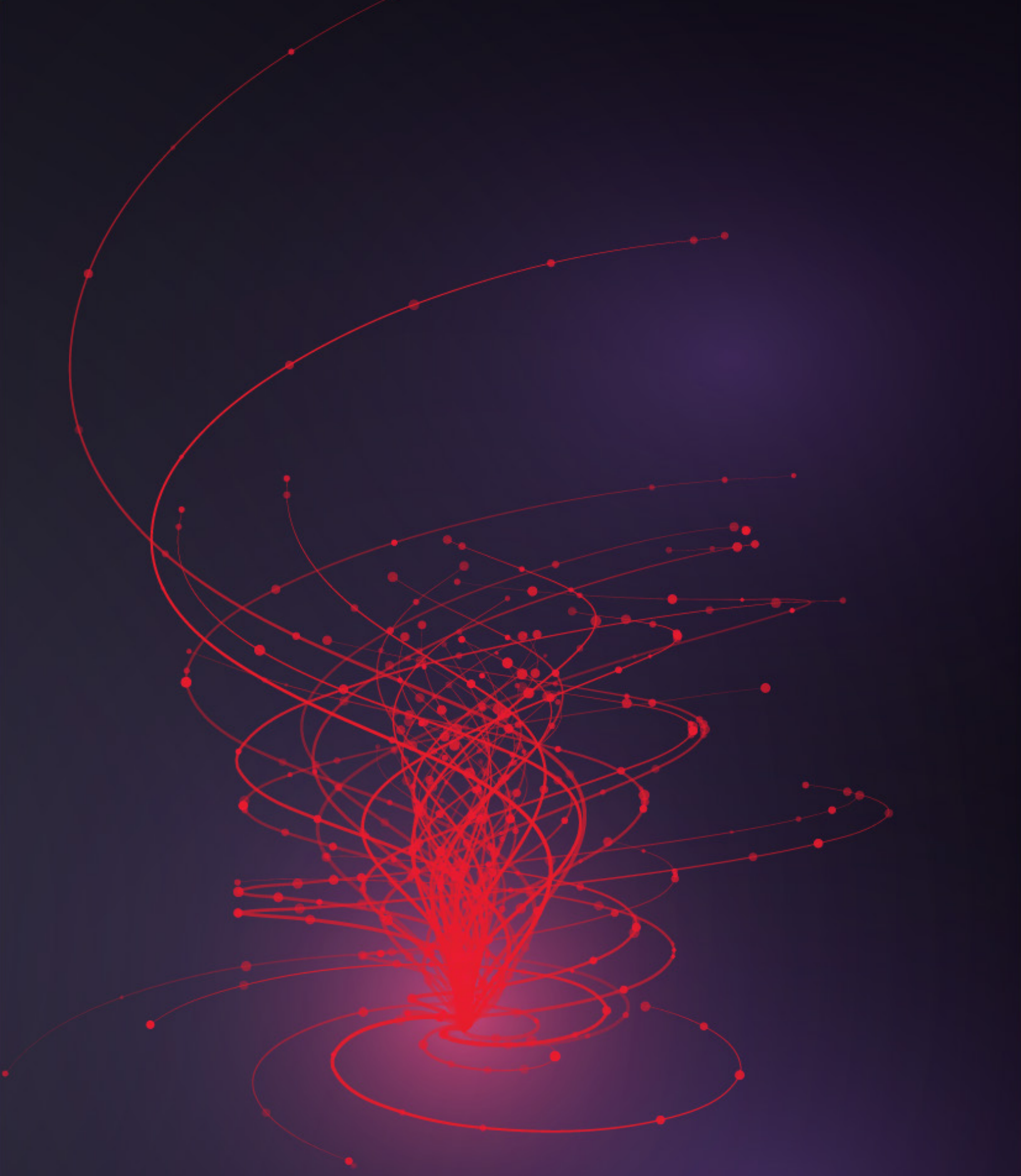
Time Dependent Dielectric Breakdown (TDDB) – A failure mechanism in MOSFETs in which the insulating gate dielectric degrades to the point where it no longer acts as an insulator, but instead starts conducting current.

Triaxial Cable – A cable designed for low (<1 nA) current measurements. The triaxial cable consists of an inner conductor surrounded by a tubular insulating layer typically made from a flexible material with a high dielectric constant, which is surrounded by a conducting guard layer that is actively driven to be at the same potential as the center conductor. The guard layer is then surrounded by another conductive layer (the shield), and covered again with a thin insulating layer on the outside.

Waveform Generator/Fast Measurement Unit (WGFMU) – The WGFMU is a measurement module that can generate arbitrary voltage waveforms (with 10 ns programmable resolution) and measure both current and voltage extremely fast (5 ns sampling rates).

About the Author

Alan Wadsworth is the North and South American Business Development Manager for Keysight's semiconductor and power products. Alan holds bachelor's and master's degrees in electrical engineering from the Massachusetts Institute of Technology and an MBA from Santa Clara University. Alan has over 30 years of experience in the semiconductor industry in both design and test.



KEYSIGHT TECHNOLOGIES

Published in USA, October 27, 2020
5992-2508EN
www.keysight.com