Improving Test Throughput with Addressable Arrays and the Agilent N9201A

Agilent N9201A Array Structure Parametric Test Option for the Agilent 4070 and 4080 Series Parametric Testers

Application Note N9201A-1

Introduction

A major issue facing advanced semiconductor manufacturing processes is the increased time required to pass through the yield ramp up phase and into the mature process phase (where the process yields are acceptable). In addition, for many advanced processes, even after they are released into the high volume manufacturing phase, the yields can fluctuate below acceptable levels for a variety of reasons (refer to Figure 1). To both ramp up the yield quickly and stabilize the yield over time, many more device parameters must be measured in a shorter time period than for older, less advanced wafer manufacturing processes. Taking these massive volumes of test data quickly is necessary in order to make timely corrections to the process recipe and return the yield to an acceptable level. Unfortunately, conventional parametric testing methodologies and equipment cannot keep up with these needs. This application note will explain a new solution employing test structure designs utilizing addressable arrays and the Agilent N9201A Array Structure Parametric Test option that can solve these previously intractable challenges.

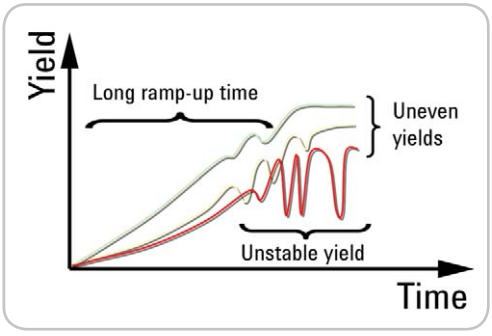

Figure 1. The yield ramp-up and process stabilization profiles across several generations of advanced processes.

Issues and requirements for quick yield ramp up and process stabilization

Advanced process evaluation faces several challenges:

1.Test structure scaling cannot track process and scribe line scaling

It is common practice to place test structures used for process monitoring and wafer acceptance testing in the wafer scribe line, since this does not use up any valuable area on the product die. However, as lithographies continue to shrink the die size also becomes smaller, which means that the scribe line width has to decrease proportionally. Unfortunately, the probe pads used to contact the test structures cannot easily scale, since their minimum size is determined by the diameter of the smallest supported probe tip (see Figure 2). Since the size of the test structures is typically much smaller than that of the probe pads, one solution is to have multiple structures share test pads and then use some sort of decoding scheme to connect to each individual test structure.

2. More test structures need to be tested in less test time

Assuming that you can overcome the test structure area limitations, you are then faced with the additional challenge of testing many more test structures in less test time. As lithographies shrink and process complexity grows, the sheer volume of test data required to guarantee process integrity (which directly equates to an increase in the number and types of test structures) increases. Conventional parametric testing methodologies cannot keep up with these testing needs and provide the timely feedback necessary to correct deviations found during process monitoring or wafer acceptance testing. This testing bottleneck impacts the yield ramp up, which in turn has a strong impact on profitability.

3. Process engineers need to be able to distinguish between systematic and random process variations

To achieve a satisfactory yield ramp up, process engineers must be able to distinguish random defects (which are statistical and inevitable) from systematic defects (which are deterministic and correctable). To differentiate between these two types of defects, it may be necessary to place and test one thousand or more devices in a single location on the wafer. Obviously, this requires a different sort of test methodology than that required by traditional single-device test structures. In addition, the size of these large test structures means that they cannot be placed in the scribe line, but must be placed on the wafer in special drop-in die.

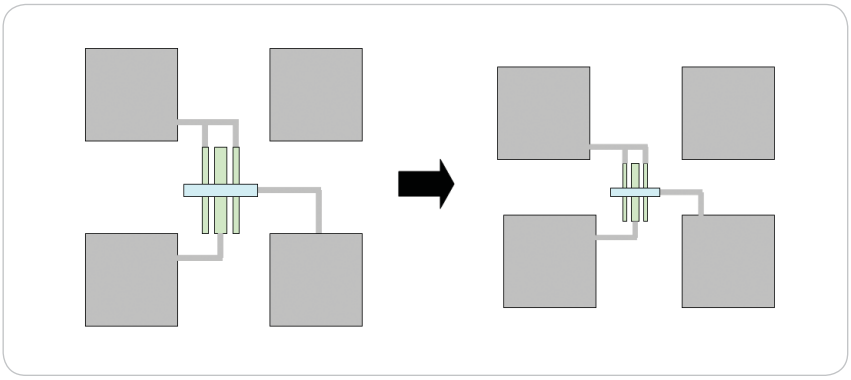

Figure 2. Probe pads cannot scale with process design rule and scribe line scaling, which limits the number of devices that can be placed into the scribe line using conventional test structure arrangements.

Addressable arrays provide a high density and high throughput test solution

Addressable array overview

Figure 3 shows the organization of independent test structures and Figure 4 shows the organization of passive array test structures. These can be compared to the block diagram of a typical addressable array structure as shown in Figure 5. The addressable array contains multiple test structures, which are accessed through decoding and switching circuitry. In contrast to an independent or passive array test structure where each device terminal in the array is connected to a probe pad, the devices in an addressable array are accessed through measurement ports. Address input signals determine which particular device connected to a measurement port is accessed at any given time. Addressable arrays can contain multiple measurement ports, so that several DUTs can be measured in parallel at each address.

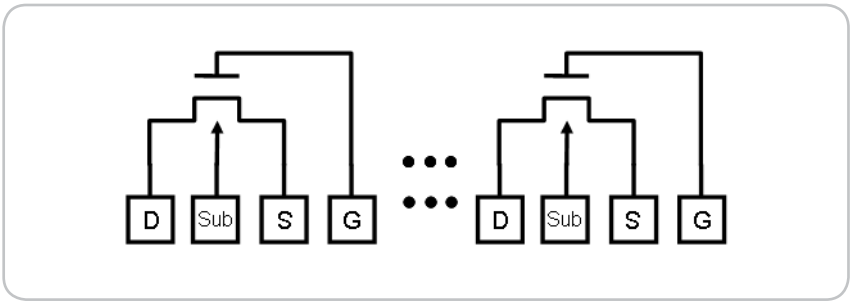

Figure 3. Independent test structure organization.

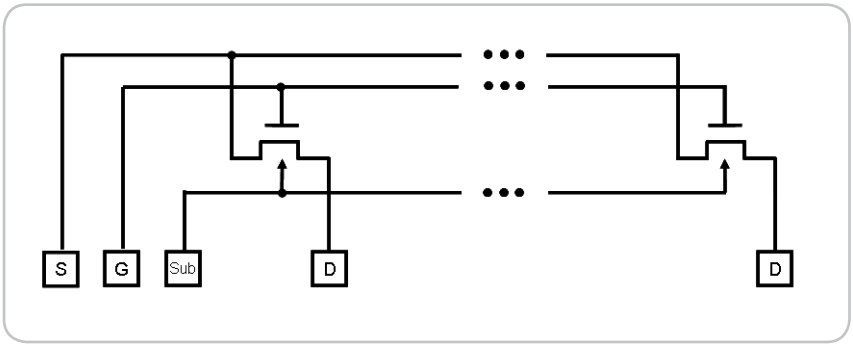

Figure 4. Passive array test structure organization.

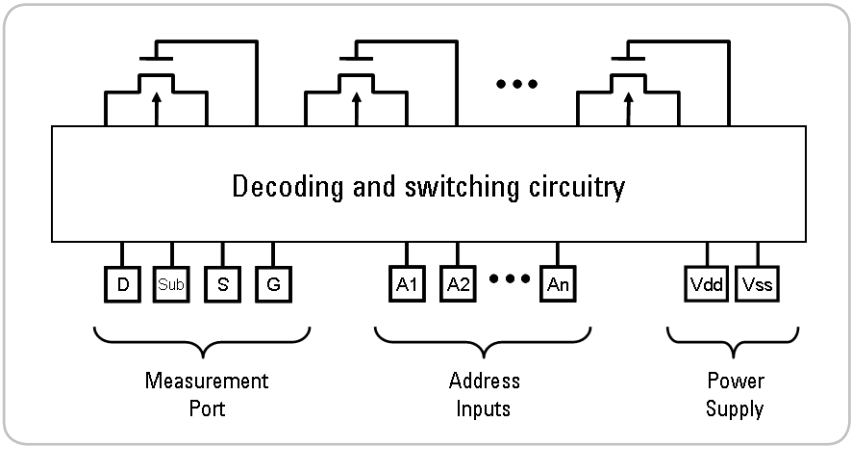

Figure 5. Addressable array example.

Advantages of addressable arrays over conventional test structures

1. More test devices per unit area

Addressable arrays can have a much higher test device to test pad ratio than can conventional test structures that require a separate test pad for every device input. As previously explained, test pads are the main constraint on efficient test structure scaling for advanced processes. Addressable arrays virtually eliminate this restriction and can easily support test structures containing 1000 or more individual devices (as is required for the yield ramp up of advanced processes).

2. Access multiple devices in a single touch-down

Conventional test structures can be arranged into modules and several of these modules can be probed in a single touch-down using a probe card and external switching matrix. However, practical limitations on probe card size and external switching matrix outputs usually restrict the number of conventional devices that can be tested in this fashion to six to eight. In contrast, an addressable array can test hundreds or thousands of devices in a single touch-down. Since the typical time to move from one wafer location to another using a wafer prober is between 200 to 500 ms, addressable arrays can test more devices in a given time period than can conventional test structures.

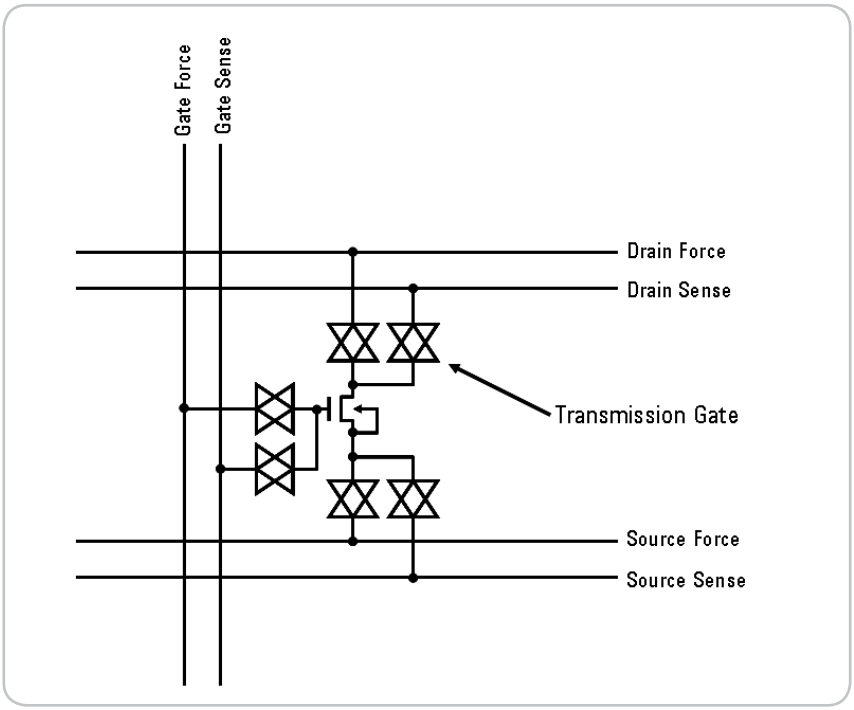

Figure 6. Typical arrangement to perform a Kelvin measurement on a MOSFET device in an addressable array.

3. Easier parallel test capability

A passive array (refer back to Figure 4) can support more devices than can an independent test structure design. However, simply sharing a probe pad (such as in the case of a common source or substrate connection for a group of MOSFETs) does not necessarily make performing parallel test easier. A common test pad places many restrictions on the ability to test multiple devices in parallel with different test conditions. In contrast, an addressable array can support independent measurement ports for each test group. This makes it much easier to test multiple devices in parallel.

4. Improved voltage forcing accuracy

Since the switching circuit in an addressable array has some series resistance associated with it (on the order of 100 to 1000 Ω), the voltage at the DUT may not be the same as the voltage applied at the pads. This is especially true if large currents are applied to the DUT. To remove this potential error, addressable arrays typically place sense signal lines between the device terminals in the array and the measurement ports to support Kelvin measurements (see Figures 6). The net result is actually a more accurate voltage applied at the actual DUT than is typically achieved when using conventional test structures, since contact and lead resistance have been eliminated. Please see Appendix A for more information.

N9201A ASPT option for addressable array testing

A picture of the N9201A array structure parametric test option is shown in Figure 7. The cabinet containing the measurement hardware connects to the Agilent 4070 or 4080 series parametric test system through the extended path ports, AUX ports or HF ports in the test head. The N9201A can support up to 40 SMUs and up to 32 address signal channels, with full synchronization between the SMUs and address channels. The block diagram is shown in Figure 8.

Figure 7. The N9201A array structure parametric test option for the 4070 and 4080 series of parametric testers.

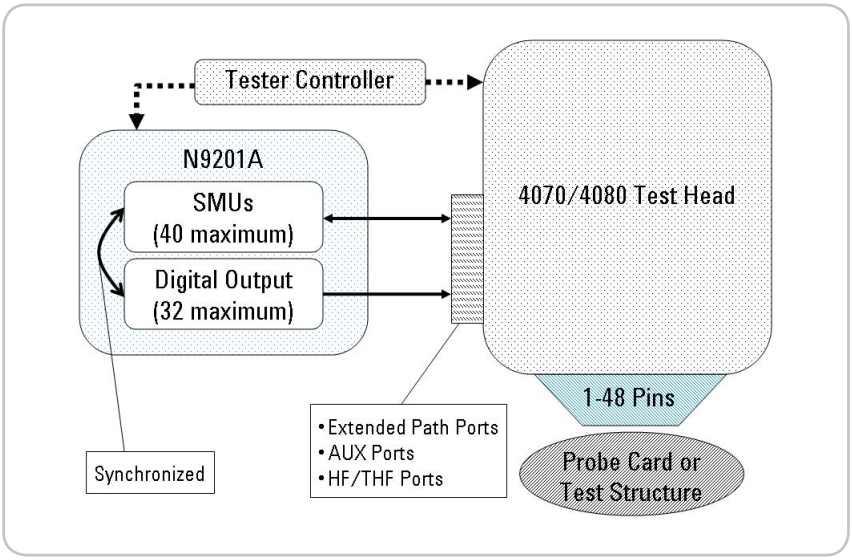

Figure 8. The N9201A option can supply up to 32 digital outputs and 40 SMUs, and it makes sure that the addresses are synchronized with the SMU measurements.

Key features and benefits of the N9201A solution

1. Synchronization between the address channels and the SMUs

To create address signals on the 4070 or 4080 Series systems without the N9201A, you would need to use the SMUs and GNDU or an external instrument such as a word generator. When using the SMUs and GNDU to generate the address signals, you have to sacrifice these resources to generate the logic high and low levels. In the case of the external word generator, because there is no automatic synchronization with the 4070/4080 test head CPU you have to be careful not to start making a measurement before the address signals have stabilized.

The N9201A option offers the immediate benefit of providing a separate address generation capability that does not waste any SMU or GNDU resources. Moreover, the N9201A system software ensures that the N9201A address output ports are fully synchronized with the measurement SMUs. The user only has to specify the address and the measurement conditions. The N9201A software will make sure that the address signals are correct and stable before starting to make measurements using the SMUs. The N9201A software will also make sure that the measurement is complete before applying the next address. Please refer to Figure 9.

2. Coordination of program memory/ Data Buffer and parallel test

The N9201A hardware consists of a series of mainframes (up to five) each of which can contain up to eight SMUs. Obviously, the number of mainframes in a given N9201A system cabinet depends upon the total number of SMUs in the system (with the maximum number of SMUs being given by: 5 mainframes x 8 SMUs per mainframe = 40 SMUs maximum). Each mainframe also has the ability to support a feature known as program memory.

Program memory allows measurement command sequences to be stored into machine memory so that the entire command sequence can later be executed without having to send the commands again over the GPIB bus. The benefit is similar to that achieved using a subroutine in a conventional program. The N9201A solution supports simple TIS commands that allow you to automatically load GPIB instructions into the program memory of the mainframes, thus freeing you from having to worry about managing the program memory for all of the mainframes separately.

The measured data collected by the instructions contained within the program memory is stored in the data buffer. This data can be read out after several blocks of the program have been executed or even while a program block is being executed, which makes the data upload to the tester controller very efficient. In addition, using the program memory and data buffer features different mainframes can separately execute different sets of instructions. By combining the program memory and data buffer features with parallel testing, the N9201A option enables you to achieve throughput rates higher than those that could be achieved via parallel testing alone. Please refer to Figure 10.

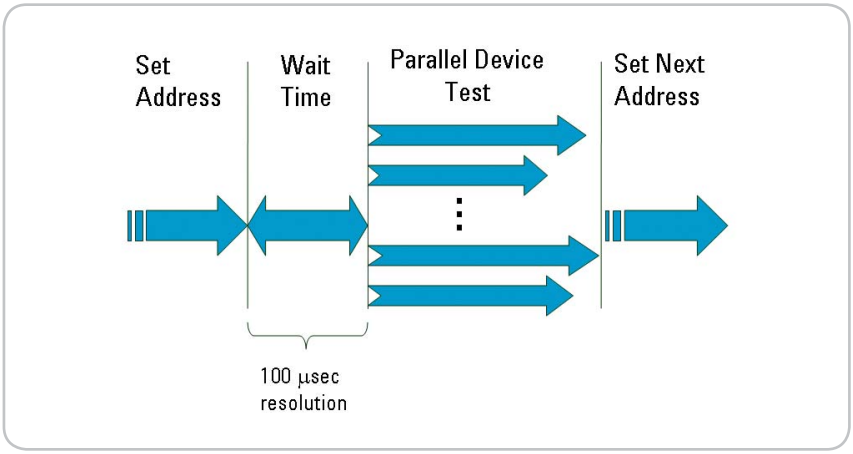

Figure 9. The N9201A option synchronizes the address outputs with the parallel measurements performed by the SMUs. The wait time between address set and measurement is programmable with 100 µs resolution.

3. Good correlation with conventional test structure measurements

The N9201A has the ability to monitor the sense signal lines in an addressable array if they are present. As previously mentioned using sense signal lines in an addressable array greatly enhances voltage forcing and measurement accuracy and improves measurement correlation with conventional test structures.

Using the Kelvin connections on the N9201A SMUs, you can directly connect the SMU force and sense lines to the force and sense signal lines (respectively) coming from an addressable array. This measurement method works well in-general and allows you to measure very small voltages very quickly. The N9201A also supports an alternative technique that uses the force lines from two separate SMUs. In this technique, the second SMU's force line is used to adjust the voltage at the DUT so as to correct for the effects of the series resistance experienced the first SMU that is actually making the measurement. The N9201A has special TIS commands that support this measurement technique, and allow the system to adjust the voltage at the device DUT very quickly to obtain an accurate and fast measurement. Each of these two measurement methods has its pros and cons, and Agilent applications specialists are available to give you advice as to the best measurement technique for a given situation.

Please refer to Appendix A for more details.

4. Compatible with conventional test structure measurement

Even when addressable arrays are in use, it may also be necessary to test conventional test structures (such as capacitors). Thus, it is important to be able to test both types of test structures using the same test system. Because the N9201A is an option for the 4070 and 4080 series of parametric testers, it is easy to test both types of test structures on a single tester.

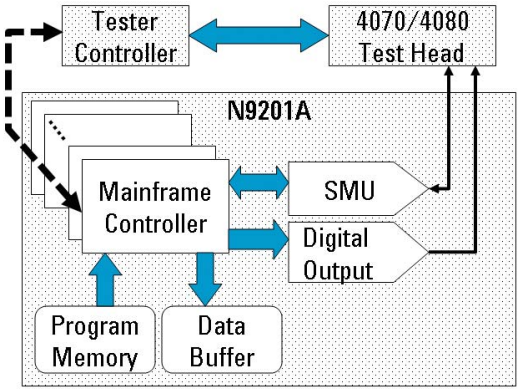

Figure 10. The N9201A option has TIS commands that allow you to load the program memory of the SMU mainframes. Combining program memory with parallel test yields better throughput than that obtainable using parallel test alone.

N9201A addressable array testing examples

The following examples illustrate the importance that proper array design plays in optimizing measurement throughput when using the N9201 option. Three examples of addressable array designs that were actually used with the N9201A will be discussed, and the measurement results obtained in each case will compared.

Array-A: A non-optimized addressable array design

Overview of Array-A

This addressable array example is a scribe line test structure for process monitoring and wafer acceptance test. It has six address lines, giving it an address depth of 64 (although only 60 of these are used). It also has two parallel measurement ports consisting of three pairs of force and sense signal lines for making Kelvin measurements on MOSFET drain, gate and source terminals (the drain and source force/sense pairs are also connected to the resistor test structures). As shown in Figure 11, one of these ports is connected to 60 MOSFET test structures, while the other port is connected to a mix of 20 MOSFET and 40 resistor test structures. Therefore, in a 25 pad test structure more than 120 test structures can be placed.

Throughput improvement using Array-A with the N9201A

Even though resistor spot measurements take much less time than MOSFET sweep, spot, and search measurements, the Array-A design cannot achieve optimum test time reductions using the N9201A solution. This is because even after a resistor spot measurement completes on port 2, the address cannot be changed until the (much longer) measurement on the corresponding MOSFET connected to port 1 completes. The net result is only about a 50% improvement in measurement throughput as shown in Figure 12.

Data correlation with conventional test structures

While the throughput improvement is only modest, the correlation of the measured data with that taken on identical test structures arranged in conventional test arrays is generally excellent.

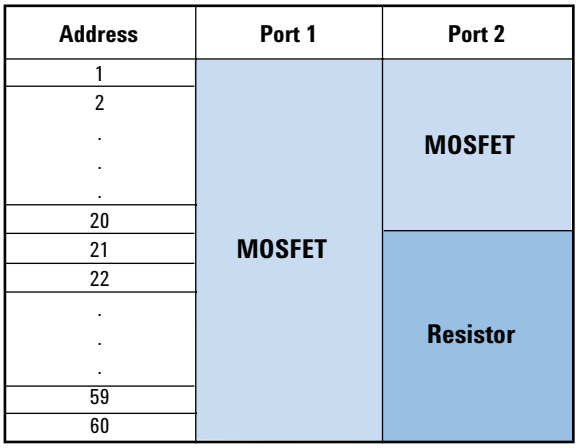

Figure 11. Array-A has two measurement ports that connect to 60 devices each.

Testing task	Conventional test structures tested with 4070			Array-A tested with N9201A		
	# of instances	Unit test time (ms)	Total test time (ms)	# of instances	Unit test time (ms)	Total test time (ms)
Index wafer prober	15	400	6,000	1	60	0
MOSFET measurements: Vtlin, Vtsat, Idsat, Ioff, etc.	80	1,000	80,000	60	1,000	60,000
Resistance measurement	40	100	4,000	0*1	100	0
		Total (sec/die):	90.00			60.06
		Normalized throughput:	100%			150%

*1. Since the resistor test time is << than the MOSFET test time, the resistance tests have no effect on the total test time.

Figure 12. Throughput comparison of conventional test structures tested using the 4070 series versus Array-A using the N9201A.

Array-B: A 4-port array optimized for the N9201A

Overview of Array-B

This addressable array is an improved design based upon the lessons learned from Array-A. It is optimized for use with the Agilent 4070 and 4080 parametric test systems and the N9201A. Like Array-A, it also has an address depth of 64. However, it has four parallel measurement ports, each consisting of three pairs of force and sense signal lines for making Kelvin measurements on MOSFET drain, gate and source terminals (similar to Array-A). This gives Array-B the capability to contain up to 240 test structures, organized as 4x30 MOSFETs, 4x20 resistors, and 4x10 miscellaneous devices (please see Figure 13). Although the size of this array may seem rather large, it would take over 30 48-pad conventional test modules (33 pads used for DC test and the rest used for other testing) to contain the same number of test structures. This gives Array-B a test structure device density that is about six times greater than that of a conventional test module layout.

Significant test throughput improvement through optimized address assignment

Array-B uses the same address for similar device types across all four of the parallel measurement ports. Thus, the test device measurement execution time for each port at each address is much more similar than for Array-A, which results in much less dead time during parallel test execution. In fact, the parallel testing efficiency approaches an ideal level. Including the elimination of the prober indexing time, the throughput improved by 4.28 times over that achieved using comparable conventional test structures (please see Figure 14).

Address	Port 1	Port 2	Port 3	Port 4
1	MOSFET	MOSFET	MOSFET	MOSFET
. . .				
30				
31	Resistor	Resistor	Resistor	Resistor
. . .				
50				
51	Miscellaneous	Miscellaneous	Miscellaneous	Miscellaneous
. . .				
59				
60				

Figure 13. Array-B has an improved design that groups similar structures into the same address space to optimize measurement throughput.

Testing task	Array-B tested with N9201A		
	# of instances	Unit test time (ms)	Total test time (ms)
Index wafer prober	1	30	30
MOSFET measurements: Vtlin, Vtsat, Idsat, Ioff, etc.	20 x 4*1	1,000	20,000
Resistance measurement	10 x 4*1	100	1,000
		Total (sec/die):	21.03
		Normalized throughput:	428%

*1. Four devices are tested in parallel.

Figure 14. Throughput improvement achieved with Array-B using the N9201A.

Array-C: A drop-in addressable array for technology development and yield ramp-up

Overview of Array-C

Array-C is a large drop-in addressable array containing one million uniformly sized MOSFETs. More information about this array can be found in the 2008 issue of VLSI Technology Digest.[1] Using this large array of test structures it is possible to compare device parameters across different die on the same wafer, different wafers on the same lot, and different wafer lots. By using this array in conjunction with the N9201A, you can distinguish between random and systematic process variations much more quickly than is possible using conventional test structures.

Addressable array multisite testing

Unlike Array-A and Array-B, Array-C does not have multiple measurement ports. Instead, a special multisite probe card is used to enable the same address signal to be applied to four different arrays at the same time. This permits measurement at four different sites on the same wafer in parallel (refer to Figure 15).[2]

Dramatic throughput improvement enables the advanced process evaluation

When used in conjunction with optimized test algorithms, Array-C attained a throughput improvement more than 100 times faster than that possible using conventional test structure designs. This demonstrates that the combination of this addressable array and the N9201A achieves results that simply would not otherwise be possible, and that for the yield ramp up phase of advanced processes this type of solution is invaluable.

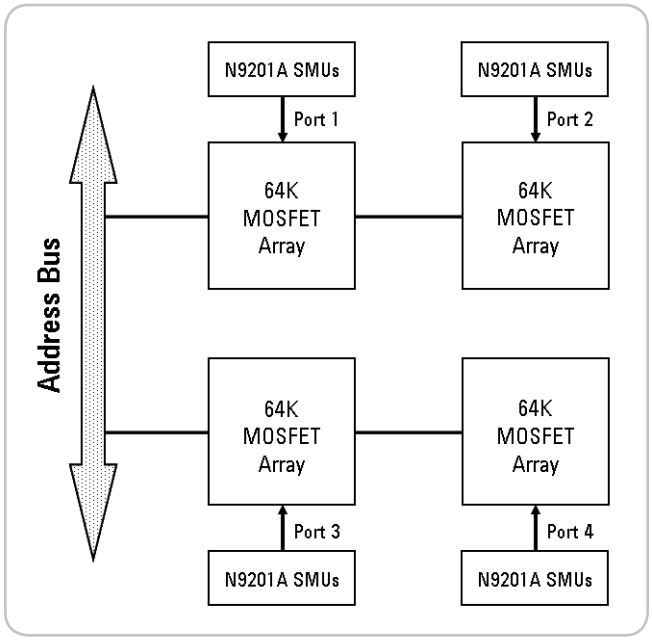

Figure 15. The Array-C architecture supports the measurement of multiple arrays on the wafer simultaneously using a multi-site probe card.

Conclusions

As the need to collect more parametric data for advanced processes in the yield ramp up phase continues to increase, addressable arrays offer an efficient means to meet this need. Addressable arrays also provide an effective way to differentiate between random and systematic process variations in production. Inevitably, high-volume wafer manufacturers will need to employ addressable arrays in one form or another.

The Agilent N9201 Array Structure Parametric Test option provides a very cost-effective means to test addressable arrays, due to its scalability and concurrent test capability.

The N9201A is well-proven, and has been successfully used to improve measurement throughput on addressable array structures. The N9201A also allows you to use the measurement resources of the 4070 or 4080 series tester concurrently.

The amount of throughput improvement obtained using an addressable array and the N9201A depends strongly upon the forethought put into the addressable array construction. The key to maximizing throughput is minimizing the amount of measurement dead time during parallel test. This means placing similar device types at the same address, so that the test times for each measurement port are roughly the same. With careful structure and test design, throughput improvements of 4-5 times (as compared to conventional test structure throughput) is easily obtainable.

1. T. Tsunomura et al., 2008 VLSI Technical Digest “Analysis of 5σ Vth Fluctuation in 65nm-MOSFETs Using Takeuchi Plot”

2. A picture of the probe card is contained in a document that can be found at: **http://www.miraipj.jp/ja/result/071218/005mirai071218.pdf**

Appendix A

The effect of switching circuit resistance on device measurements

Addressable array switching circuits typically have series resistances on the order of 100-500 Ω, which can easily impact measurement results. For example, suppose you are trying to measure the drain saturation current (Id_{sat}) of a MOSFET in an addressable array with $R_{sd} = R_{ss} =$ 100 Ω. Assuming an Id_{sat} of approximately 100 µA at $V_{ds} = V_{gs} = 1$ V, Ohm's law tells us that the actual V_{ds} and V_{gs} bias voltages at the DUT will be reduced by approximately 10 mV. Therefore, it is not possible to make a truly accurate Idsat measurement (refer to Figure 16).

To correct for the series resistance, addressable arrays typically add a sense signal line to monitor the actual voltage applied to the DUT terminal. There are two measurement techniques used to make these types of measurements, the Kelvin SMU Method and the Monitor and Adjust Method. These two methods are shown in Figures 17 and 18 (respectively).

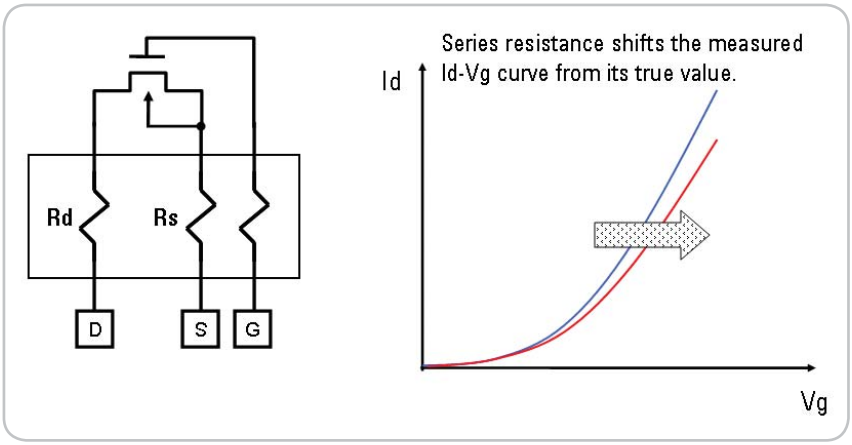

Figure 16. The series resistance in the signal lines going to the MOSFET distorts the actual voltage applied to the DUT.

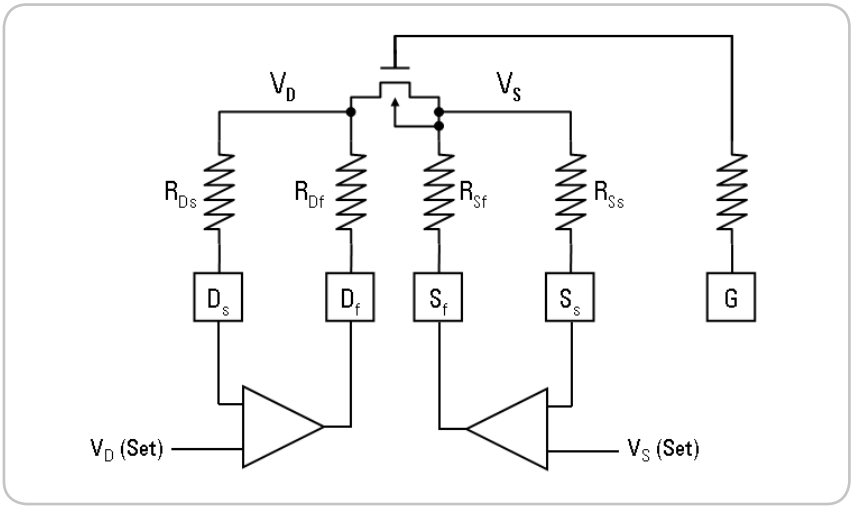

Figure 17. Kelvin SMU Method – The force and sense inputs of the Kelvin SMU can be connected directly to the force and sense pads (respectively) of the addressable array to remove the effects of series resistance in the lines.

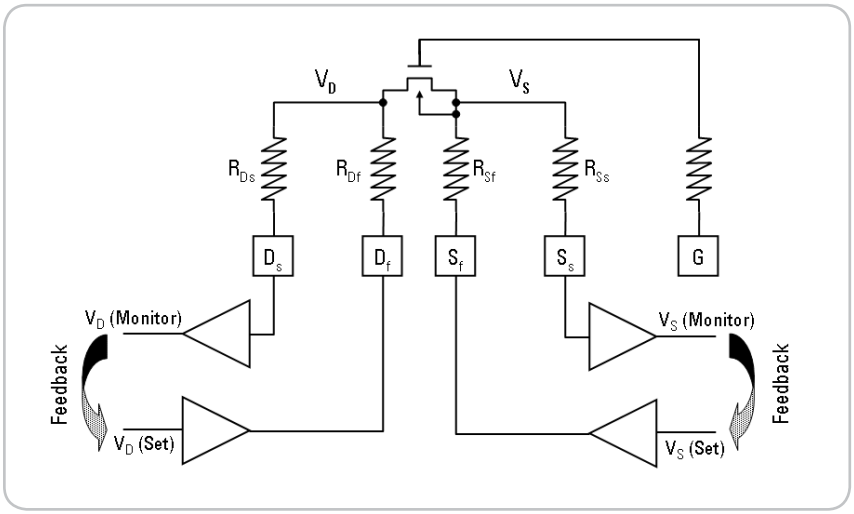

Figure 18. Monitor and Adjust Method – In this method a separate measurement resource is used (SMU or VMU) to monitor the actual DUT voltage and the applied voltage is adjusted until the voltage at the DUT equals the desired value.

Agilent Email Updates

www.agilent.com/find/emailupdates
Get the latest information on the products and applications you select.

Remove all doubt

Our repair and calibration services will get your equipment back to you, performing like new, when promised. You will get full value out of your Agilent equipment throughout its lifetime. Your equipment will be serviced by Agilent-trained technicians using the latest factory calibration procedures, automated repair diagnostics and genuine parts. You will always have the utmost confidence in your measurements.

Agilent offers a wide range of additional expert test and measurement services for your equipment, including initial start-up assistance onsite education and training, as well as design, system integration, and project management.

For more information on repair and calibration services, go to

www.agilent.com/find/removealldoubt

www.agilent.com

For more information on Agilent Technologies' products, applications or services, please contact your local Agilent office. The complete list is available at:
www.agilent.com/find/contactus

Americas

Canada	877 894 4414
Latin America	305 269 7500
United States	800 829 4444

Asia Pacific

Australia	1 800 629 485
China	800 810 0189
Hong Kong	800 938 693
India	1 800 112 929
Japan	81 426 56 7832
Korea	080 769 0800
Malaysia	1 800 888 848
Singapore	1 800 375 8100
Taiwan	0800 047 866
Thailand	1 800 226 008

Europe & Middle East

Austria	01 36027 71571
Belgium	32 (0) 2 404 93 40
Denmark	45 70 13 15 15
Finland	358 (0) 10 855 2100
France	0825 010 700* *0.125 €/minute
Germany	07031 464 6333** **0.14 €/minute
Ireland	1890 924 204
Israel	972 3 9288 504/544
Italy	39 02 92 60 8484
Netherlands	31 (0) 20 547 2111
Spain	34 (91) 631 3300
Sweden	0200-88 22 55
Switzerland	0800 80 53 53
United Kingdom	44 (0) 118 9276201

Other European Countries:
www.agilent.com/find/contactus

Revised: July 17, 2008

Product specifications and descriptions in this document subject to change without notice.

Printed in USA, August 11, 2008
5989-9418EN

Agilent Technologies