KEITHLEY A GREATER MEASURE OF CONFIDENCE

Test Structure Design for Parallel Testing

Randall G. Lee
Keithley Instruments, Inc.

Parallel testing provides higher throughput than conventional sequential testing. Although parallel testing can sometimes be performed successfully on existing test structures, efficient test execution without signal loss generally requires attention to various test structure details. Frequently, optimizing the test structures for parallel test will increase throughput significantly and improve measurement integrity.

Common Substrate Issues

The semiconductor wafers produced in most processes have a common substrate. (Wafers produced by dielectric isolation processes are an exception.) Wells with a polarity opposite to that of the substrate are isolated—for example, separate n-doped wells in a p-doped substrate produced by a CMOS process. However, wells having a polarity identical to that of the substrate—for example, p-doped wells in a p-doped substrate are all shorted together. As a result, simultaneously forcing different voltages at different points can introduce significant errors as a result of:

- Different voltages causing current flow and a voltage gradient across the substrate.
- The voltage gradient causing uncertainty about exact substrate voltages under the gates of transistors under test.

Parasitic Voltage-drop Issues

Semiconductor test structures are generally much smaller than the probe pads used to connect the tester to these structures. As a result, the total area dedicated to a test structure is roughly the same as the area occupied by its probe pads. Understandably, during test structure design, substantial effort is devoted to minimizing the number of probe pads.

Probe pad count is often minimized by using common pads. Common pads are probe pads that connect to more than one device under test (DUT). The most frequent application for common pads is to connect the source terminals together for a set of transistors (***Figure 1***).

However, connection of multiple DUTs to one probe pad can require substantial lengths of metal line between the DUTs, resulting in substantial parasitic resistances. In turn, currents flowing through the metal line can introduce substantial voltage drops.

In ***Figure 1***, R1, R2, and R3 represent parasitic resistances and voltage drops in the line connecting the common source pad to the transistor source terminals:

- If transistors are tested one at a time, the parasitic voltage drop in the common-source line to a transistor is the product of the source current to the transistor and the cumulative line resistances, which has the following implications:
  - Voltage drop for T3 = I3(R3)
  - Voltage drop for T2 = I2(R2 + R3) at T2
  - Voltage drop for T1 = I1(R1 + R2 + R3)
- If all three transistors are tested in parallel, the voltage drops are the products of the cumulative line resistances and the cumulative source currents to the three transistors.
  - Voltage drop for T3 = (I1 + I2 + I3) R3, where I1, I2, and I3 are the source currents for transistors T1, T2, and T3 respectively
  - Voltage drop for T2 = (I1 + I2)R2 + (I1 + I2 + I3)R3
  - Voltage drop for T1 = I1(R1) + (I1 + I2)R2 + (I1 + I2 + I3)R3

Voltage Drop Calculations for the Unoptimized Test Structure. The following

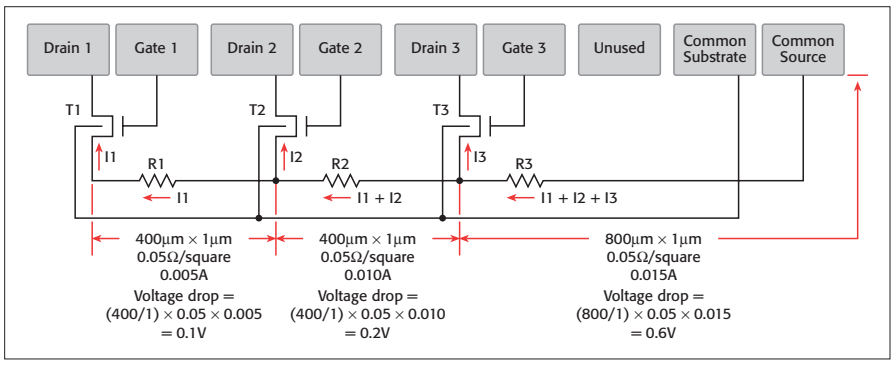

Figure 1. Unoptimized test structure – three transistors with common source and substrate connections made through nine test pads.

assumptions are used to calculate the parasitic voltage drops in ***Figure 1***:

- Pad dimensions are 100µm × 100µm
- Pad spacing is 100µm
- Source line sheet resistivity is 0.05Ω/square
- Source currents are 5mA to each transistor when tested in parallel

The calculated parasitic resistance [(length/width) × sheet resistivity] of R1 and R2 is 40Ω each; R3 is 80Ω. The resulting voltage drops due to these resistances are shown in ***Figure 1***. When the three transistors are tested in parallel, the cumulative voltage drops affect V_{ds} values as follows:

- V_{ds} across T3 = 0.6V less than the voltage forced on the Drain 3 pad.
- V_{ds} across T2 = 0.8V (0.2V + 0.6V) less than the voltage forced on the Drain 2 pad.
- V_{ds} across T1 = 0.9V (0.1 + 0.2 + 0.6V) less than the voltage forced on the Drain 1 pad.

These results clearly show the problem with parasitic resistance. A 0.6–0.9V drop due to parasitic resistance will be highly significant for a process designed for a 3.3V power supply. Even if the transistors were tested one at a time, similar calculations would show that the parasitic voltage drops would only be reduced to:

- 0.2V for T3
- 0.3V for T2
- 0.4V for T1

These voltage drops are still unacceptable, although significantly less than when the transistors are tested in parallel. The example in ***Figure 1*** suggests a need for the following changes:

- A common-source pad located as close as possible to the transistors.
- Individual connections to the common-source pad.

Test Structure Optimization

Figure 2 shows the same basic circuit as in ***Figure 1***, but with a modified pad order and layout that significantly reduces parasitic resistances. More specifically, the changes include the following:

- The three transistors no longer share a common-source metal line; rather, the transistors are connected individually to the common-source pad.
- T1 and T2 essentially extend from the common-source pad, and T3 is connected as closely as possible.
- The drain pads are as close as possible to the common-source pad.
- The gate and substrate pads are now further from the transistors.

Regarding the last bullet point, gate and substrate currents are generally small—on the order of a few microamps. Therefore, even a 100Ω series resistance would cause a parasitic voltage drop of only a few hundred microvolts.

Voltage drops for the optimized test structure. ***Table 1*** lists the most important characteristics (changes highlighted in italics) of the optimized test structure in ***Figure 2***. The resulting parasitic resistances are 1Ω for R1 and R2, and 10Ω for R3. The new parasitic voltage drops for parallel testing are shown in ***Figure 2***.

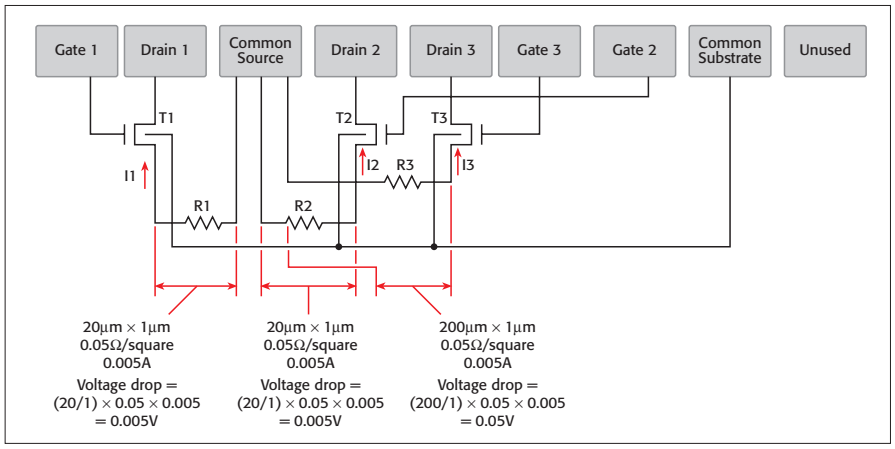

Figure 2. Optimized test structure for the example of Figure 1 – the layout was changed to minimize series resistances.

Table 1. Characteristics of the test structure in Figure 2 (italics indicate changes from Figure 1).

Transistor location	*T1 — Between the Drain 1 pad and the common-source pad.* *T2 — Between the Drain 2 pad and the common-source pad.* *T3 — Between the Drain 2 pad and the Drain 3 pad (by necessity)[1]*
Pad characteristics	Number of pads — 9 Pad dimensions — 100µm (microns) × 100µm (microns) Pad spacing — 100µm
Common-source lines	*The source of each transistor is connected individually to a common-source pad, in each case with a 1.0µm-wide metal line.[2] The lengths of the individual source lines transistors are as follows:* *T1 to common-source probe pad — 20µm* *T2 to common-source probe pad — 20µm* *T3 to common-source probe pad — 200µm*
Source line resistivity	0.05W/square (sheet resistivity)
Source currents	5mA to each transistor

Table Notes:

1. T3 could perhaps be placed next to T2, but a similar series resistance would result—between the Drain 3 pad and the T3 drain terminal, instead of between the common source pad and the T3 source terminal.
2. The 1.0µm-wide metal line could be widened/paralleled for further improvement. Refer to "Additional Improvements" below.

When the three transistors are tested in parallel, the calculated source-to-source pad voltage drops are the same as the individual voltage drops calculated in ***Figure 2***. Therefore, the effects on V_{ds} values are as follows:

- V_{ds} across T3 will be 0.050V less than the voltage forced on the Drain 3 pad.
- V_{ds} across T2 will be 0.005V less than the voltage forced on the Drain 2 pad.
- V_{ds} across T1 will be 0.005V less than the voltage forced on the Drain 1 pad.

These results are a dramatic improvement over the results for the unoptimized test structure.

Additional Improvements

It's still possible to improve on the layout in ***Figure 2***. To reduce the series resistances

further, the metal line widths could be increased significantly. There is no need to keep the metal line width at one micron, because the line need not run outside of the probe pads in this arrangement. For example, the line widths could easily be increased to ten microns, thereby reducing parasitic resistances and voltages by a factor of ten.

Adding similar parallel lines in other metal layers will also reduce the series resistance[1]. If it is not possible to run parallel metal lines under probe pads, the metal line width can be increased significantly where lines run between probe pads.

Making Efficient Use of Assets

Traditionally, test structures are arranged functionally. For example, a set of three test structures might consist of the following:

- Structure A — A probe-pad set connected to a group of transistors.
- Structure B — A probe-pad set connected to a group of capacitors.
- Structure C — A probe-pad set connected to a group of resistors.

While such an arrangement seems logical, it does not necessarily provide maximum throughput for parallel testing. Alternate arrangements can often improve throughput by optimizing the use of multiple instruments in parallel. Consider the following:

A Keithley Series S600 tester can contain up to eight SMUs. A parallel execution thread can control multiple pairs of SMUs.

A parametric tester generally contains a single capacitance meter; it is addressed via the GPIB bus and can be assigned to an additional, independent execution thread.

Optimizing the use of test equipment is just as important as optimizing test structure designs.

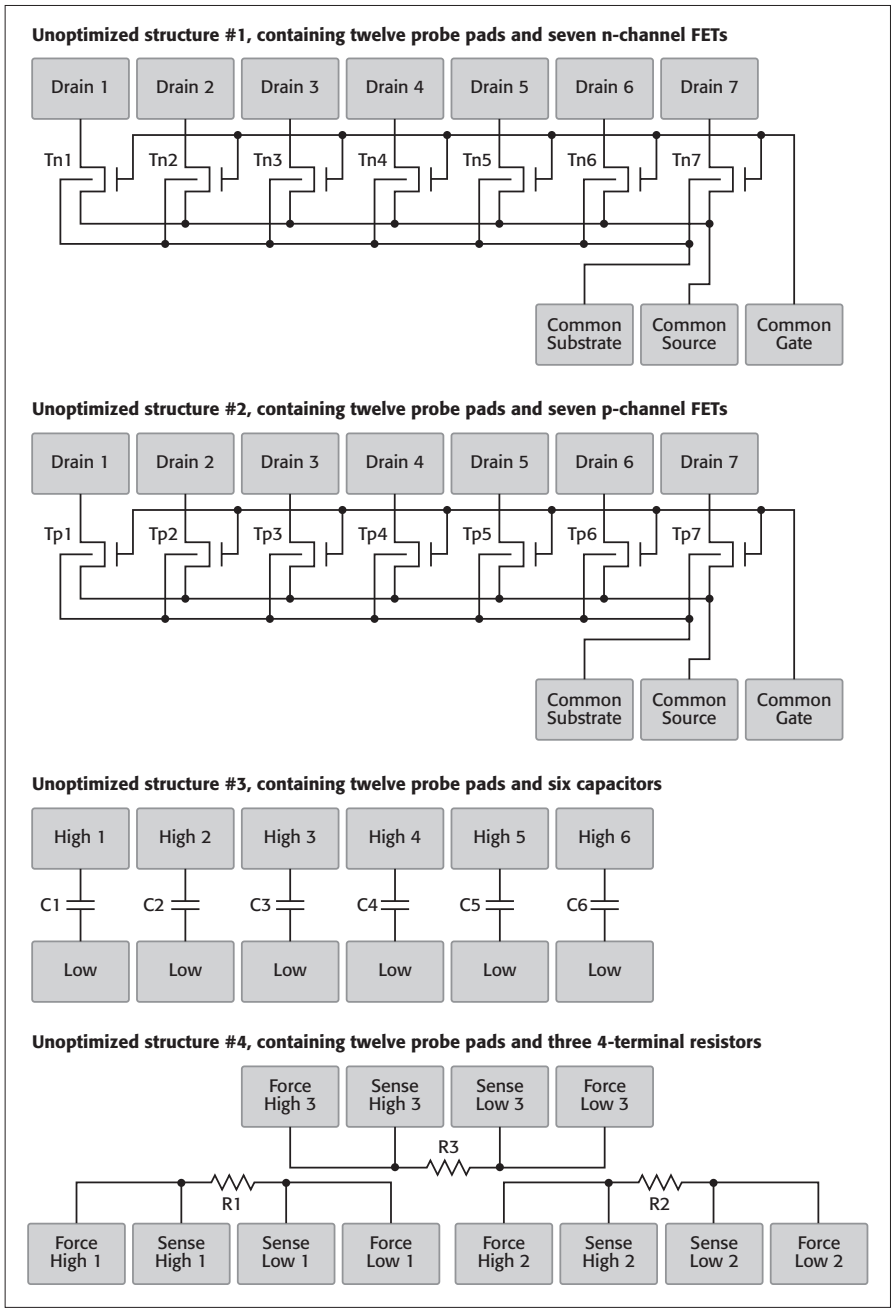

Figure 3. Set of four unoptimized test structures, each containing twelve probe pads.

Testing of Unoptimized Test Structures

Consider the four unoptimized test structures in ***Figure 3***. They contain four sets of twelve probe pads that are connected as follows:

- Unoptimized structure #1 — Seven n-channel transistors
- Unoptimized structure #2 — Seven p-channel transistors
- Unoptimized structure #3 — Six capacitors
- Unoptimized structure #4 — Three 4-terminal resistors

Throughput considerations for the unoptimized test structures. The type of test structure design represented by ***Figure 3*** uses scribe line space very efficiently, requiring only 48 pads. However, the structure is not optimized for throughput for the following reasons:

- In unoptimized structure #1, the use of a common gate pad for the n-channel transistors means that each transistor must be tested separately, and parallel execution threads offer no advantage with this design; the same issue applies to the p-channel transistors in unoptimized structure #2.
- Because the system has only one capacitance meter, each of the six capacitors in unoptimized structure #3 must be tested sequentially.
- In unoptimized structure #4, the three

1 Because all three metal source lines run independently, running them in parallel causes no additional voltage drop to any of the transistors.

Table 2. Time to test the unoptimized test structures.

Unoptimized structure number	Test number	Devices tested in parallel	Individual device-test and prober-movement times					Parallel test time (seconds)*
			n-channel FET, 0.5s	p-channel FET, 0.5s	Capacitor, 0.3s	Resistor, 0.03s	Prober move, 0.5s	
1	1	Tn1	0.5					0.50
	2	Tn2	0.5					0.50
	3	Tn3	0.5					0.50
	4	Tn4	0.5					0.50
	5	Tn5	0.5					0.50
	6	Tn6	0.5					0.50
	7	Tn7	0.5					0.50
Move to unoptimized structure #2							0.5	0.50
2	1	Tp1		0.5				0.50
	2	Tp2		0.5				0.50
	3	Tp3		0.5				0.50
	4	Tp4		0.5				0.50
	5	Tp5		0.5				0.50
	6	Tp6		0.5				0.50
	7	Tp7		0.5				0.50
Move to unoptimized structure #3							0.5	0.50
3	1	C1			0.3			0.30
	2	C2			0.3			0.30
	3	C3			0.3			0.30
	4	C4			0.3			0.30
	5	C5			0.3			0.30
	6	C6			0.3			0.30
Move to unoptimized structure #4							0.5	0.50
4	1	R1				0.03		0.03
	2	R2				0.03		0.03
	3	R3				0.03		0.03

*Parallel test time is the largest individual test time for the devices that are tested.

resistors can be tested in parallel using three execution threads, each one using two SMUs; one SMU forces current, while the other measures differential voltage. (However, the resistor test is generally the fastest of three types of tests, so the advantage is minimal.)

Throughput time for the unoptimized test structures. For illustration purposes, assume that the individual device test times and prober-movement are as follows:

- Full transistor test (V_T, gM, I_{sat}, I_{Dlin}, GIDL current, gate leakage, I_{SUB}, V_{BDSS}, V_{BDII}, etc.) – 0.5 seconds
- Capacitor test – 0.3 seconds
- Resistor test – 0.03 seconds
- Prober movement from one structure to another – 0.5 seconds

Given these assumptions, it would take the tester 10.39 seconds to test all four unoptimized structures sequentially. ***Table 2*** details the individual test times for these structures. In this scenario the unoptimized structure #4 is tested sequentially, since there is minimal advantage to testing the resistors in parallel. If the resistors were tested in parallel, the total test time would be 10.33 seconds, a negligible improvement of only 0.06 seconds (i.e., less than 1%).

Testing of Optimized Test Structures

Consider the alternative set of test structures in ***Figure 4***. These three structures contain the same devices as the four structures in ***Figure 3***. However, the structures are optimized for high throughput using parallel testing. Each of the three structures contains 16 probe pads for a total of 48 probe pads – the same total number as in the set of four unoptimized structures. The devices are distributed among the three optimized structures as follows:

- Optimized structure #1
  - Four n-channel transistors
  - Four p-channel transistors
  - Two capacitors
- Optimized structure #2
  - Three n-channel transistors
  - Three p-channel transistors
  - Two capacitors
- Optimized structure #3
  - Three four-terminal resistors
  - Two capacitors

Test execution for optimized structure #1. This structure has been arranged as follows:

- The transistors have been connected so that both n-channel and p-channel transistors can be tested in parallel; two parallel execution threads can be defined for these transistors using four SMUs for each thread[2]
  - One execution thread and one executor (controlling four SMUs in this case) would be assigned to test one of the n-channel transistors at a time
  - Another execution thread and the second executor (controlling a second

2 Keithley Model 60110-SMUs in an execution thread must be specified in pairs. For this reason, four SMUs must be assigned even though only three are needed. This constraint does not apply when using the Model 60111-SMU.

Table 3. Time to test the optimized test structures.

Unoptimized structure #	Test #	Devices tested in parallel	Individual device-test and prober-movement times: n-chan. FET, 0.5s	p-chan. FET, 0.5s	Cap. 0.3s	Res. 0.03s	Prober move, 0.5s	Parallel test time (sec.)[1]	Conservative derate multiple[2]	Derated parallel test time (sec.)
1	1	Tn1, Tp1, C1	0.5	0.5	0.3			0.50	1.25 (100% ÷ 80%)	0.625
	2	Tn2, Tp2, C2	0.5	0.5	0.3			0.50	1.25	0.625
	3	Tn3, Tp3	0.5	0.5				0.50	1.25	0.625
	4	Tn4, Tp4	0.5	0.5				0.50	1.25	0.625
Move to optimized structure #2							0.5	0.50	N/A	0.500
2	1	Tn5, Tp5, C3	0.5	0.5	0.3			0.50	1.25	0.625
	2	Tn6, Tp6, C4	0.5	0.5	0.3			0.50	1.25	0.625
	3	Tn7, Tp7	0.5	0.5				0.50	1.25	0.625
Move to optimized structure #3							0.5	0.50	N/A	0.500
3	1	C5, R1, R2, R3			0.3	0.03, 0.03, 0.03		0.30	1.03 (100% ÷ 97%)	0.309
	2	C6						0.30	1.00	0.300
Total test time, seconds										5.984

Table Notes:

1. Parallel test time is the largest *individual* test time for the devices that are tested. The values in this column assume that each individual test runs as efficiently in parallel mode as in sequential mode.
2. Conservative derate multiplier allows for slightly longer individual test times when devices are tested in parallel.

set of four SMUs) would be assigned to test one of the p-channel transistors at the same time

- A third executor would be assigned control of a GPIB capacitance meter, which allows testing of a capacitor while the other two executors are testing transistors
  - Because the first pad set contains only two capacitors, one of the capacitors would be tested in parallel with the first pair of transistors
  - The second capacitor would be tested in parallel with the second pair of transistors
  - No capacitors would be tested in parallel with the last two pairs of transistors

Throughput time for the optimized test structures. Assume the same individual device test times as those for the unoptimized test structures:

- Full transistor test (V_T, gM, I_{sat}, I_{Dlin}, GIDL current, gate leakage, I_{SUB}, V_{BDSS}, V_{BDII}, etc.) – 0.5 seconds
- Capacitor test – 0.3 seconds
- Resistor test – 0.03 seconds
- Prober movement from one structure to another – 0.5 seconds

Based on the above assumptions, the total calculated test time for optimized structure #1 would be approximately the time required to test four transistors or 2.0 seconds, assuming 100% efficiency. If one conservatively assumes only 80% efficiency for parallel testing, the total test time for this structure would be 2/0.8 = 2.5 seconds.

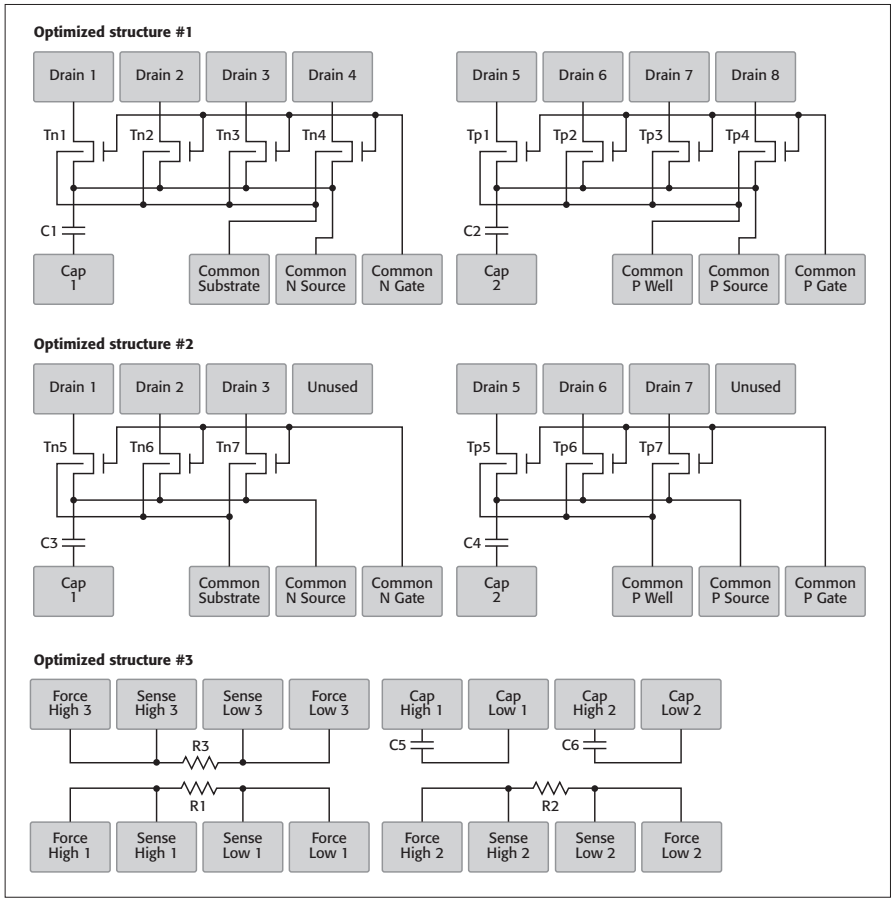

Figure 4. The devices of Figure 3 after rearranging into three asset-optimized, 16-pad test structures.

Optimized structure #2 is similar to optimized structure #1, except that it contains only three n-channel and three p-channel

transistors. Again, two parallel execution threads controlling four SMUs each would be defined, so that one p-channel and one n-channel transistor can be tested in parallel. Again, there are also two capacitors in this pad set. Each capacitor is tested in parallel with a transistor pair using a GPIB capacitance meter. Based on the assumed 0.5-second transistor test time and 0.3-second capacitor test time, the total test time for optimized structure #2 would be approximately the time required to test three transistors, or 1.5 seconds assuming 100% efficiency. If one conservatively assumes only 80% efficiency for parallel testing, the total time for this set would be 1.5/0.8 = 1.875 seconds.

Optimized structure #3 includes the three resistors and the remaining two capacitors. In this case, three execution threads could be defined, each containing two SMUs. In each pair of SMUs, one SMU would force current and the other would measure the differential voltage drop across the resistor. A fourth executor would be defined as a GPIB capacitance meter. In this case, each capacitance measurement would take much longer than the parallel testing of the three resistors, so the total test time would be the sequential time to test two capacitors, or 0.6 seconds. If one conservatively assumes 97% efficiency for the parallel test (three resistors and one capacitor in parallel) and 100% efficiency for the second capacitor alone, the total time for this set would be 0.3/0.97 + 0.3 = 0.609 seconds.

Total throughput time for optimized test structures. Based on the calculations above for individual optimized structures, plus two prober movements between structures, the total, conservatively derated time to test the optimized structures would be 5.98 seconds. Compare the detailed time breakdown for the optimized structures in ***Table 3*** with the detailed time breakdown for the unoptimized structures in ***Table 2***.

Conclusions

Even assuming conservatively derated times to test the optimized structures (5.98 seconds total), these examples show that parallel testing of optimized test structures, instead of sequentially testing unoptimized test structures, would result in a test time reduction of more than 40%:

(10.39 – 5.98) / 10.39 = 0.42

This improvement is achieved without compromising efficient use of space on the wafer. The total pad count in both optimized and unoptimized designs is the same, suggesting that the total scribe line area will also be the same.

This is the last of the Keithley parallel test article series. More information on wafer level parallel parametric test can be found in Keithley's Parallel Test Technology handbook, available at http://www.keithley.com/at/508. KEITHLEY

About the Author

Randall Lee is a Senior Industry Market Manager in Keithley Instruments' Semiconductor Test Group where he is responsible for parametric tester product marketing and management. He received a BS in Electrical Engineering from Rensselaer Polytechnic Institute, and has 25 years of experience in semiconductor lithography and wafer characterization and test.