

INTERNATIONAL
TECHNOLOGY ROADMAP
FOR
SEMICONDUCTORS

2009 EDITION

TEST AND TEST EQUIPMENT

THE ITRS IS DEvised AND INTENDED FOR TECHNOLOGY ASSESSMENT ONLY AND IS WITHOUT REGARD TO ANY COMMERCIAL CONSIDERATIONS PERTAINING TO INDIVIDUAL PRODUCTS OR EQUIPMENT.

TABLE OF CONTENTS

Introduction and Scope of the 2009 edition	1
Key Drivers, Difficult Challenges, and Future Opportunities.....	1
Key Drivers.....	4
Difficult Challenges (in priority order)	7
Future Opportunities	9
Test and Yield Learning.....	10
Electrical-Test -Based Diagnosis	10
Failure Analysis.....	11
Test Cost Focus Topic.....	12
Base Cost Trend	14
Channel Cost Trend.....	14
Power Cost Trend	15
Interface Cost Trend	15
Multi-site Trend	15
Other Cost Trends.....	16
Important Areas of Concern.....	16
Adaptive Test	17
Adaptive Test definition.....	17
Process Variability – its impact on Testing	18
Directions for Adaptive Test in the next 5–10 years	19
Adaptive Test Building Blocks.....	20
Test Technology Requirements.....	21
Introduction	21
System Integration—SoC and SiP Test Challenges and Implications	22
Logic.....	28
High Speed Input/Output Interface	29
Memory	34
Analog and Mixed-signal.....	35
Radio Frequency	36
Reliability Technology Requirements	37
Burn-In Requirements	38
Test Mechanical Handling Requirements	40
Device Interface Technology Requirements.....	41
Probe Cards	42
Test Sockets	45
Specialty Devices	48

LIST OF FIGURES

Figure TST1	Test Cost Drivers	12
Figure TST2	Test Cell Cost / Unit versus Interface Cost Trend	14
Figure TST3	Importance of Multi-site Efficiency in Massive Parallel Test	16
Figure TST4	High-level diagram of Adaptive Test Flow	18
Figure TST5	Organization of Cores for System Integration and Applications	22
Figure TST6	Potential Solution for the Problem of the Test Data Increase	23
Figure TST7	Impact of Repeated Use of Same Cores	24
Figure TST8	Potential Solutions for Test Time Reduction	24
Figure TST9	High Speed Interface Trend	30
Figure TST10	High Speed I/O Jitter Test Accuracy Requirements Scaling with Frequency ...	33
Figure TST11	The Production Process with WLBI Compared with Package Burn-in	39
Figure TST12	Probing and Wirebond Contacting a Bond Pad	44
Figure TST13	Contactors Types	47
Figure TST14	Image Sensor Cell	49

LIST OF TABLES

Table TST1	Summary of Key Test Drivers, Challenges and Opportunities	3
Table TST2	Multi-site Test for Product Segments	15
Table TST3	“Stages” of Adaptive Test	19
Table TST4	Implications of Adaptive Test	21
Table TST5	System on Chip Test Requirements	24
Table TST6	Logic Test Requirements	29
Table TST7	Vector Multipliers	29
Table TST8	Memory Test Requirements	34
Table TST9	Mixed-signal Test Requirements	36
Table TST10	RF Test Requirements	36
Table TST11	Burn-In Test Requirements	39
Table TST12	Test Handler and Prober Difficult Challenges	41
Table TST13	Prober Requirements	41
Table TST14	Handler Requirements	41
Table TST15	Probing Difficult Challenges	43
Table TST16	Wafer Probe Technology Requirements	45
Table TST17	Test Socket Technology Requirements	48

TEST AND TEST EQUIPMENT

INTRODUCTION AND SCOPE OF THE 2009 EDITION

The 2009 edition of the Test Roadmap contains some significant changes to many of the tables; includes a new section on Adaptive Test; includes some discussion of 3D silicon devices; and added accelerometers to the Specialty Devices section. A survey on Cost of Test was completed in 2009 and the results are included in the Cost of Test Focus topic section. The high speed interfaces section was heavily updated. Other sections of the chapter contain only minor revisions.

Adaptive test has been being adopted as a method to reduce overall product test cost and improve product yields. The Test chapter now devotes an entire section to adaptive test concepts, challenges, impact to engineering and factory methodology, IT infrastructure requirements; and some implementation examples.

3D (Three Dimensional) silicon devices have been added to the key drivers for 2009. The concept of 3D devices is that a single design is implemented across multiple die connected by TSVs (Thru Silicon Vias) in order to optimize the cost and performance of devices. The SoC methodology of integrating a very complex system on a single chip can require many mask layers to be added to a technology to accommodate the requirements of the SoC. The 3D concept allows each die domain to be process optimized (i.e., Logic, DRAM, NVM, Analog), but introduces design and test complexities.

The test parallelism table has remained controversial even though new device categories have been added since 2005, so typical functional pin count assumptions have been added as notes. However, as many devices are being tested at wafer probe using reduce pin interfaces; all possible parallelisms cannot be included in the table.

The SoC table has been revised to reflect the latest trends and the chapter section has been appropriately changed. Formulas based changes in the Logic Table reflect the impact of the 2009 MPU and Consumer transistor count changes in the ORTC tables. Memory changes were also driven by the 2009 ORTC density roadmap.

Carrier frequency range in the RF table has been broken into general and special radio categories and the rapid increase in modulation bandwidth required for technologies such as UWB has moved out in time. The remainder of the table reflects small changes in numeric values and some cell color changes.

Prober, probecard, handlers and test socket tables have had significant changes. The prober table has had a near complete rewrite, so changes are not easily summarized. Probecards now reflect a special LCD display driver probe requirements section to specify the unique challenges of testing these devices. Handlers broke the high power (>10W) device requirements into a med-hi (10–50W) range and a high range (>50W). Test sockets now covers blade rubber contacts and also included a table showing socket self-inductance must be reduced below 0.3nH to properly support 6GHz; and below 0.1nH for support of 20GHz signals.

There were either no changes, or only cell color changes or small numerical changes to the Vector multiplier, Mixed signal, Burn-in and Probing difficult challenges tables.

This document represents significant contributions from a large number of participants representing a wide cross-section of the industry as noted in the acknowledgements.

KEY DRIVERS, DIFFICULT CHALLENGES, AND FUTURE OPPORTUNITIES

Starting with the 2005 edition of the ITRS Test Chapter, this important section of the test chapter was refined by splitting the Difficult Challenges section into Key Drivers and Difficult Challenges and adding a Future Opportunities section. This split distinguishes the drivers, which are primary boundary conditions that define the scope of solutions for upcoming manufacturing test for semiconductor components, from key technical and business challenges. At a high level, these boundary conditions actually represent expectations or even requirements of the test process, while the challenges represent current and upcoming key roadblocks, strategic inflection points, or opportunities for the future. The key drivers determine many of the concepts and table values in this chapter. The difficult challenges define areas where more development or understanding is needed in order to cost effectively meet the semiconductor roadmap.

For many years, the mission of semiconductor manufacturing test has been described as “screening defects” and to a lesser extent or within certain business segments “speed binning” or “speed classification.” It is interesting to note that some of the most important test challenges are now actually centered on some of the more subtle historical missions of

2 Test and Test Equipment

manufacturing test—reliability and yield learning. It is also important to note that the impact of these challenges affect not only on the manufacturing test process itself, but are essential to the entire semiconductor business, both in terms of enabling the cadence or timely delivery of future processes and cost effective products, but also in terms of meeting customer expectations for reliability.

Within the Difficult Challenges section, the challenges are listed in order of perceived importance or priority. For example, test for yield learning is followed by screening for reliability and is then followed by increasing systemic defects. In contrast, there is no specific intent in the ordering of the Key Drivers, as they are all boundary conditions or requirements that the semiconductor test solutions must meet. *Table TST1* summarizes all of the key test drivers, challenges, and opportunities.

Table TST1 Summary of Key Test Drivers, Challenges and Opportunities

Key Drivers (not in any particular order)	
<i>Device trends</i>	<i>Increasing device interface bandwidth (# of signals and data rates)</i>
	<i>Increasing device integration (SoC, SiP, MCP, 3D packaging)</i>
	<i>Integration of emerging and non-digital CMOS technologies</i>
	<i>Complex package electrical and mechanical characteristics</i>
	<i>Device characteristics beyond one sided stimulus/response model</i>
	<i>3 Dimensional silicon - multi-die and Multi-layer</i>
	<i>Multiple I/O types and power supplies on same device</i>
<i>Increasing test process complexity</i>	<i>Device customization during the test process</i>
	<i>“Distributed test” to maintain cost scaling</i>
	<i>Feedback data for tuning manufacturing</i>
	<i>Dynamic test flows via “Adaptive Test”</i>
	<i>Higher order dimensionality of test conditions</i>
<i>Continued economic scaling of test</i>	<i>Physical limits of test parallelism</i>
	<i>Managing (logic) test data and feedback data volume</i>
	<i>Defining an effective limit for performance difference for HVM ATE versus DUT</i>
	<i>Managing interface hardware and (test) socket costs</i>
	<i>Trade-off between the cost of test and the cost of quality</i>
	<i>Multiple insertions due to system test and BIST</i>
Difficult Challenges (in order of priority)	
<i>Test for yield learning</i>	<i>Critically essential for fab process and device learning below optical device dimensions</i>
<i>Detecting Systemic Defects</i>	<i>Testing for local non-uniformities, not just hard defects</i>
	<i>Detecting symptoms and effects of line width variations, finite dopant distributions, systemic process defects</i>
<i>Screening for reliability</i>	<i>Implementation challenges and efficacies of burn-in, IDDQ, and Vstress</i>
	<i>Erratic, non deterministic, and intermittent device behavior</i>
<i>Potential yield losses</i>	<i>Tester inaccuracies (timing, voltage, current, temperature control, etc)</i>
	<i>Over testing (e.g., delay faults on non-functional paths)</i>
	<i>Mechanical damage during the testing process</i>
	<i>Defects in test-only circuitry or spec failures in a test mode e.g., BIST, power, noise</i>
	<i>Some IDDQ-only failures</i>
	<i>Faulty repairs of normally repairable circuits</i>
	<i>Decisions made on overly aggressive statistical post-processing</i>
Future Opportunities (not in any order)	
<i>Test program automation (not ATPG)</i>	<i>Automation of generation of entire test programs for ATE</i>
<i>Simulation and modeling</i>	<i>Seamless Integration of simulation and modeling of test interface hardware and instrumentation into the device design process</i>
<i>Convergence of test and system reliability solutions</i>	<i>Re-use and fungibility of solutions between test (DFT), device, and system reliability (error detection, reporting, correction)</i>

ATE—automatic test equipment ATPG—automatic test pattern generation BIST—built-in self test HVM—high volume manufacturing
MCP—multi-chip packaging MEMS—micro-electromechanical systems

4 Test and Test Equipment

KEY DRIVERS

As previously mentioned, the key drivers in the Test chapter are considered *boundary conditions* within which the semiconductor test function must operate and still deliver acceptable component quality and reliability at continued test cost scaling.

DEVICE TRENDS

- *Device Interface Bandwidth*—In the 2001 edition of the ITRS Roadmap, a change in direction was denoted to deliver increasing system performance with rapidly increasing (compared to previous trends) component I/O speeds. Beginning with the 2003 edition, the Test working group noted that while the clock speeds are not climbing as rapidly as the 2001 ITRS expected, the penetration of Gb/s interfaces is permeating throughout a wider spectrum of the semiconductor component market segments (memories, CPUs, chipsets, etc.). The design clock speed roadmap was changed starting with the 2008 roadmap edition to reflect the lower GHz due to architecture and transistor scaling. High speed serial and differential I/O protocols will continue to rise in speed and continue to force innovation in the DFT and HVM test processes to maintain continued cost scaling.
- *Increasing Device Integration*—SoC and SiP integration continue to rise in prevalence throughout a number of business segments. Increased device integration forces a re-integration of test solutions to maintain scaling of test costs and product quality. The optimized test solutions for stand-alone RAMs, cores, and other blocks typically do not scale linearly without modification, additional DFT, or new partitioning to the integrated device test solutions. In particular, additional DFT in-die or even in-package may be required to provide access to and testing of embedded blocks and cores, or additional distribution or re-integration of the HVM test instrumentation may be required. Techniques for known good die (KGD) that provide high quality dice for multi-die packaging also become very important and an essential part of the test techniques and cost trade-offs.
- *Integration of Emerging and Non-Digital CMOS Technologies*—In the recent past, mixed-signal device circuits have been increasingly delivered on die integrated to digital CMOS, posing key challenges for ATE, instrumentation, and test manufacturing flow integration choices. It also poses new challenges and opportunities for DFT innovation, where it had not been previously prevalent (as it had in digital logic and memories). Integration of RF circuits is already starting to experience a similar rise, as are more radically different types of semiconductor devices, such as MEMS (which are already showing up in integrated forms) and optical (which are just around the corner from integration to larger CMOS digital die). As in previous types of integration, the test mission for these devices revolve around access to potentially embedded blocks and integration of radically different test methods into a cost effective manufacturing process.
- *Package Form Factor and Electro-Mechanical and Thermal Characteristics*—The envelope of package form factors is pushing both outward and inward, with higher complexity form factors in MCP and SiP for multi-die on the high end, but also chip-scale packaging pushing towards the small end for systems targeted for the smallest platform form factors (e.g., handhelds). In addition, multi-functionality of die packaging with inclusion of heat sinks and spreaders, and perhaps even voltage regulation and power management functionalities, appears to be on the increase. With increasing power a major concern, form factor thermal transfer characteristics and uniformities (thermal gradient coefficient-junction to ambient, junction to case [θ_{JA} , θ_{JC}], etc.) become even more critical to the test process. These expansions of the form factor technology envelopes necessitate improvements and delivery of key test sub-systems: handling for packaging test, Ohmic contacting technologies, and test sockets (the latter described later in the chapter for the first time).
- *Device Characteristics beyond One-Sided Stimulus Response Model*—The history of semiconductor testing and ATE architecture and usage is built upon a foundation of deterministic device behavior. In addition, digital CMOS testing has historically been based on simplifying the environmental topologies of Vcc, temperature, and frequency, so that worst case (one-sided), one iteration tests guaranteed the performance over a wider and continuous range of environment topologies. When these topologies were driven by only two or even three variables, they were manageable and fairly easy to characterize and to optimize for HVM test. However, a number of device characteristics threaten the dimensionality of this one-sided test paradigm, threatening a non-linear effect on non-recurring engineering (NRE) development costs for test as well as potentially limiting continued test cost scaling for HVM. Device features such as extended self-repair and correction and built-in variable multi-power management modes potentially add tremendous complexities to the device test conditions contour. In addition, from the device design, architecture, and behavior front, there looms the increasing future possibility of non-deterministic device behaviors (e.g., asynchronous logic architectures and fault tolerant devices) where correct die to die device behavior (from end user system standpoint) under similar conditions is

not deterministic within the time/vector synchronization standpoint. These behaviors, while correct from an end use customer and system standpoint would break the historical HVM test stimulus-response model, where one set of digital logic 1s and 0s delivered on the same synchronization, would suffice to test such behaviors. While these device architectures and behaviors are not widely prevalent today, they do pose either a major paradigm shift or innovation challenge for the semiconductor test process in the not too distant future.

- *3D silicon devices*—Developments in TSVs have enabled the ability to connect die together using tens, hundreds, thousands, or even millions of connections that could be as small as $2\ \mu\text{m}^2$ per connection. Because of this, it is now possible to design a system that spans multiple die domains and is partitioned across multiple die to optimize product cost and performance. A typical monolithic integration put many board components on a single die. 3D integration is simply taking the many components that are on a board, re-architecting the system to exploit the performance improvements that can be achieved from using 3D interconnects, properly partitioning the entire system for cost and manufacturability; and stacking the die vertically. As an example, an MPU can have the logic cores on one die, the RAM cache on a second die, and NVM storage on a third. The MPU example would require three dimensional design capabilities so the thousands of cache connections (TSVs) to the cores are defined to be in the same X and Y locations. Feed thru vias to the NVM memory must go from the logic die thru the cache die to the NVM memory die. In addition to connecting the die, since each die must be able to be tested on its own at wafer probe test, test insertion must be complete for a single die domain. As die are stacked on top of each other, connections are made with the TSVs and the multi-die stack must be able to be tested as a semi-completed system using the test insertion. Vectors and paths are likely to be different from those used for testing of the individual die. A further complication arises as the TSVs will likely not have any ESD protection because ESD structures will greatly increase the die area cost of 1000's of TSVs. Physically probing non-ESD protected pads or TSV nodes is likely to damage the die. For this reason, it is recommended that the test solution NOT contact the TSVs during test, but to use an alternative method such as boundary-scan or other non-contact means to access the TSVs pads.
- *Multiple I/O types and power supplies on same device*—Integration of devices in SoC and SiP will drive and increased number of power supplies in order to optimize the power consumption per core. Logic, analog, memory, mixed signal, and high speed I/O each optimally operate under different power supply conditions. Using multiple supplies versus attempting to generate or regulate voltages on chip saves die area and power and helps in achieving greater noise isolation across the cores. Single ended and differential I/O will be used to optimize power while not limiting the data transfer performance.

INCREASING TEST PROCESS COMPLEXITIES

- *Increased Device Customization and Line Item Complexity (during the test process)*—Increasingly, the test process is expanding to include not only the metrologies surrounding the question “is this a good die?” but also actual process steps that modify, differentiate, or customize specific die. Examples of these include memory block (and other) redundancy/repair, burning of on die configuration fuses, programming of read-only memories (ROMs), or other programming of product features. In addition, some business segments are seeing substantial increases in product line items derived from the same die base, with differentiation and physical segregation as part of the manufacturing test flow. All of this places increasing demand on the manufacturing test process and expands equipment (e.g., handling, fuse blowing, etc.) and factory automation and integration infrastructures.
- *Increased “Distributed” Test to Maintain Cost Scaling*—Distribution of test content and manufacturing instrumentation and equipment platform requirements enabled by a wider application of DFT techniques will continue to be a widespread industrial test process integration scheme used to optimize test manufacturing capital expenditures, operational test cost, and outgoing product quality. Distribution and differentiation of different types of machines for design verification versus manufacturing test will continue. Typical vectors for this differentiation will center around the speed, accuracy, number, and types of channels on verification versus manufacturing tools. Integration of a wider array of digital CMOS circuits, mixed signal and even optical and or mechanical silicon together in future products will continue to make distributed test solutions more attractive than to increase the platform complexity of ATE towards a “do all” platform for manufacturing test. Hand in hand with this will be increases in the complexity (number of steps and interactions) within the manufacturing test process flow itself.
- *Increased Test Data Feedback to Tune Manufacturing*—Test data usage for purposes beyond identifying whether a given die is good or defective has become essential for several reasons and drives a need for expanded, revamped, and better integrated test data systems and infrastructure. In one sense the output of the component test process is data, i.e., the results of the various applied metrologies of the test process across the product dice

6 Test and Test Equipment

and manufacturing lot populations. The need for better integrated usage of this test output (data) for fab process yield learning, maverick material identification, and feedback within a more distributed manufacturing test are all becoming more critical and even essential applications moving forward, not just nice to have.

- *Dynamic test flows via “Adaptive Test”*—An emerging method to reduce overall test cost is to break the paradigm of running a constant test flow and instead, instituting a method whereby results of recently tested die or units can be used to determine the probability of need for a particular test. If the “importance” of a test or group of tests is deemed statistically low, then the test(s) can either be temporarily dropped or have other tests substituted. Probability triggers can also be used to increase the tests that are executed.
- *Higher Order Dimensionality of Test Conditions*—Historically, component manufacturing test has used testing at a simple matrix of points or corner conditions of two or three environmental variables (typically Voltage, Temperature, and Frequency) to guarantee the wider multi-dimension contour and interior space of product specifications beyond the contour outlined by the test points. Components in a broad range of markets from battery application platforms to client computing and servers are adding numerous schemes of power management that quickly add exponential complexity to the worst-case test conditions contour. For example, the complexities of component schemes may include multiple or variable power modes (sleep, hibernation, etc) or even multiple and *in situ* responsive variable Vcc and frequency control systems to achieve optimized power-performance in the application. The challenge of determining, characterizing, and optimizing (reducing) the larger set of environmental test points for component manufacturing test is daunting with the additional dimensions of these variables. Keeping pace with this complexity requires additional validation efforts and innovative methods to validate the quality of the more complex environmental test points. This is needed to ensure the lack of holes and predictability of more complex worst case conditions in the product test validation and test development phase. This increased complexity of test environment set points will also challenge the continued economic scaling of manufacturing test cost by the product of the applied test content and the number of iterations of its application at the various set points in the manufacturing flow.

CONTINUED ECONOMIC SCALING OF TEST

- *Physical Limits of Further Parallelism*—Over recent generations, continued increases in parallelism (in number of DUTs tested in parallel at a test insertion), particularly for commodity memories, but also for digital logic, have been a primary means of continuing the economic scaling of test in the context of devices with more transistors, increased functionality, and higher I/O and core speeds. In the current test tool and interface hardware integration paradigms, further increases in DUT test parallelism are reaching non-linear limits and the impact will be further seen in future generations. These are driven by the practical limits of how many electrical channels can be squeezed into the physical space between the parallel DUTs and the test instrumentation while yet maintaining acceptable physical and electrical proximities between the two. These approaching physical limits will require alternate means be used or expanded in order to maintain continued economic scaling of test, or alternately, new paradigms of DUT, handling, contacting, and test instrumentation integration must be developed that enable further increases in DUT parallelism beyond what is currently envisioned.
- *Managing (Digital Logic) Test Data Volume*—Increased digital logic die complexity and content drives proportional increases on the test data volume (number and width of vectors). Unconstrained, this additional test data volume drives increases in test capital and operational costs by requiring additional vector memory depth per digital channel of the test tools (ATE) and by increasing test time per DUT. Currently, a number of logic test vector compression schemes are being developed and applied in a variety of ways on the test databases themselves (for scan based tests) or via compression hardware (DFT) on the product die itself. Moving forward, compression will become more ubiquitous across component business segments, driven by the increasing product complexities and higher levels of product integration (e.g., SoC, SiP) and may ultimately require increases in the rate of compression (i.e. the compression ratio of the test database versus uncompressed).
- *Effective Limit for Speed Difference of HVM ATE versus DUT*—In recent years, a primary means of achieving economic scaling of test has been an increasing gap between application device speed versus manufacturing test instrumentation and applied test. This is one key aspect of distributed test, or test partitioning, and is enabled by a variety of DFT techniques such as I/O loopback and special test modes on the DUT. Maintaining this delta between application speed and design verification ATE versus HVM ATE enables a continued economic scaling for test, but there are limits to how much slower or how much less accuracy (e.g., for signal edge placement) is needed in combination with DFT to assure adequate quality of the DUT in the end user customer application. As device speeds and I/O edge rates increase, the manufacturing test instrumentation will lag but will also need to slowly increase, but probably in larger step function increments than the DUT families themselves, as

opportunities for ATE fleet replacements or expansions offer the most optimal opportunities to keep abreast in cost and quality effectiveness.

- *Managing Interface Hardware and (test) Socket Costs*—The portion of costs based on test and probe interface hardware and test sockets is an increasing proportion of the overall test cost. There are a number of factors driving this, such as higher speed (Gb/s), more complex DUT I/O protocols, increased DUT parallelism, higher signal and power pin counts, and increased power delivery and signal channel fidelity requirements. While this might be an alarming trend by itself, this trend needs to be considered in the context of the overall test process cost, and whether or not it enables a continued overall economic scaling of manufacturing test. Managing interface hardware costs within an acceptable range may also be dependent on technology drivers and boundary conditions, e.g., the continued extension of the use of FR4 materials within an acceptable numbers of layers as the primary materials platform for such hardware.

DIFFICULT CHALLENGES (IN PRIORITY ORDER)

(1) TEST FOR YIELD LEARNING

Test's peripheral role as a feedback loop for understanding underlying defect mechanisms, process marginalities, and as an enabler for rapid fab process yield learning and improvement has traditionally been considered a secondary role to screening hard defects. With the increasing reduction in feature (and defect) sizes well below optical wavelengths, rapidly increasing failure analysis throughput times, reduction in failure analysis efficacy, and approaching practical physical limits to other physical techniques (pica, laser probes), the industry is reaching a strategic inflection point for the semiconductor business where the criticality of DFT and test enabled diagnostics and yield learning becomes paramount. In other words, the yield learning rates of the past process generations are not sustainable by historical fault isolation and failure analysis methods per se. Rather, they need to be augmented by more universal deployment of enabling on-die circuitry (DFT, etc.) across and throughout products, as well as improvements in the on-die instrumentation and diagnostic software tools with respect to fault isolation specificity. Where it may have been sufficient to isolate to failing bit in an array or failing gate in logic in the past, there is a real business need to be able to isolate electrically to at least the failing transistor or interconnect in the future, or suffer the economic consequences of reduced yield improvement learning rates on new process technologies, perhaps even to the point of lowering the asymptotic maximum achievable yields per die size on future processes. Given the paramount importance of this "mission" for component test, the Test team has not only summarized its relevance here as the top challenge for test, but also included a detailed section on this topic in the early part of the test chapter.

(2) DETECTING SYSTEMIC DEFECTS

The industry faces new manufacturing-imperfection-related test and yield-learning challenges that result from changing processing technology, changing circuit sensitivities and design modeling limitations.

1. Increasing design-process interactions are increasing the prevalence of systematic defects. Such defects may occur only in certain circuit/layout configurations, due to, for example, pattern density, pattern proximity, and imperfections in optical proximity correction (OPC) algorithms. Although they are systematic, they may appear random because of their rarity and the complexity of the conditions required for their occurrence.
2. Process-technology advancements change the population of physical defects. For example, changes in materials, such as those needed to support high-k dielectrics and metal gates, change defect populations.
3. Changing circuit sensitivities and increased variability are likely to make defects that were benign in the past become killer defects in the future. For example, shorter clock cycles mean defects that cause picoseconds of delay are more likely to cause circuit failures. Furthermore, power-optimized and/or synthesized designs will result in fewer paths with significant timing margin, which implies that random delay-causing defects will more likely cause failure. Similarly, increasing noise effects, such as crosstalk and power/GND bounce, decrease noise and timing margins and again increase circuit susceptibility to defects.
4. Modeling complexity and aggressive design threaten the ability of EDA/design to ensure that a circuit meets power and performance specifications under all process conditions. That decreased ability, combined with increasing process variability, may increasingly result in circuits with parametric failure modes that reach the test floor.
5. Smaller transistors may increase the importance of degradation mechanisms, such as Negative (Positive) Bias Temperature Instability.
6. Radiation-induced soft error rates are increasing to where, in addition to SRAMs, latches and flip-flops are likely to need protecting, at least for chips targeting enterprise applications. Other problems now increasingly observed

8 Test and Test Equipment

include multiple-adjacent-cell SRAM upsets due to single radiation events and erratic shifts in minimum-operating-voltage, problematic for low power applications.

All aspects of the test process, including defect modeling, test generation, test-coverage evaluation, DFT solutions, test application and diagnosis, must handle these realistic and changing populations of manufacturing/operating imperfections. Promising strategies include out-of-spec testing such as low-VDD or temperature, statistical methods, adaptive test, and realistic defect-based modeling/targeting.

(3) SCREENING FOR RELIABILITY

A lesser-publicized mission for semiconductor test beyond the primary “screening defects” or telling a good unit from a bad one has been to screen out infant mortality of the product population to acceptable levels. Another way of describing it more aligned to the “defect” mission of Test is to call this essential function “screening $t > 0$ defects” where “0” on the timeline represents the date product moves to the customer from the component provider. Historically, different semiconductor business segments have used a variety of techniques from burn-in to IDDQ, to voltage stressing during the manufacturing process to identify and screen sufficient numbers of the less reliable sub-populations or product dice to meet customer quality expectations. From a similar set of causes, notably increasing background leakage currents, and reducing product operational margin (collapsing Vdd/Vcc with process scaling), all of these techniques are becoming both less effective and more expensive to varying degrees. Where burn-in equipment and techniques had remained essentially unchanged and re-used over many process generations from the early 80s to the mid 90s, beginning in the late 90s burn-in production systems (in product segments that leverage this technique) have been one of the areas of the largest increases in test capital and interface hardware costs. At the burn-in elevated voltage and temperatures needed to accelerate latent defects, leakage levels are much higher than under normal application conditions. In addition, reducing product Vcc and temperature margins limit the range these conditions can be used to accelerate the latent defect populations. IDDQ, which has been widely used in product segments from ASIC to SoC to commodity memories, has been greatly challenged by the very same trend in proportional increases in background leakage, which greatly reduces the signal to noise ratio of “normal” static current levels versus DUTs that contain latent defects. In fact, this has been a challenge at least since the 250 nm DRAM half pitch, where many companies, began using more advanced techniques, for example, IDDQ “delta” and other Boolean static Icc/Idd combinations and comparisons for results of multiple tests to maintain efficacy against latent defects and reduce invalid yield “overkill” cost effects. Commodity memories, which have tended to have somewhat lower intrinsic leakage levels per technology generation, have managed to extend and depend on the various IDDQ techniques a little farther than have other product families, but even they are now forecasting a real drop off in effectiveness for IDDQ in the next one to two technology generations. Similarly, voltage stress, or applying Vcc/Vdd and patterns well above nominal range to accelerate latent defects, also has been losing effectiveness as the differential between V_{stress} and $V_{nominal}$ has continued to get smaller at each generation. In the long term, new techniques for providing this reliability screening function will likely be needed in some business segments and product types sometime in the next few generations. Some of the newer concepts under exploration are improved Boolean and distribution algorithms among various test results both intra-die test (results from the same die) as well as inter-die (wafer neighborhood analysis, analyses within lots, adaptive test limits, analyses across lots, etc.). Another vector that will likely play a larger role moving forward is correction, whether by self-test and self-correction, which is being implemented on some embedded RAMs today, or by error coding detection and correction (ECC) techniques. The challenge here will not be on memories, but rather how and when similar capabilities or alternate approaches to self-correction might be practically applicable and affordable for logic.

(4) POTENTIAL YIELD LOSSES

Manufacturing yield loss occurs whenever any test or inspection process rejects as faulty a device that would function correctly in the target system. Causes of yield loss include:

- Tester inaccuracies (timing, voltage, current, temperature control, etc.)
- Over-testing (e.g., delay faults on non-functional paths)
- Mechanical damage during the testing process
- Defects occurring in test-only circuitry, e.g., BIST
- Some IDDQ-only failures
- Faulty repairs of normally repairable circuits
- Overly aggressive statistical post-processing

Tester-inaccuracy related yield losses are being mitigated to some extent by the use of alternative test methods to do at-speed functional test. DFT methodologies must mature to provide better coverage of the “collateral” defects currently best identified by at-speed functional test vectors through advanced pattern application methods and novel fault models. Care

must be taken to avoid adding to the over-testing burden in this development path. Further work on appropriate fault models is also required.

Tester inaccuracies in parametric measurements (timing, voltage, current, and temperature control) and DUT parametric drift during test may require guardbands that can be large enough to affect yield. Increasing DUT parametric variation with advanced lithography is pushing the need for more effective BIST or other alternative test strategies such as on-product parametric measurement circuits.

Similar yield loss issues have recently surfaced with inaccurate launch-capture delay tests. Delay path measurement errors of fifteen picoseconds have been observed in delay path measurements—this is 10% of recently announced internal clock periods, indicating a possibility of either yield loss or test escapes, as there is as yet no known way of adding margin testing to delay path measurement. Delay path measurements also contribute to yield loss if inappropriate delay paths are measured. Tools must be developed to avoid yield loss due to measurement of false paths.

IDDQ failures are known to indicate a variety of interconnection and other defects. If, however, IDDQ-only failures are not severe, experience has shown that only a few percent change with burn-in or otherwise prove to be reliability issues. This can be reckoned as yield loss in non-critical applications.

Statistical post-processing is becoming a significant tool for identifying potential failures. However, since it consists of discarding “good die in bad neighborhoods,” for example, it is inherently a potential yield loss mechanism.

In a larger sense, however, the concept of yield loss may be understood to include discarding large numbers of properly fabricated devices or interconnect because a small number of improperly fabricated devices are faulty. Yield loss of this sort is currently mitigated in high-density memories by providing redundant rows and columns; similar strategies have been used successfully at system levels using higher levels of reconfiguration and redundancy. Continued development of reconfigurable circuits for defect avoidance or repair focusing on repair techniques that do not cause collateral damage is needed.

Research and development regarding ways to incorporate on-line testing and repairs has a rich history in systems technology. Migration of the results of these efforts into ICs has the potential of reducing yield loss reckoned in this larger sense. More efforts in pursuit of robust design, online testing, and *in situ* repair technology are required.

FUTURE OPPORTUNITIES

TEST PROGRAM AUTOMATION (NOT ATPG)

While the EDA industry has been providing a range of capabilities around scan DFT and more recently including scan compression and even some DFT synthesis capabilities, for example, for embedded memories, there is an area where the semiconductor suppliers software NRE costs present a productivity and automation opportunity: generation of actual test programs. Across the industry today, the product or test engineering communities spend much of their time creating new customer software code bases, that is, unique and custom individual production test programs. There are a lot of innovative software products for different pieces of this, provided primarily by ATE suppliers along with the ATE, as well as a lot of additional software tooling done in house by the semiconductor companies themselves. There are even historical or new test program data structure standards (test data format (TDF) or standard test interface language (STIL)) that provide some common structural opportunities. However, the basic fact remains is there are no widely available turn-key software products that produce and validate whole main code test programs and yet most of the 1000s of these in use across the industry today do basically the same kinds of things on similar tools across companies.

SIMULATION AND MODELING

Extending the design phase simulation systems and modeling to cover not only die packaging, but also test interface hardware and perhaps even the ATE instrumentation circuits themselves (both parametric and logical) will enable shorter product and manufacturing test validation cycles. This extension will also enable higher reliability pre-silicon verification of the DUT DFT scheme integrated to the manufacturing test instrumentation. Providing such a seamlessly integrated simulation and modeling environment in the pre-silicon design phase would better enable not only accuracy of product performance, but also help tune the test process, and reduce design turns due to minor discrepancies in DFT and test interface modes and hardware.

CONVERGENCE OF TEST AND SYSTEM RELIABILITY SOLUTIONS

Some of the most widely proliferated DFT techniques (such as scan) were actually originally motivated by the need to provide better system reliability for high-end enterprise computing systems. The same scan hardware was re-usable for manufacturing test screening purposes and has proliferated throughout the industry. In the future, there are a number of physical and circuit phenomena that are challenging lower level intra-die reliability such as alpha particle and cosmic ray

10 Test and Test Equipment

induced soft errors, erratic circuit behaviors, and other increasingly intermittent non-predictable lower level behaviors. In response, there is likely to be further extensions of device functionalities to provide additional detection and correction. Like scan, the development and deployment of lower level intrinsic mitigation schemes designed to provide improved system reliability, such as error correction used on RAMs today, and other capabilities of the future, are likely re-usable for test purposes as well (detecting, correcting defects, even systemic ones, on top of the more infrequent or non-predictable intermittent behaviors they would be originally designed to provide protection against). At the very least there will be integration issues with the test process (for example, for redundancy/repair for memories) but more optimistically they could likely be aligned to provide more effective, more efficient, or reduced overhead to future semiconductor test processes.

TEST AND YIELD LEARNING

In addition to the normal sorting function, test provides the essential feedback loop for understanding the characteristics of manufactured chips. Test must increase its capabilities to support failure root cause determination, cost-effective defect isolation, process measurements and design-process sensitivities.

ELECTRICAL-TEST -BASED DIAGNOSIS

Test-based-learning is needed for both (1) defects and (2) parametrics and variability. Defect-learning is needed for both random and systematic defects (see “Defects and Failure Mechanisms”). Test-structure-based defect learning methods suffer from both traditional area-related limitations and by the number of physical design configurations they are able to cover. As susceptibility to localized failures becomes a function of complexities that include OPC algorithms, nearby shapes and neighborhood densities, it is increasingly attractive to base learning on product-test diagnosis, which inherently reflects the physical design configurations of importance to the product.

Parametric-related feedback is needed for (1) device and interconnect parameters and (2) design-process interactions. Measurement of device and interconnect parameters have traditionally relied upon test structures, especially scribe line FETs and interconnect resistance and capacitance monitors. Increasing across-chip variation (intra-die variability) enhances the negative impact of scribe-line-to-chip offsets. Moreover, test structures are limited by the number of physical and electrical configurations they are able to cover. As circuit parametrics are increasingly affected by such configurations, it becomes attractive to base learning on product test. Monitor circuits such as thermal and VDD sensors and process-monitoring ring oscillators are increasingly being embedded and distributed on-chip and can be used to help diagnose parametric fails and understand variability. Such circuits may increase in importance as modeling capabilities relied upon during design fail to keep pace with technology changes and more parametric problems reach the test floor. Product test is uniquely well-suited to provide feedback on design-process interactions, including those leading to noise-related fails, such as power-grid-droop and crosstalk fails. Product-test-based parametric yield-learning is a relatively immature area in need of significant progress.

Electrical-test-based diagnosis methods are needed for circuit types including memories, logic and scan chains. Memory techniques, the most mature, are no longer adequate alone. RAMs have been superseded by microprocessors as technology leaders, they fail to utilize all metal layers and they represent only limited layout configurations. Logic diagnosis has been an active area of recent research and development, but still is challenged in terms of accuracy, resolution and large data requirements. Integration of layout information is commonly employed. Successful techniques for integrating in-line inspection results have been developed, but suffer from in-line data availability and ever-smaller, potentially non-visible, killer defects. Volume diagnostics, which combines results from many failing die, is a promising strategy for addressing remaining challenges. Multi-die statistical analyses, such as those employed in volume diagnostics, hold promise also for distinguishing systematic from random failure mechanisms. Identifying systematic failure mechanisms is important for process improvement, DFM guidance (including hot spot identification) and ensuring test quality. Challenges in making the distinction include setting expectations for random failures against which systematic are compared. Expectations based on design analysis, including traditional critical area analysis, suffer from the need to make failure-likelihood assumptions or depend upon defect models that may not be accurate given complicated design-process interactions. Another emerging role of tools is to provide guidance as to whether or not PFA is appropriate for a given failing part of group of parts, especially given the increasing prevalence of non-visible failure mechanisms (non-visible defects or parametric problems) and the related fact that PFA success rate and cost varies by failure mechanism. The yield impact of broken scan chains has made their diagnosis a recent focus, but more widely applicable and accurate methods are needed. Moreover, defects can affect clock and other infrastructural nets, which are difficult to diagnose with today’s methods. The tools must handle all realistic physical defects, including resistive bridges, resistive contacts/vias and opens. Methods are also needed to handle diagnostics for fails detected by all major test methodologies, including scan-based and BIST-based test; functional; IDDQ and delay test.

The tools and methodologies should support several levels of software-based diagnosis:

1. Production-worthy data collection, trading off resolution against test-cost overhead. Concerns may include test-data compression and BIST approaches. An absolute minimum requirement is failing-block, e.g., core or IP, identification. Average test time overhead should be less than 1%.
2. Extensive data gathering on selected engineering or monitor wafers or lots. Granularity must be sufficient to build an accurate defect-type Pareto and support tool-commonality analysis. Throughput time must be short enough to provide timely feedback to the fabrication process on sufficient volume and must support both time-zero and reliability failures. Tools should identify not just failing nets, but failing layers. Such analyses may involve integrating layout information and/or in-line test results into fault localization. Typical test time should be on the order of seconds. During early production, a more significant part of the material may be exposed to such extensive logging.
3. Individual die analysis that identifies defects to a single transistor or section of conductor no longer than 10 μm and identifies the failing layer. Such analysis may involve special-purpose diagnostic-resolution-enhancing ATPG and fail-data collection and/or analog re-simulation and may be followed-up by failure analysis. Analysis time here may be considerably longer than in the previous two cases.

Data-gathering infrastructure must support increasing diagnostic needs. Specifically, ATE should allow for unlimited collecting of scan data at the model scan vector rates predicted in *Table TST6* for DFT testers. DFT techniques such as BIST and test compression must be designed with special consideration to support the necessary data gathering. IDDQ measurement devices need to support the accuracy levels required by diagnostics. Diagnostic data collection allowing localization to a single or few failing net candidates should not add significantly to overall test time. In addition, factory integration issues must be addressed. Data capture and management capabilities must support increasing reliance on statistical analysis and data mining of volume production data for yield-learning. Secure mechanisms for yield-data flow for distributed design, manufacture and test, including fabless/foundry and 3rd party IP, are needed. Standard test data formats, such as STDF-V4-2007 scan fail data, and infrastructure to support their transmittal, are needed to support automation and sharing of data. Distributed design, manufacture and test also creates an emerging role for methodologies and tools to help determine in which area problems lie, e.g., design house, foundry or test house.

FAILURE ANALYSIS

While electrical-test-based learning is increasingly important, Failure Analysis [FA] is still required, especially during technology learning when defect types are not well-known, and for understanding important fails, such as test escapes and system/field fails. CMOS technology migration is severely challenging the traditional FA process. Fault isolation, de-processing, and physical characterization will increasingly be too slow and difficult for routine analysis. To keep pace, improvements and breakthroughs to existing tools/techniques are required. FA capability gaps are detailed in the following prioritized list.

1. **Circuit node probing**—In-chamber (SEM) and atomic-force-microscope probing at the nanoscale is now required to characterize minimum sized transistors and SRAM cells at first contact level. This is driven by the need to measure increasingly subtle and non-visible defects and individual transistor parameters. Major limitations include adequate SEM image quality at low accelerating voltages required to avoid device damage, probe drift, probe contact resistance.
2. **Fault isolation precision**—Present tools for isolation of electrical faults and defects in die are severely limited by sensitivity and spatial resolution. For example, a majority of techniques (e.g., TIVA, PEM) rely on infrared light that is by definition constrained to about one micron spatial resolution – already more than an order of magnitude worse than minimum feature sizes. SEM or FIB based methods (e.g., voltage contrast, EBIC) have much better resolution but only on exposed conductors and thus are not effective for global isolation on fully processed devices. Near-field solutions are available/emerging (e.g. photonic, magnetic, AFM) but techniques are limited to samples with surface or near-surface features exposed. As such in-situ high-precision, non-invasive probing capability is required to provide electrical stimulus to multiple nano-scale conductors for tracing signals back to root-cause faults.
3. **Packaging analysis**—Fault isolation and imaging of package level defects is impacted by new and emerging technologies like organic laminates, chip-scale packaging, stacked die, and package on package configurations. Improvements in key methods like magnetic current imaging, X-ray tomography, and CSAM are critical, as are continued emphasis on test and fixtures.
4. **CAD/EDA tools**—Failure analysis is heavily reliant on scan methodology to lead directly to a fault location or to complement other localization tools. Continued improvements in accuracy, applicability to ‘soft’ defects and AC defects, and seamless integration are required. CAD navigation must be both spatial and time-based, i.e., linked to simulation waveforms.

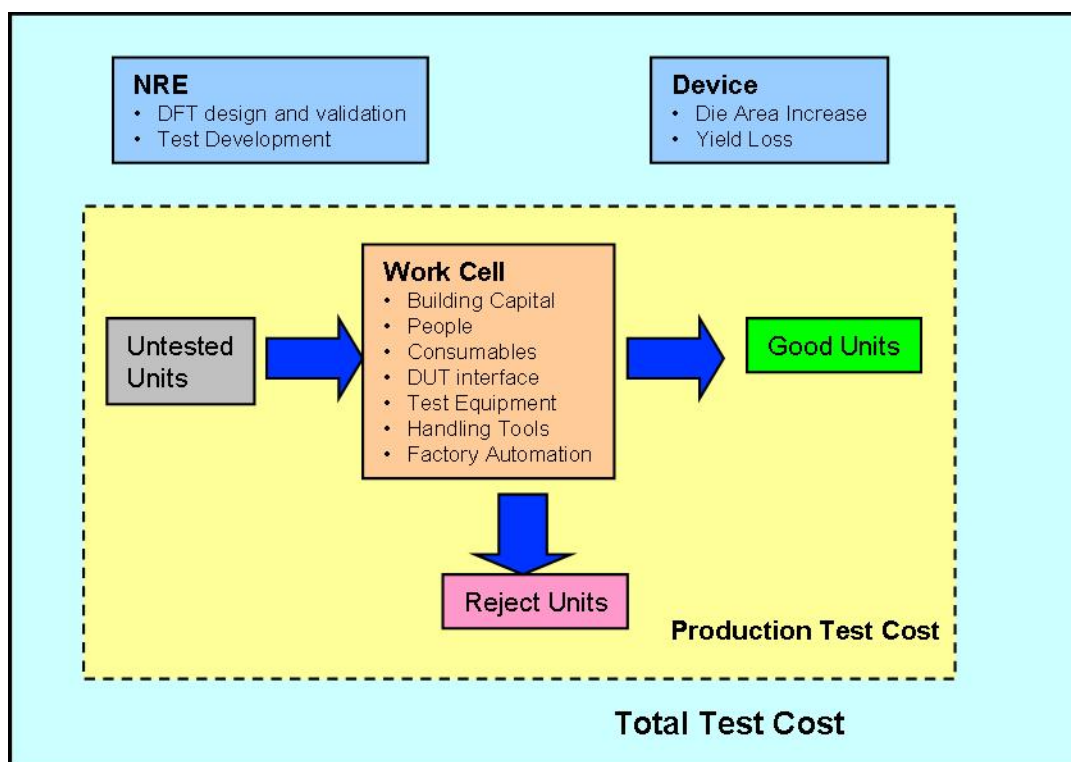
12 Test and Test Equipment

5. **On-chip timing measurement**—Lower supply voltages are causing hot electron based photon emission to exponentially drop in intensity and shift to longer wavelengths. Improvements in time resolved emission (TRE) technology and solid immersion lens (SIL) optics are required to maintain capability. Emerging laser-based probes like laser voltage probing (LVP) are promising. Rapidly increasing active power is also challenging cooling solutions to not compromise optical access to the chip backside or induce vibration. Unless improvements are made in these areas, a radically different technology will be required.
6. **Sample Prep and Fixtures**—A number of developments in packaging have made backside access for timing and fault isolation difficult to impossible. For example, control of stress-relief induced warpage and cracking is required to enable package and die preparation and fixtures for tools like TRE and OBIRCH. Stacked die packages and 3D chips will further challenge sample preparation methods.

The development of new capabilities for failure analysis has become increasingly expensive and high-risk, especially as throughput decreases and Failure Analysis value-proposition goes down. As a result, tool installation is more centralized, reducing total market potential and attractiveness, especially for smaller suppliers and start-ups. The existing model of incidental (or accidental) tool development and R&D investment driven by one or two companies may need to be replaced by a shared consortium whereby cost and risk are spread across a number of stakeholders.

TEST COST FOCUS TOPIC

Significant progress continues in the reduction of manufacturing test cost, however much work remains ahead. Semiconductor test technology continues the trend towards higher channel integration and higher degrees of multi-site testing, enabled by advanced probe card technologies, new handling technologies, and design for testability techniques. Evidence of this trend is clear in the growing number of low performance ATE systems and (BOST) load-board solutions in standalone memory test and logic test. Even though a lot of efforts reduce cost of test, 30% of the respondents already consider the cost of test as one of their biggest concern and 85% expect cost of test to become their biggest concern going forward. Significant work remains to translate similar improvements to the broad market of analog and RF. Current DFT techniques are point solutions in certain areas but not generic solutions. The continuing focus on cost of test will result in a better understanding of cost trade-offs between test methodologies, ATE architectures, and distributed test across



multiple insertions among other considerations, resulting in overall test cost reduction.

Figure TST1 Test Cost Drivers

The cost of semiconductor test to the organization has many drivers as shown in Figure TST1. To better understand various aspects related to cost of test, ITRS conducted an extensive survey in 2009. Few market participants deploy very elaborate cost models beyond the major metrics cost per unit, cost as percentage of total product cost, and the cost per second. The metrics such as total capital expenditures or the test cost per transistor or bit are not used to evaluate the cost effectiveness of test. The importance of the specific drivers varies substantially from device to device. Test development costs are more important for products with a lower volume. The cost of DFT area depends on whether the product is pad limited or core limited. The acceptable cost of test is very market specific and must be determined by balancing the value of test with its cost. The survey demonstrates that the main value of test is to obtain quality and support the yield and yield ramp-up. Most participants consider their quality best in class (57%). Moreover, more than a third of the participants feel that test is important to the competitiveness of their products (i.e., higher quality products or lower cost products). Only 6% of the participants don't emphasize the potential value of test but declare it a necessary evil.

The participants of the ITRS survey identify the ATE capital expenditures and interface expenditures as the two major cost drivers (22% vs. 21%), followed by test program development and silicon debug. Moreover, going forward, the group identifies new defects and reliability problems, device performance metrics, Known-Good-Die requirements, and the test requirements due to 3D stacking as the major cost drivers

CURRENT TOP COST DRIVERS

1. ATE Capital Expenditures
2. ATE Interface Expenditures
3. Cost of Test Program Development
4. Test Time and Test Coverage

FUTURE COST DRIVERS

1. New Defects and Reliability Problems
2. Device Performance Metrics
3. Known Good Die Requirement
4. Test Requirements of Packaging

ATE capital cost has traditionally been measured using a simple cost-per-digital pin approach. Although this is a convenient metric, it is misleading because it ignores base system costs associated with equipment infrastructure and central instruments as well as the scaling that occurs with reducing pin-counts and number of sites. Moreover, it is not aligned with the current trend in ATE platforms, where the same base infrastructure can be used for very diverse sets of test channels. The following equation expresses test cell capital cost in terms of the relevant cost drivers of future test technology:

$$C_{CELL} = C_{BASE} + C_{INTERFACE} + C_{POWER-SUPPLIES} + C_{TEST-CHANNELS} + C_{OTHER}$$

In this equation, C_{BASE} equals the base cost of a test system with zero pins/channels (e.g., it includes the cost of the mechanical infrastructure, back-plane, tester operating system software, and central instruments). $C_{INTERFACE}$ includes all costs required for interfacing with the device, e.g., the cost of interface electronics, sockets, and probe-cards (including spare probe-cards). C_{POWER} equals the cost of the power supplies. $C_{TEST-CHANNELS}$ is equal to the cost of the instruments (such as digital, analog, RF, memory test instruments). C_{OTHER} includes the remaining costs (e.g., floor-space). Practical considerations may limit the overall performance breadth that can be cost-effectively achieved by a given C_{BASE} infrastructure and should be taken into consideration in the overall test cell planning. For example, a low-end system may have an air-cooled infrastructure, whereas the high-end system will use liquid cooling. Test scenarios are evaluated by dividing the capital cost and performance metrics. For example, an important figure of merit is the *Units per Hour per Cost (UPH/\$M)*, that is, the number of shipped devices per hour (*throughput*) over the total cost.

Figure TST2 indicates the rapidly rising interface cost—which must be contained over time to avoid dominating the overall test cell cost. The next sections will describe these cost trends in more detail.

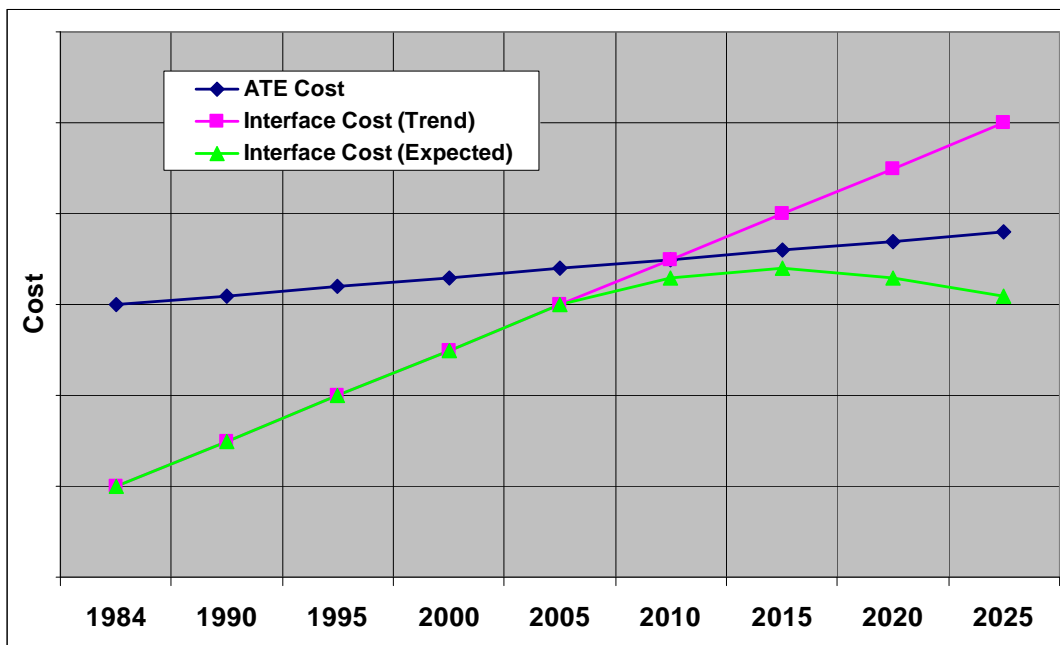


Figure TST2 Test Cell Cost / Unit versus Interface Cost Trend

A web survey was conducted to determine the top test cost reduction techniques that are currently used and those that may be deployed in the future. The results were:

CURRENTLY DEPLOYED COST REDUCTION TECHNIQUES

1. Multi-site & reduced pin-count
2. Structural Test & Scan
3. Compression/BIST/DFT
4. Yield Learning
5. Concurrent Test

COST REDUCTION TECHNIQUES THAT MAY BE DEPLOYED IN THE FUTURE

1. Wafer-level at-speed testing
2. Advanced embedded instruments
3. Adaptive Test
4. New contacting technologies
5. Build-in Fault Tolerance

BASE COST TREND

The total base is expected to decrease slightly over time. Platform strategies will extend the lifetime of the base infrastructure. Moreover, cost may move from the base infrastructure to the instruments. Multi-site test increases throughput and distributes the base cost across multiple dies, thereby reducing the base cost per site (and making the base cost less of a concern). For successful cost scaling using multi-site test it is important that the ATE infrastructure allows dedicated resources because shared resources may limit throughput. The trend of massive parallel test in memory will continue. Moreover, new probe card technologies and handler technologies, will enable massive parallel test in other segments (for both wafer and package test).

CHANNEL COST TREND

Continuing reduction in channel cost is essential for successful cost scaling using multi-site test: A dominating channel cost per site reduces the advantage of distributing the base-cost among many sites, whereas sharing expensive channels across multiple sites limits throughput. The channel cost is expected to decrease through continued integration within the tester electronics, and also by increased DFT adoption that reduces the ATE pin's performance requirements. Additionally, reduced pin-count test strategies utilizing small test ports can reduce the channel cost per site.

The relatively high cost of analog and RF test instruments, and the long test times associated with testing of these circuits, remain key challenges. DFT methodologies for analog and mixed-signal test are required.

The cost of testing high-speed I/O is becoming significant. In telecom applications, SONET data rates will increase from 2.5 Gb/s to the range of 10–40 Gb/s. In computing applications, serial ATA will increase from 1.5 Gb/s to 6 Gb/s in 2008, and PCIe will increase from 2.5 Gb/s to 5 Gb/s around 2008. Moreover, techniques like simultaneous bi-directional signaling may become significant, adding to the test cost. Looking forward, high-speed I/O DFT techniques and new test methodologies will become a more important part of the solution to control the cost of test.

POWER COST TREND

With increasing degrees of multi-site, the cost of power supplies will increase. Especially if reduced pin-count techniques are deployed, the power cost per site may dominate the channel cost per site. The cost increase in power supplies may be contained by innovations in power supply and power delivery technology. Note that some DFT techniques increase demand on power supplies to achieve shorter test times.

INTERFACE COST TREND

Controlling the interface cost is essential for successful cost scaling using multi-site test: A dominating interface cost that increases exponential with the number of sites may defeat the purpose of increasing the number of sites. The interface cost becomes very challenging with high bandwidth (2 Gbit/s) and/or high multi-sites (128 sites). There is a need to develop consistent cost models that cover the wide range of probe card technologies in the market place. Long probe card lead times cause significant cost problems, especially for the advanced technologies. Probe card lead times will be reduced by a factor of 2 within this roadmap’s horizon. For certain products, it may become economical to skip wafer test or only do a simple low performance test. Looking forward, high-speed I/O DFT techniques will become a more pervasive part of the solution to control interface cost.

MULTI-SITE TREND

As discussed in the previous sections, the most important way to reduce cost of test is increasing the number of sites. The effectiveness of increasing the number of sites is limited by (1) a high interface cost, (2) a high channel and/or power cost, and (3) a low multi-site efficiency M :

$$M = 1 - \frac{(T_N - T_1)}{(N - 1)T_1}$$

Where N is the number of devices tested in parallel ($N > 1$), T_1 is the test-time for testing one device, and T_N is the test time for testing N devices in parallel. For example, a device with a test time T_1 of 10 seconds tested using $N=32$ sites in $T_N=16$ seconds has a multi-site efficiency of 98.06%. Hence, for each additional device tested in parallel there is an overhead of $(1-M) = 1.94\%$.

Table TST2 Multi-site Test for Product Segments

Shared ATE resources that reduce the channel cost may cause a low multi-site efficiency (for example, this is apparent in mixed-signal / RF test). Moreover, as one continues to increase the number of sites, a low multi-site efficiency has a larger impact on the cost of test. For example, 98% efficiency is adequate for testing two and four sites. However, much higher efficiency is needed for testing 32 sites. At 98% efficiency going from testing a single site to testing four sites will increase a 10s test time to 10.8s. However, going from testing a single site to testing 32 sites will increase a 10s test time to 16.4s, that is, significantly reducing the potential advantage of multi-site test (Figure TST3).

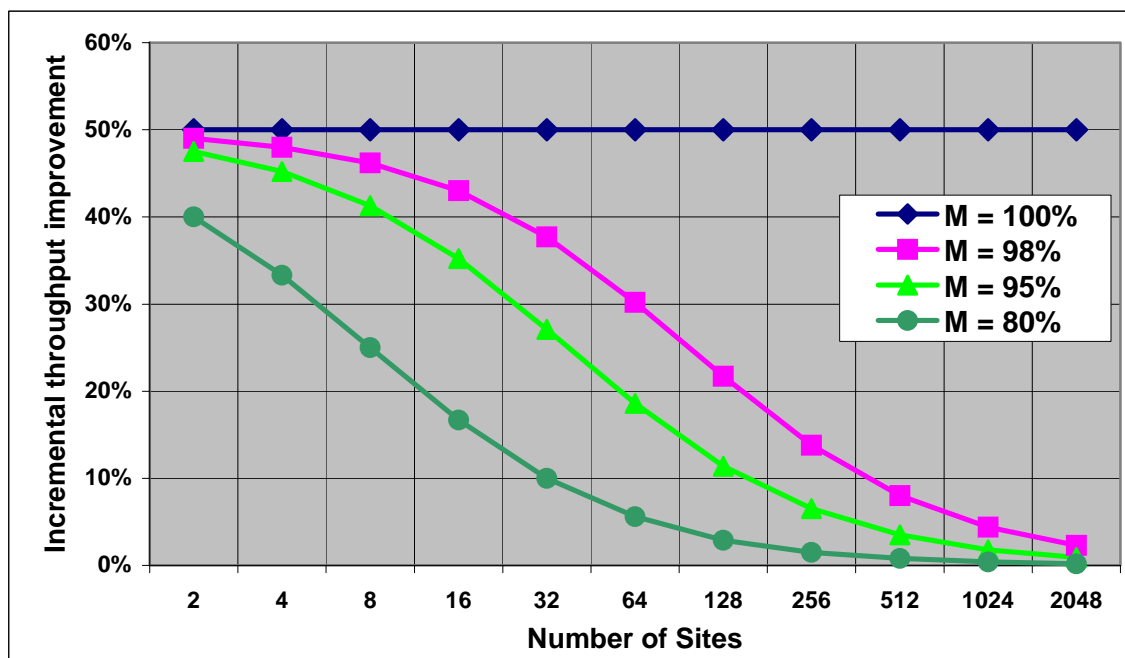


Figure TST3 Importance of Multi-site Efficiency in Massive Parallel Test

Table TST2 presents the expected trend in the number of sites for an arbitrary device in each product segment. A custom economic model should be deployed to identify the optimized roadmap to scale test cost of a custom device. Note that there are multiple trajectories/approaches that achieve the test cost targets.

OTHER COST TRENDS

The dramatic increase in SiP solutions that integrate memory, logic, and RF have further increased pressure to reduce the cost of test for mixed technology designs as well as improve the wafer test quality of KGD.

Test development time and cost will be reduced further by DFT techniques; test standards (to support test content reuse, test program inter-operability, compression, and manufacturing agility); automatic generation of test patterns (such as structural test approaches), and the programs that use them.

Stand alone memory DFT, like BIST and built-in self repair (BISR), will become pervasive as DFT will be essential to control the cost of test. For certain segments, new manufacturing process flows may become economically justified. For example, BIST can be used to test low performance followed by a second insertion on a high performance tester.

Based upon the test survey, the top 5 techniques to reduce cost of test are reduced-pin count and multi-site testing, structural test & scan, compression/BIST/DFT, yield learning and concurrent test. Survey respondents did not rank adaptive test, embedded instruments, nor standardization efforts (e.g., CAST) in the top 5 cost reduction techniques currently deployed. However, adaptive test, new contacting technologies (e.g., MEMS, contactless probing), wafer level at speed testing, build-in fault tolerance, and embedded instruments were identified as techniques that might offer test cost savings in the future.

IMPORTANT AREAS OF CONCERN

- Increases in the number of sites place severe demands on ATE architectures and probe card technologies. Research and development must continue to bring to market cost-effective probe technologies directed at product offerings and multi-site test trends. Lead times for advanced probe cards are an issue. RF radiation may become an issue for massive RF parallel test.
- The relatively high cost of analog and RF test instruments and the long test times associated with testing these circuits remain key challenges. To enable parallel test, multiple instruments are required with fast execution of DSP test algorithms like fast Fourier transform (FFT) or other correlation tests. A secondary consideration for mixed-signal multi-site test is the load board circuits required for package test, especially for complex packages. DFT techniques for mixed-signal and RF devices remain development needs. Because of the high costs, mixed-signal resources (and post-processing steps) are frequently shared, significantly reducing multi-site efficiency.

- High-speed serial interfaces are penetrating ASIC and SoC markets. Jitter testing results in high test-times and equipment capital costs. As the number of interfaces increase, the cost problem will increase linearly. New test methods need to be developed to manage the cost scaling.
- New DFT techniques are required to test pins that are not connected during massive parallel test. Moreover, in setups where reduced pin-count techniques are deployed, power supply cost may dominate channel cost. The power supply cost may need to be contained by innovations in power supply and power delivery technology.
- Although a low multi-site efficiency is not a problem for low site counts, for massive parallel test the impact can defeat the purpose of multi-site test. To continue multi-site scaling, ATE architectures may require dedicated channels/instruments per site, because shared channels/instruments may limit multi-site efficiency. To allow dedicated channels/instruments in a cost effective way, the channel cost needs to be reduced for certain segments.

ADAPTIVE TEST

The ever growing need to dynamically and cost effectively tailor IC testing to discriminately manage manufacturing process variations while simultaneously eliminating redundancies (within constraints of pre-defined rules) to a level far beyond what is achievable through the static optimization predominantly applied in the industry, will increasingly drive the adoption of dynamic “adaptive test” methodologies. Benefits arising from Adaptive Test methods include lower test costs, higher yields, better quality & reliability and an improved data collection environment for more effective and rapid yield learning. By virtue of its evolutionary nature, adaptive test improves detection of maverick or defective parts, increases efficiency of production test floors and better meets customer expectations.

ADAPTIVE TEST DEFINITION

Adaptive Test is a broad term used to describe methods that change test conditions, test flow, test content and test limits (potentially at the die/unit or sub-die level) based on manufacturing test data and statistical data analysis. This includes feed-forward data from inline test and early test steps to later test steps and feed-back data from post-test statistical analysis that is used to optimize testing of future products. Adaptive Test also includes real-time data analysis that can perform Statistical Process Control (SPC) and adjust test limits and content during production testing on-the-fly. (e.g., Parts Average Testing algorithms) Analysis driving Adaptive Test will occur both real-time (in parallel with testing), near-time (at the end of sample testing and at the end of wafer test and lot test) and off-line. Although some simple applications have been applied for some time, Adaptive Test will increasingly be applied and will require updated software algorithms and improved statistical analysis methods and expanded database infrastructure.

A high-level flow diagram of Adaptive Test is shown in Figure TST4. Note that each test step has “real-time” and “near-time” (e.g., end of wafer) analysis capability and can access information from a shared database and off-line analysis system. “Real-time” analysis capability could be during the test of an individual die, between consecutive dies or within a very short time window (e.g., within 5 seconds). This real-time analysis must not significantly slow down testing. This system enables full feed-forward and feed-backward of data and analysis results from any test step to any other one. Although not shown in this figure, this system would enable multiple product test flows (or routings) for a product based on data analysis. Note that this flow is not restricted to IDMs – it is envisioned to be applied across multiple companies through the supply chain.

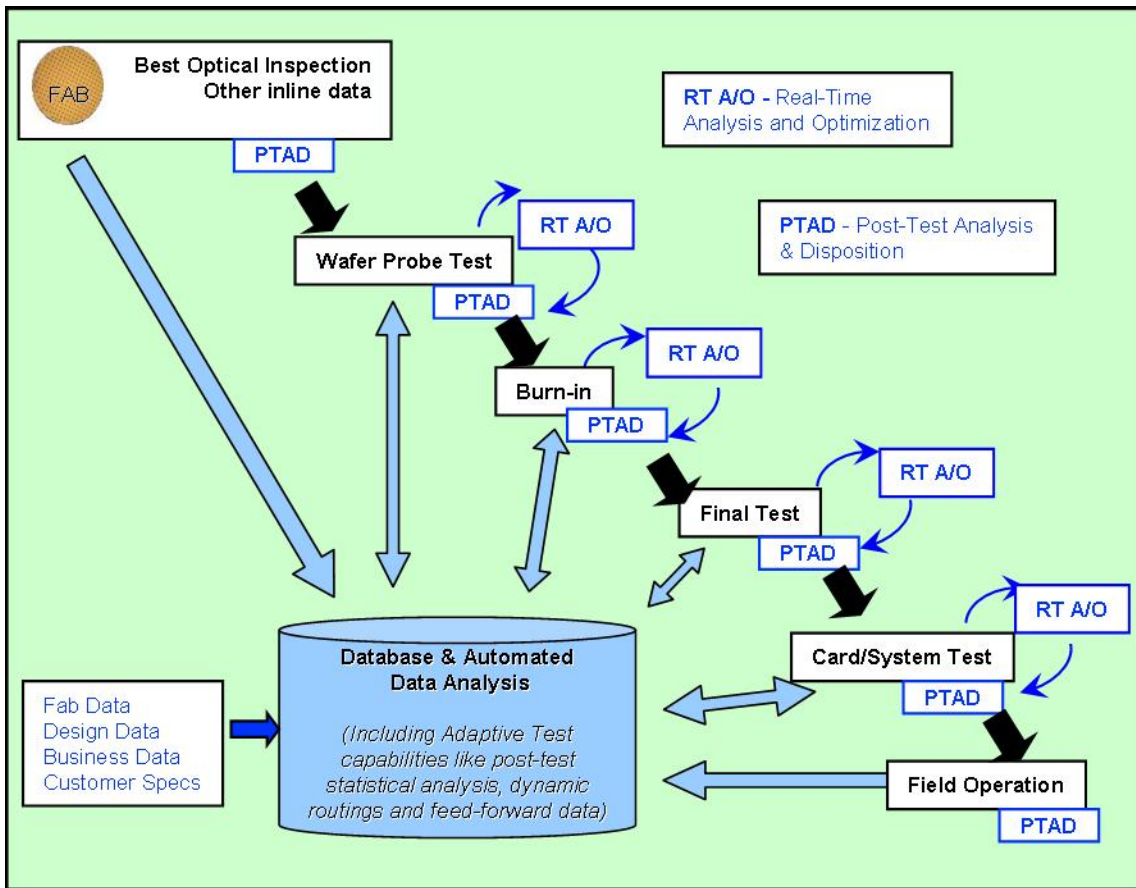


Figure TST4 High-level diagram of Adaptive Test Flow

PROCESS VARIABILITY – ITS IMPACT ON TESTING

The effects of process variation remain the overarching theme of the ITRS. For example, the ITRS Design chapter uses variation/variability and uncertainty over one hundred times. If design could control for all variation and uncertainty, then test would leverage these controls and largely ignore variation at test. Similarly, if systems could accommodate the variability of its components, then conventional test guard-banding would deliver useable parts. These levels of control and accommodation are unknown and hence test must identify and screen for variation by designing new test methods capable of constantly adjusting test conditions, test content and test flows.

Table TST3 “Stages” of Adaptive Test

<i>Stages of Adaptive Test</i>	<i>Implementation</i>
<i>No Adaptive Test</i>	<ul style="list-style-type: none"> • <i>Static Adaptation only -- test limits and test content and test flows are statically set and occasionally updated based on off-line analysis. (perhaps every few months)</i>
EMERGING METHODS <i>(The following three stages may not necessarily be implemented in the order shown.)</i>	
<i>Adaptive Test Limit Setting</i>	<ul style="list-style-type: none"> • <i>A procedure or algorithm is defined which sets test limits. (algorithm is statically defined)</i> • <i>During test execution, real-time (and possibly historical) DUT response data is used as input to the limit setting algorithms.</i> • <i>Many types of on-tester and off-tester outlier screening techniques can be applied at this stage.</i>
<i>Data Feed-forward and Feed-backward</i>	<ul style="list-style-type: none"> • <i>Data from any test step is available to be used at any other test step.</i> • <i>Data analysis of historical data at one test step is used to set limits, test content and flows at any other test step.</i>
<i>Adaptive Test Content & Test Flows</i>	<ul style="list-style-type: none"> • <i>A procedure or algorithm is defined which sets test content. (algorithm is statically defined)</i> • <i>During test execution, real-time (and possibly historical) DUT response data is used as input to the test content/flow setting algorithms.</i> • <i>Multiple Test flows can be dynamically chosen based on automated statistical data analysis.</i>
FUTURE METHODS	
<i>Adaptive Test based on DUT Response Models</i>	<ul style="list-style-type: none"> • <i>A “DUT response model” (which has many different limit setting schemes) is used instead of a static algorithm.</i> • <i>Over time, Adaptive Test will expand its input source from DUT responses to other types of process and business data.</i> • <i>In effect, there is a decision-tree structure that in the extreme could result in each DUT being uniquely tested by a flow dictated by the observed response which may include data from multiple insertions.</i> • <i>The more likely outcome is DUTs are identified as members of a common sub-group and the test flow is dynamically adjusted using the group of DUTs and their common response.</i>

EXAMPLE APPLICATIONS

Below is a short list of Adaptive Test example applications. These applications are shifting test execution from manual and static methods to automatic adaptation at any level (e.g., lot, wafer, die) and minimizing direct input by product and test engineering staff.

- 1. Real-time monitoring of test results that can dynamically change test flows**—adding or removing tests and selectively performing more yield learning characterization and diagnostic data collection. This application supports real-time Statistical Process Control.
- 2. Post-test statistical analysis of test results**—(e.g., for entire wafer or lot) to better identify outliers or maverick events that could cause DPM or reliability failures. This post-test analysis can also be used to route target dies through unique test flows.
- 3. Feed-forward of test results from one test step to later test steps**—to optimize testing or to enable more focused screening. (e.g., inline testing to wafer test, wafer test to package-level testing, burn-in results to package-level testing, package-level testing to card/system testing)
- 4. Feed-back test results to drive changes for future devices**—for example, off-line analysis would be used to optimize test flows, test content and measurement routines. Off-line analysis combines data from many sources including historic data collected, test capacity, required turn-around times, DPM requirements, expected yields and parametrics.
- 5. Production test monitoring and alerts**—including real-time statistical analysis of all available data to enable more complex production test operations than has historically been performed. (e.g., real-time parametric data analysis to identify subtle parametric shifts that may be due to marginal wafer probe contacting)

DIRECTIONS FOR ADAPTIVE TEST IN THE NEXT 5–10 YEARS

Fundamentally, the goal of Adaptive Test is to fully automate the process of integrating data from all sources – design, models, manufacturing, historical, etc. – combining and reducing it, and making the most intelligent decision possible about how to expend test resources for the best outcome for every individual device. Work will continue on algorithm

20 Test and Test Equipment

development, integrated and peripheral instrumentation, data quality assurance, data communication and delivery, and flexible test flows and manufacturing floors, so that every die is screened appropriately and with the least cost.

A functional abstraction for Adaptive Test allows others to build new and expand upon existing Adaptive Test applications. By necessity, abstraction removes significant detail to highlight the essential characteristics of the system being described. Answers to two related questions form the key to understanding Adaptive Test; "What components are reviewed and approved by the in-house group that controls changes to the test program?" and "What components in the test flow are data-driven by the observed DUT response?" The abstraction isolates two functional elements of Test: DUT response modeling and limit setting. A test program executes combinations of the two to disposition the part, e.g., final binning or continues testing.

ADAPTIVE TEST BUILDING BLOCKS

Adaptive Test depends on a set of infrastructure building blocks to enable dynamic optimization of test content and/or flow. The key elements are:

- Data (various sources, may be real-time, historical or feed-forward, and structured to feed decision algorithms)
- Data Structure & Control Path (deliver targeted data to the analysis engine and decisions to the action point)
- Decision Algorithms (models to automatically adjust test limit, flow, or content based on input data.)
- Analysis Engine (Executes model/algorithm on the data structure to return control signals based on specific product parameters. The engine may be separate or embedded in test program or station controller software.)
- Station controller/test program hooks to communicate with or run the analysis engine and take action.

Adaptive Test will drive additional infrastructure and business processes **that must enable the following:**

- Full traceability (details of how Adaptive Test was applied for each product/part tested)
- Standard models and algorithms (Parts Average Testing, Linear models, Machine learning, sampling, ...)
- Recipe management (model form and parameters, target products/lots)
- Analysis functions (to quantify trade-offs in quality, reliability, test time, yield loss, risk assessment)

IMPLICATIONS OF ADAPTIVE TEST FOR TEST DATA AUTOMATION

- Test data storage for large volume of data coming from the complete manufacturing flow
- Real-time retrieval of historic test data (die level feed forward, real-time lot statistics, feed-back)
- Real-time data integrity validation
- File formats and data representation standards for response and traceability
- Algorithm representations

IMPLICATIONS OF ADAPTIVE TEST FOR ATE

- Define/create a (standard) interface at the ATE for a data path transfer of the real-time test inputs/results
- Create a (standard) real-time interface at the ATE for controlling the test flow and hardware configuration for an adapted test flow.
- Define the interface for adjusting test program content real-time (between units or lots).
- Automatic adjustments to hardware configuration for optimized test flow

Table TST4 Implications of Adaptive Test

Challenge	Required Direction
<i>IT Infrastructure</i>	<ul style="list-style-type: none"> • <i>Infrastructure to enable the Adaptive Test flow shown in Figure TST4.</i> • <i>End-to-end supply chain data integration – including data from Fabs, Test Houses and other Subcons.</i> • <i>Develop supply chain data integration and processes which automatically detect supply chain issues and implement corrective actions in near real-time.</i> • <i>Integrate multiple databases, flexible logistics system, full part tracking at each test step, and feed-forward/feed-backward data flows.</i>
<i>Traceability</i>	<ul style="list-style-type: none"> • <i>Enable full traceability of Adaptive Test parameters (limits, content, flows, rules) for each die. (accessible anytime in future)</i>
<i>Real-Time Communications</i>	<ul style="list-style-type: none"> • <i>Develop tester-to/from-data analysis engine communication – without significantly impacting test time.</i>
<i>Development of Improved Models & Algorithms</i>	<ul style="list-style-type: none"> • <i>Development of methods where the models are not fixed – instead the models are dynamically adjusted based on DUT responses.</i> • <i>Develop peripheral coverage metrics and associated quality impact of dropped or modified tests.</i> • <i>More encompassing fault coverage metrics are required – particularly for analog circuits.</i>

TEST TECHNOLOGY REQUIREMENTS

INTRODUCTION

Over the past 25 years, semiconductor test technology requirements have been driven primarily by relentlessly increasing performance and transistor counts. A fundamental shift is underway driven by the emergence of new market demands (for example, mobility, security, ease of use, ease of system management, low power, etc.). This in turn is fueling the integration of different semiconductor technologies in more ways and in a greater set of applications than ever before. This in itself is a huge challenge to test as it is ultimately the application requirements or specifications that determine test technology requirements, but it would be impossible to capture a comprehensive set of applications and their associated test requirements trends within this chapter. Therefore, core semiconductor technology building blocks have been identified to provide a framework for describing the test challenges and trends associated with each core technology as well as for describing the test challenges associated with integrating these core technologies together either as a SoC or a SiP.

Each core semiconductor technology has certain applications associated with it and some of these will be used as a basis for extracting long-term trends. In particular, the ITRS publishes key technology attribute trends for CPU, ASIC, DRAM, and Flash memory. These will be referenced where appropriate in the core technology sections. Figure TST5 shows the core semiconductor technologies addressed in this chapter as well as examples of associated applications. The application mapping is intentionally loose as many of the examples listed may contain multiple core technologies. The core technologies are differentiated mainly by their inherent functional differences and thus their different test requirements. Two emerging core technologies that are not included in this revision are MEMS and optical.

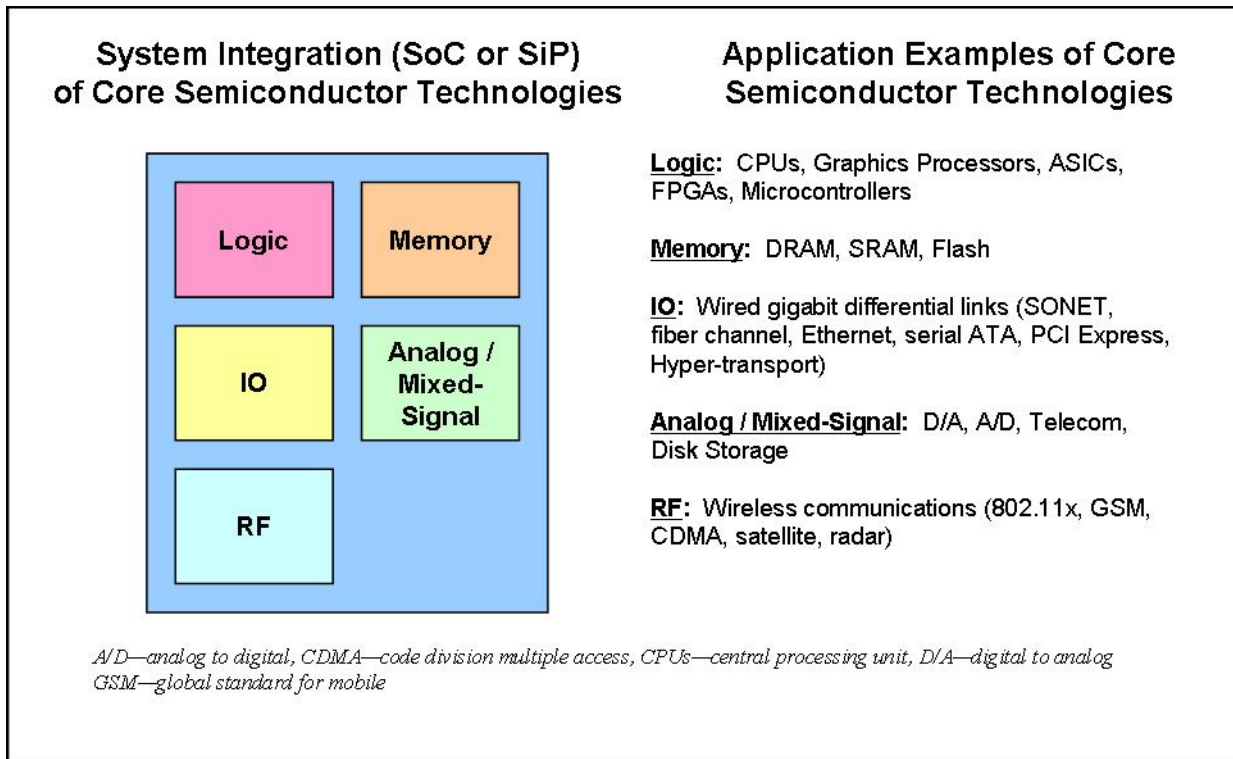


Figure TST5 Organization of Cores for System Integration and Applications

In the recent past, these core semiconductor technologies and applications have demanded distinctly different test solutions, each having specific test equipment and interface tooling markets. Increasing integration is blurring these boundaries. It has also raised the stakes for DFT as successful integration is determined not just by “can it be done?” but also “can the integrated whole be tested economically?” The remainder of the Test Technology Requirements section will address the test challenges associated with increasing integration followed by the test requirements of each constituent core technology.

SYSTEM INTEGRATION—SoC AND SiP TEST CHALLENGES AND IMPLICATIONS

While possibly equivalent in theory, SoC and SiP are very different from each other in terms of which technologies tend to be more easily integrated in package versus on chip and each has very different test implications. Recent advancements in assembly and packaging technologies coupled with the difficulty of optimizing the same wafer fabrication process for different core semiconductor technologies have provided a lot of momentum for SiP, causing some to forecast that SiP will be dominant. It may be that wafer fabrication process improvements and design/DFT needs could push SoC to the fore front or there could be hybrids of the two. One thing is clear: integration is a trend that will continue. The only questions are how fast and in what forms. The next two sections will discuss the test challenges and implications associated with SoC and SiP respectively.

SYSTEM ON A CHIP

A SoC design consists of multiple individual design blocks, or cores, using different technologies (logic, memory, analog, RF, etc). This assortment requires a diversity of solutions to test the specific technologies corresponding to these embedded cores. Increasingly, SoC design is relying on a database of pre-existing IP cores that encapsulate the design itself, its embedded test solution, and the interface to other cores. SoC test implies a highly structured DFT infrastructure to observe and control individual core test solutions. SoC test must include the appropriate combination of these solutions associated with individual cores, core test access, and full-chip testing. One fundamental challenge of SoC test is the need to combine test requirements from multiple sources with differing testability methods. Another challenge of SoC test is to highly optimization both quality and cost. The overall quality and cost of SoC should be evaluated and be adjusted to an acceptable level for customers. Some hierarchical approaches or parallel approaches will be required.

The quantitative trends and requirements in *Table TST5* were estimated according to the trends that are shown in the SoC-PE driver model of the ITRS Systems Driver Chapter. The trends include the number of transistors, flip-flops, and

memory bits. Although the estimation has been done assuming a specific SoC model, the requirements should also be applicable for various types of SoC designs.

REQUIREMENTS FOR LOGIC CORES

Sophisticated DFT methods such as random pattern logic BIST or compressed deterministic pattern test are required to reduce large amount of test data for logic cores. The adopted method should consider the pros and cons regarding DFT area investment, design rule restrictions, and associated ATE cost. DFT area mainly consists of the test controllers and test points, which can be kept constant over time by using a hierarchical design approach.

A tradeoff between test quality and test cost is a great concern. ATPG should support not only stuck-at and transition faults but also small delay and other defect-based faults to achieve a high-level of test quality. Test pattern count will increase over the roadmap as logic transistor count increases. To avoid rising test cost, the test application time per gate should be kept constant. Therefore various approaches, such as test pattern reduction, scan chain length reduction and scalable speed-up of scan shift frequency, should be investigated.

Another problem caused by the increase of test patterns is the volume of test data. Even assuming tester memory size will be doubled every three years, a high test data compression ratio will be required in near future. Therefore test data reduction will remain a serious issue that must be tackled. One possible solution for simultaneous reduction of test application time and test data volume is the repeated use of the same IP cores in a design. Figure TST7 shows the impact of the repeated use of constant test application time and test data reduction requirement using same IP cores. It shows the constant test application time requirement can be relaxed to a certain extent, but not enough to solve the problem completely. Therefore, other approaches, such as the combination of BIST and compressed deterministic pattern test or an innovative DFT scheme should be also targeted.

The increase of power domains may require some additional test patterns. However, the increase of test patterns will be linear to the number of power domains, so it won't have severe impact on overall test pattern counts. Nevertheless the increase of power domains may deteriorate the effectiveness of simultaneous test of identical cores. The impact of this effect will be investigated for future editions of the roadmap.

First Year of IC Production	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Required compression	80	200	300	500	750	1300	2000	3300	4800	7300	12000	20000	35000	67000	83000	104000
1-dim: Test-cube compression (100X)																
2-dim: Spatial compression(500X)																
3-dim: Time correlations compression(1000X)																
Multi-dim: compression(5000X)																

This legend indicates the time during which research, development, and qualification/pre-production should be taking place for the solution.





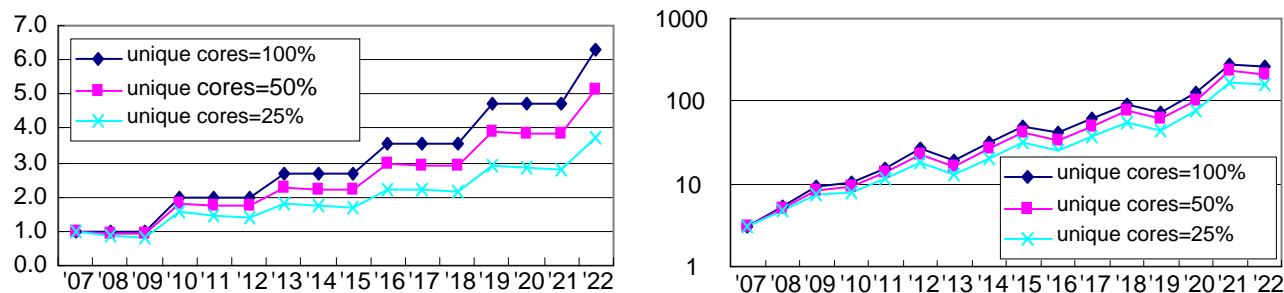
Research Required 
 Development Underway 
 Qualification / Pre-Production 
 Continuous Improvement 

Figure TST6 Potential Solution for the Problem of the Test Data Increase

Figure TST6 shows a potential solution for the problem of the test data increase. The current compression technologies mainly utilize the fact that each test vector has many ‘X-values’ (don’t care bits that don’t contribute to the increase of test coverage), and more than 100X compressions have been achieved. However, even a 100x compression won’t be enough as shown in Table TST5 SoC, therefore, more sophisticated technologies will be required in future. The similarity of test vectors applied on scan chains will allow a chance of achieving higher compression ratio. The similarity of test vectors applied in time space may also allow further compression. Thus, utilizing multi-dimensional similarity will be a potential solution.

24 Test and Test Equipment



(a) Test Time per Gate (Relative)

(b) Required Test Data Volume Compression

Figure TST7 Impact of Repeated Use of Same Cores

It should be noted that keeping the test application time per gate constant does not immediately mean a stable test application cost. Therefore some approaches to reduce the ATE cost, such as the increase of parallel sites, the use of low-cost ATE, or a speed up of test, are also required to establish the scalable reduction of test cost per transistor.

First Year of IC Production	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Required Test time reduction	1x	2x	2x	2x	2.5x	2.5x	2.5x	3.2x	3.2x	3.2x	4.2x	4.2x	4.2x	5.6x	5.6x	5.6x
Multi-site Test																
Core-Parallel Test																
Test Vector Reduction (include compression)																
Test per clock																

This legend indicates the time during which research, development, and qualification/pre-production should be taking place for the solution.

Research Required	Black
Development Underway	Cyan
Qualification / Pre-Production	White
Continuous Improvement	Grey

Figure TST8 Potential Solutions for Test Time Reduction

Figure TST8 shows a potential solution for the problem of test time increase. Increasing the number of multi-site testing is an effective solution for reducing test time per chip. Concurrent parallel test in the core hierarchy has the potential of reducing test time. ATPG/DFT level reduction technologies should be developed in future. “Test per clock” means a test methodology that is quite different from the scan test (i.e. non-scan test). The test is done at each clock pulse and the scan shift operation is not needed. There are some researches regarding the methodology, however, more research is required for industrial use.

High-level design languages are being used to improve design efficiency, and it is preferable that DFT is applied at high-level design phase. DFT design rule checking is already available to some extent. Testability analysis, estimation of fault coverage, and DFT synthesis in high-level design that includes non-scan-design approaches are required in the next stage. Yield-loss is a concern. As test patterns excite all possible faults on DUT, it will lead to the excessive switching activity, which does not occur in normal functional operation. This will cause excessive power consumption making the functional operation unstable, and eventually may make the test fail, which will cause over-kill. In addition, signal integrity issues due to resistive drop or crosstalk can also occur which would make the functional operation unstable or marginal and eventually cause failures. Therefore, predictability and control of power consumption and noise during DFT design is required.

Table TST5 System on Chip Test Requirements

REQUIREMENTS FOR EMBEDDED MEMORY CORES

As process technology advances, the total capacity of memory bits increases and will cause an increase in area investment for BIST, repair and diagnostic circuitry. As the density and operating frequency of memory cores grow, memory DFT technology will require functionality and performance innovations as follows:

- To cover new types of defects that appear in a nanometer process, testing algorithms should evolve from a generic fixed one to either a selective combination of generic algorithms and test conditions, or a dedicated optimal algorithm for a given memory design and defect set. Furthermore, a highly programmable BIST that enables flexible composition of the testing algorithms should be developed.
- Practical embedded repair technologies, such as built-in redundancy allocation (BIRA) and built-in self-repair (BISR) technologies should be developed. BIRA analyzes the BIST results and allocate redundancy elements for yield improvement. BISR performs the actual reconfiguration (hard-repairing) on-chip.
- On-line acquisition of failure information is becoming essential for yield learning. Failure types, such as bit, row, and column failures or combinations of those failure types, need to be distinguished on-chip without dumping a large quantity of test results. A built-in self-diagnostic (BISD) feature could enable this functionality by analyzing memory output and pass the results to ATE.
- All the above features need to be implemented in a compact size, and operate at the system frequency.

Based on *Table TST5*, the embedded memory test, repair and diagnostic logic size will be up to 35 K gates per million bits in 2007. This contains BIST, BIRA, BISR, and BISD logic, but does not include the repair programming devices such as optical or electrical fuses. The ratio of area investment to the number of memory bits should not increase over the next decade. This requirement is not easily achievable. In particular, when the memory redundancy architecture becomes more complex, it will be difficult to implement the repair analysis with a small amount of logic. Therefore, a breakthrough in BIST, repair and diagnostic architecture is required. Dividing BIST, repair and diagnostic logic of memory cores into a high-speed portion and a low-speed portion might reduce the area investment and turn-around-time for timing closure. A high-speed portion that consists of counters and data comparators can be embedded in the memory cores which will relax the restrictions for system speed operation in testing mode. A Low-speed portion that consists of the logic for scheduling, pattern programming, etc. can be either designed to operate at low-speed or shared by multiple memory cores, which will reduce area investment and ease logical and physical design work. To automate such a design methodology, the interface for high-speed memory test operation must be unified and standardized over all necessary memory types and configurations. A lot of small-size memory cores are very often seen in modern SoC; however, they require a larger amount of DFT gates than for a single memory core of the same total bit count. Therefore consolidating memory cores into to a smaller number of memory blocks can reduce memory DFT area investment drastically. Testability-aware high-level synthesis should realize this feature in the memory cell allocation process and consider the parallelism of memory access on system operation.

REQUIREMENTS FOR INTEGRATION OF SOC

Reuse of IP core is the key issue for design efficiency. When an IP core is obtained from a third party provider, its predefined test solution must be adopted. Many EDA tools already leverage a standard format for logic cores (for example, IEEE1500¹); and this format must be prevailed and extended to other core types, such as analog cores. DFT-ATE interface is going to be standardized (for example, IEEE1450²), and it should include not only test vectors but also parametric factors. Test quality of each core is now evaluated using various types of fault coverage such as stuck-at fault, transition delay fault, or small delay fault coverage. A unified method to obtain overall test quality that integrates the test coverage of each core should be developed. Conventionally, functional test has been used to compensate structural test's quality. However, automated test for inter-core or core interface should be developed near future. SoC level diagnosis requires a systematic hierarchical diagnosis platform that is available for learning the limiting factors in a design or process (such as, systemic defects). It should hierarchically locate the defective core, defective part in the core, and the defective X-Y coordinate in the part. A menu of supported defect types must be enhanced to meet with the growing population of physical defects in the latest process technology. Smooth standardized interfaces of design tools with ATE or FA machines are also required. Volume diagnosis is required to collect consistent data across multiple products containing the same design cores, which is stored in a data base and is analyzed statistically using data mining methods. The menu of data items is very crucial for efficient yield learning, but it is a know-how issue now.

¹ 1500-2005 IEEE Standard Testability Method for Embedded Core-based Integrated Circuits.

² P1450.6, Draft Standard for Standard Test Interface Language (STIL) for Digital Test Vector Data—Core Test Language (CTL).

SYSTEM IN A PACKAGE

In contrast to SoC, SiP offers the option of testing components prior to integration. This is important since integrating one bad component could negate several good components in the SiP, severely limiting SiP yield. In addition, this component testing must typically be done at wafer probe test since integration occurs at assembly and packaging. A key challenge then is identifying good die prior to integration. The term “known good die,” or KGD, was coined during the mid-1990s to designate bare die that could be relied upon to exhibit the same quality and reliability as the equivalent single chip packaged device.

In most instances, testing and screening the device in a single chip package format achieves the outgoing quality and reliability figures for IC products shipping today. Wafer probe test is not generally suitable for performance sorting, reliability screening, or effective parallel contacting, so it is generally more efficient to do these tests at the package level using test and burn-in sockets, burn-in chambers, and load boards. Consequently, KGD processing implies that die will be up-binned at probe or with a subsequent insertion of die level tests and screens to meet acceptable quality and reliability targets. The key short term challenges are to determine the quality and reliability targets required in different market segments, develop cost effective tests and reliability screens that can be applied at the wafer or die level, and to develop quality and reliability methods that provide high confidence regarding quality and reliability levels achieved. Longer-term challenges will be to move to a complete self-test strategy with error detection and correction available in the end application.

SIP TESTING AND EQUIPMENT CHALLENGES

SiP products can present many unique challenges to backend manufacturing flows because SiP products often contain die from more than one supplier. This can create problems in the areas of:

- development of a package test strategy to realize both cost and DPM goals
- production flows to accommodate the necessary reliability screening methods (burn-in, voltage stress, etc) of diverse product/process technologies
- failure analysis methodologies for fault localization in order to resolve quality problems and systematic yield issues

SiP test at the package level closely resembles the test problems of complex SoC products, that is, a variety of IP, each with specialized test requirements, which must be consolidated into a single consistent test flow. In the case of SoC, because everything is on one chip and designed together, various block test strategies can be consolidated via the use of test shell wrappers, test control blocks, etc. using strategies such as defined in the IEEE 1500 specifications. In the case of SiP, die suppliers may be reluctant to provide information needed to access special test modes (sometimes considered confidential, especially for commodity memory products) and the individual die may not have the necessary test infrastructure overhead to implement test strategies commonly used for SoC.

Even in the case of SiPs that use only KGD, a certain amount of testing is necessary after final assembly to ensure that the die have been assembled properly. When final assembly may include die thinning and stacking, which can damage/change KGD die, additional testing may be necessary. For the case of fault localization, the ability to narrow the failure to a specific die, and further to a small region of that die, may require full understanding of the detailed test strategies for that die, even if not necessary in normal production..

In the case of reliability screens, some die may require burn-in while others may require only voltage stress. Stress conditions for one die may be inconsistent (or even detrimental) to other die in the same package. Resolution is more difficult since the different die in a SiP product often have totally different processes. One solution is to avoid reliability screens after final packaging but this can increase overall costs (for example, wafer level burn-in is typically more costly than package level burn-in).

When heterogeneous die are assembled into a multi-chip package, several test insertions on different platforms may be required to test the assembled module fully. The multiple test insertions may result in test escapes or yield fallout due to mechanical damage. New testing equipment will be required to accommodate contacting the top side of the package for package stacking. For wafer stacking technologies, better redundancy/repair technologies are needed so that the final stack can be “fixed” to achieve yield/cost targets. Design and production of electronic systems that can detect failed components and invoke redundant elements while in service is a key challenge for SiP reliability.

WAFER TESTING AND EQUIPMENT CHALLENGES/CONCERNS

The probe card technologies in common use today are less than ideal as a “final test” environment. Since much of the performance based speed critical, RF, delay and analog testing is presently performed at package level, a critical challenge for KGD processing is the development of cost-effective, production worthy, reliable and accurate methods of rapidly identifying devices that are defective or will fail early in an application before those devices are transferred to the next level assembly.

Test time for certain technologies, such as display drivers or state of the art DRAM is exceedingly large. Because of the limitations in the wafer probing process, the test throughput is much less than packaged components. The challenges for fully testing DRAM die in a cost effective manner at the wafer level include development of technology that can probe multiple die on a wafer without overlapping previously probed die or stepping off the wafer, and to avoid wasting test time and power on all previously rejected and obviously non-functional die.

WAFER TEST FOR RF DEVICES

A key challenge for applying KGD processes to RF die is development of high performance, fine pitch probe cards. Because of the small size of RF die, the pad pitch is very small. As an example, the pad pitch in some products can go below 75 μm , which is the limit of the actual probe technology today.

In order to obtain good signal integrity during RF probing, a configuration of GND-Signal-GND for RF signals is required. A key challenge for KGD processing of RF devices is to ensure that the GND-Signal-GND configuration is designed into the die to maintain the RF path at controlled impedance, given proper probe card design and RF probing techniques.

RELIABILITY SCREENING AT THE WAFER OR DIE LEVEL

Voltage and temperature over time are the known stresses for accelerating silicon latent defects to failure. These are more readily applied at package level than at wafer or die level. Applying these stresses prior to packaging the die is a key challenge for KGD.

Development of a cost-effective full-wafer contact technology with the process capability required for manufacturing is a key challenge for the industry. Contact process capability is a function of not only the contactor technology performance but also the burn-in stress requirements for a given product.

STATISTICAL PROCESSING OF TEST DATA

Techniques using statistical data analysis to identify subtle and latent defects are gaining favor in the industry, especially for device types with low shipping volumes, part number profusion and short product lifetimes that make burn-in an untenable option and for products where intrinsic process variation makes separating good die from defective die impossible using traditional on-tester limits. The advantages of reliability screening at a test insertion instead of burn-in are savings in time, fixtures, equipment, and handling. The KGD implications are that screens can be performed at the wafer level with standard probes and testers, so every device can be considered fully conditioned in compliance with data sheet specifications and shipped quality and reliability targets for that process regardless of the final package in which the device is to be shipped. Using off-tester statistical methods the test measurements (for example, Idd, Vddmin, Fmax) of each die are recorded instead of being binned. These measurements can be recorded for different test conditions, pre and post stress testing, and at different temperatures. Pass or fail criteria are determined based on statistical analysis of the measurements recorded using off-tester post processing algorithms. Outliers to the statistical distribution are graded based on their statistical likelihood of being system failures or early life failing devices, and the inkless wafer maps are modified accordingly. The challenge for testing using statistical methods is to meet an acceptable trade-off between the potential failing population and the intrinsic yield loss.

SUBSEQUENT PROCESSING AFFECTS THE QUALITY OF THE DIE

The processing that occurs during assembly can damage some technologies. Wafer thinning is one example: when DRAM wafers are thinned, a shift in the refresh characteristics has been observed. A die from a wafer that was fully tested at wafer level may fail the exact same test after being thinned and assembled into a SiP or MCP. The thermal processing steps that are present in the assembly process can also lead to a change in the refresh characteristics of individual bits. This phenomenon, known as variable retention time (VRT), is impossible to screen prior to the assembly process.

A key challenge is to re-establish the quality levels achieved by the die supplier. This can be accomplished through additional post assembly testing, invoking redundant elements in the individual failing die within the multi chip package, or using components that are specifically designed for multi chip applications.

LOGIC

The focus of this section is the testing of CMOS digital logic portions of highly complex digital logic devices such as microprocessors and more generally the testing of logic cores that could stand-alone or be integrated into more complex devices. Of primary concern will be the trends of key logic test attributes assuming the continuation of fundamental roadmap trends. The “high volume microprocessor” and “Consumer SoC” devices are chosen as the primary reference because the most trend data is available for them. Specific test requirements for embedded memory (such as cache), I/O, mixed-signal, or RF are addressed in their respective sections and must also be comprehended when considering complex logic devices that contain these technologies.

HIGH VOLUME MICROPROCESSOR TRENDS DRIVERS

The trends in the first part of the tables are extracted from other parts of the ITRS, and are reproduced here to form the foundation of the key assumptions used to forecast future logic testing requirements. The first two line items in [Table TST6](#) show trends of functions per chip (number of transistors) and chip size at production. Chip size is held relatively constant aside from incremental yearly reductions within a process generation. The next line item reflects a trend toward multiple core designs to address, in part, what has become the primary microprocessor scaling constraint - power consumption. The ITRS currently assumes a doubling of cores with each process generation. The last two line items in this part of the logic tables are the nominal device Vdd range and off-chip data rate trends.

SYSTEM TRENDS DRIVERS

System trends drivers are very important to consider when projecting future test requirements. For example, one of the most critical system constraints is power consumption. The proliferation of mobile applications, lagging battery technology improvements, system power dissipation issues, and increased energy costs are all contributing to a practical cap on device power consumption. The era of device power increasing unconstrained with increasing performance is over. This does not necessarily mean that performance will be similarly capped, but this is one of the main challenges to be overcome if Moore’s Law is to continue. Innovations in transistors, process technology, design architecture, and system technologies could all have a major impact.

One system technology innovation that could impact test would be the integration of voltage regulation on-chip/package. Increasing chip power and increasing number of cores make this ever more likely for at least two reasons. The first reason is that eventually the package limits power consumption by constraining the number of power/ground pins and the maximum current per pin. These constraints can be greatly eased with on-chip regulation since you can then deliver power to the chip at a significantly higher voltage. The second reason is that multi-core architectures may necessitate more sophisticated independent power delivery to each core in order to optimize fully the power consumption. Eventually it is likely that this will need to be done on-chip. Overall, this trend would simplify the problem of delivering power to devices under test, but it also could create many new test issues since precise voltage control and current measurement have always been an important aspect of testing high power devices.

Another important system trend is the continuous increase of chip-to-chip data bandwidth required over time. This translates into increasing chip I/O data rates and or an increasing numbers of I/O pins. In order to reliably achieve speeds much greater than one giga-transfers per second (GT/s), it is necessary to incorporate high-speed serial signal techniques such as differential signaling or embedding the clock together with the data. For example, this is already occurring with PCI Express and Serial ATA interfaces and will proliferate to all I/O ports (e.g., front side bus) for microprocessors over the ITRS timeframe. Refer to the High Speed Input/Output Interface section for more detailed discussion on the test requirements and challenges for testing these interfaces.

DFT TRENDS DRIVERS

In order for test cost not to increase proportionally with chip scale trends, continuing improvements to DFT coverage and effectiveness will be crucial. The general trend toward multiple reusable cores offers many exciting DFT possibilities. Could the cores be tested in parallel? Could the cores test each other? Could one or more cores be redundant so that “bad” cores could be disabled by test or even in end-use? Also, could there be opportunity for general purpose test cores—sort of an on-chip ATE? It is very difficult to determine exactly how this will evolve and how this will impact manufacturing test requirements. However, there are some clear trends:

- Structural, self test and test data compression techniques will continue and will be important for containing test data volume increases as well as constraining the device I/O interface during test.
- DFT will continue to be essential for localizing failures to accelerate yield learning.

- DFT is required to minimize the complexity of testing embedded technologies such as memories and I/O. For example, memory BIST engines are routinely integrated into the device to alleviate the need for external algorithmic pattern generator capability and, similarly, I/O DFT features are increasingly being employed to alleviate the need for sophisticated I/O testing capabilities in high volume manufacturing test.
- DFT will increasingly be required to ensure device deterministic behavior or accommodate device non-deterministic behavior. Power management features, I/O communication protocols, and device self repair mechanisms are a few examples of desirable non-deterministic behaviors.

Table TST6 Logic Test Requirements

HIGH VOLUME MICROPROCESSOR TEST REQUIREMENTS

The second parts of the Logic Test Requirements tables enumerate the test requirements over the ITRS horizon. I/O data rates are bounded on the low end by the need to provide slow speed access for structural or DFT-based testing and on the high end by the native speed of the interface. There will be considerable overlap of I/O types, including on the same device, through at least the near term horizon. Power consumption has been effectively capped at a maximum of 300 W for client microprocessor applications and a maximum of 400 W for server microprocessor applications through the end of the roadmap. Similarly, equipment vector memory requirements are expected to grow modestly over time.

The Giga-Pin-Vectors (GPV) shown in the tables are based on the number of logic gates assuming Stuck-At-Faults (SAF) only. Table TST7 gives multiplier ranges that can be used if detection of other fault types is to be included in test.

IMPORTANT AREAS OF CONCERN

1. Power Consumption and Thermal Management—While it is assumed that device power consumption will be significantly constrained through the roadmap horizon, new innovations in power delivery and thermal control at test may be required to stay ahead of system design optimization.
2. High Speed I/O Interface—Test equipment solutions are needed for device characterization and DFT / test equipment solutions that scale economically are needed for high volume manufacturing test.
3. Multi-core Trends—Create both test challenges and opportunities. Multi-core designs may exacerbate the previous two areas of concern and add test complexity, but also may offer exciting DFT opportunities such as redundancy that could significantly aid test.

Table TST7 Vector Multipliers

Fault Type	Vector Multiplier	
	Min	Max
BF (Bridging Fault)	1.3	1.3
TF (Transition Fault)	3	5
SD (Small Delay)	2	40

HIGH SPEED INPUT/OUTPUT INTERFACE

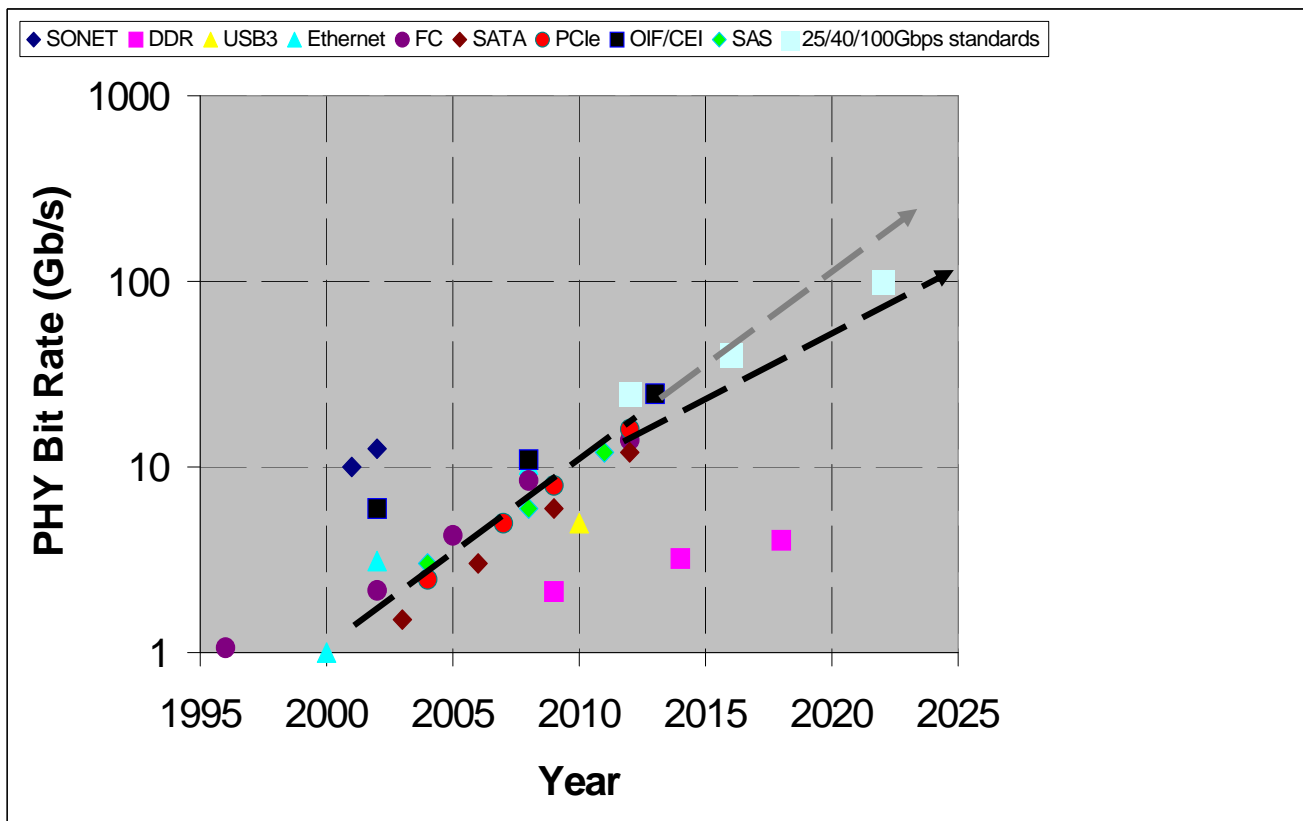
High frequency I/O technology continues to show significant growth in speed and port count in computing and networking applications, beyond its true serial communication origin. It is now a very typical IP block found in micro processors, peripheral interface chips, and inside peripheral devices. Gigahertz interface standards such as PCIe, hyper transport, DDR, FB DIMM, USB, Infiniband, SATA, SAS, Fiber channel, Gigabit Ethernet, XAUI, SONET, OTU, and OIF have been gaining in application popularity. Because of the diversity of applications, it is difficult to outline the future trends in a simple manner. Devices pushing the technology envelope represent a very small percentage of the IC industry. There are very high volume devices that do NOT necessarily push the speed limits, but they are driven by very a different cost structure than the high-end low-volume devices and therefore a very different tolerance level on test cost. There is another category of devices that require large port counts in a single device, which deviates from the normal “serial” concept of this technology, and demanding different test considerations. Because of the above noted reasons, the technology trends are illustrated as a scattered plot in Figure TST9.

The booming telecomm industry of 1998~2001 drove the standard serial communication technology to 13Gbps, and towards 40Gbps using more exotic processes. Gigahertz interface speed experienced an exponential increase (as shown in Figure TST9 plotted in logarithmic scale) in computing and network applications; partially as a result of leveraging early telecomm experience. However, because of the much different cost structure, processing, packaging and test technologies,

30 Test and Test Equipment

used in high speed interfaces, the adoption was far from simple. For example, most of the Gigahertz interfaces are built in SOC type of devices with mainstream CMOS processes. The large scale integration not only presents a challenge in designing reliable high performance interface IP targeted for a very noisy SoC environment, but also on testing of such kind of high performance interfaces. The other challenge derived from high integration level is the trend for lower Gbps-per-mW, which is becoming a major competitiveness measure in this area. This power reduction requirement challenged the design margins. This impacts the test area by an intricate balance of “guarantee by design” and “production test coverage”. As an integrated IP block, Gigahertz interface testing is now tied to the whole SoC chip level testing. For most of these applications, it means high frequency/performance testing is tied together with large amount of CMOS digital pins/logics. Test cost tolerance, high frequency instrument availability, and signal integrity restriction on test hardware design are the new areas for development in the last few years.

As in any technology roadmap analysis, sustainability of the exponential growth trend depends on changes in technology requirements and foundation technology development. As forecasted in 2007, the telecomm, datacom, and storage industries reached a temporary gigahertz plateau and it will require a serious undertake in technology to go beyond the somewhat established 13Gbps range. As of 2009, shrinking process technology is starting to support growth on data rates beyond 13Gbps; such as 14Gbps FC and 16Gbps PCIe. Some next generation focused common electrical interface (CEI) of OIF are expected to reach up to 25Gbps. Unlike the long haul telecomm applications, some of the Gigahertz interface (especially the chip-to-chip interfaces) have chosen to grow in port count (like a bus) until the higher speed technology becomes more cost efficient for use (for example, CEI/OIF, and 40G/100G Ethernet (IEEE 802.3ba standard)). Some applications have opted to change the coding scheme which eases the required generational speed increase, so the expected 3 year doubling requirement is somewhat reduced (for example, PCIe, and FC). Future data rate progression will be slower than that seen in recent history but still requires doubling every few years.



pattern type of synchronization and the typical PRBS based patterns used for BIST. From the basic at-speed testing instrument stand point, it is safe to say that ATE test solutions for 6Gbps have been established. Of course, there will always be applications that push beyond the ATE speed limit (such as for high-end PHY) or port count limit (such as for network switching devices).

At the present time, high-speed production test solutions range from internal digital loopback, DFT assist digital loopback, external wired loopback, external active loopback, high end ATE pin cards, test modules, golden device on test board, to add-on external instrumentation. Each of these approaches has their own pros and cons, which each company must consider. The trade off is of course on silicon area used for DFT vs. implementation costs vs. test equipment and interface cost. The tolerance to defect rate of different products is also a major determining factor for the different test choices.

Another challenge to test instruments is the wider adoption of on-chip equalizations. In order to keep the overall system cost down, low cost PCB lamination materials such as FR4 will still be the material of choice for most of the telecomm backplane and computing applications. However, the preference for FR4 forms a bottleneck in spectral bandwidth. Several techniques are being developed to extend the transmission data-rate under this constraint, such as transmit-side equalization (single/multi-tap pre-emphasis/de-emphasis) and receiver-side adaptive equalization (analog equalization and digital feedback equalization). In order to test these types of features, test instruments need to go beyond generating and capturing binary data. In relatively lower speed applications, these types of equalization functions were handled with analog instruments such as AWG and sampling oscilloscope. However for speeds beyond 3~6Gbps, most of these equalization features are covered in overall close loop BER test, with bandwidth limiting media mimicking the final applications. This limits the ability to isolate and spec individual components of the on-chip equalization circuit. In practice, it is also challenging to make the production test PCB match the end application. On-chip data eye analysis in the receiver is becoming more standard in order to facilitate this type of testing. More research is needed to properly specify these features, as well as on test, DFT and BIST.

Test equipment manufacturers are refining the instruments for GHz testing. It is becoming a norm to expect at-speed instruments with flexible clocking and device synchronization. Jitter measurement and tolerance testing capability is also becoming more readily available. On the other hand, because of the heavily used equalization schemes being introduced on the GHz interface I/Os, it is more and more difficult to isolate the stand alone performance of the transmitter and receiver. With strong Tx equalization and Rx adaptive equalization, the transmitter eye quality alone is hard to quantify without knowing the transmission media model and the receiver equalization designs. These pose challenges to the test equipment manufacturers on how to handle the increasingly blurred Tx and Rx performance specification boundary with only a system level BER focused metric. Of course, viable solutions need to be cost effective. Currently, solutions to this problem range from early stage mid volume thorough testing for prove in; guarantee by design and thorough characterization; to proprietary outlier detection techniques.

In the long run, the existing DFT features need to be extended beyond the current functional PRBS BIST approach to provide more performance related parametric coverage. It is envisioned that on-chip instruments and built-in design verification techniques will evolve and coexist with off-chip test equipment. An economically ideal distribution of on-chip and off-chip test coverage has yet to be determined, and can vary from application to application. The overall goal is to minimize manufacturing test cost and efficiently test high port count devices.

IMPORTANT AREAS OF CONCERN

1. **Data Rate Increase**—In the high volume computing and networking applications, the Gigahertz interface is still keeping its exponential pace but with a reduced slope in logarithmic scale, as shown in Figure TST9.
2. **Port Count Increase**—Low voltage CMOS technologies and low output voltage swings enable massive integration into large ASIC and SoC devices. In 2009, up to 250 pairs of 7Gbps backplane style SerDes are found in some applications. However, for a large percentage of applications the port count will be limited to 32 or less. The port count is still increasing and will be limited by chip level power consumption and the actual increasing single lane Gbps data rate.
3. **Cost Factor**—Traditionally, most multi-gigabit transceivers were designed as high-performance, high-priced, and high-margin devices with a low level of integration and relatively low production volume. With the introduction of low cost, low power CMOS macro cells, Gigahertz interfaces have become valued additions to many high-volume and low priced commodity devices. In addition to high port count, a cost efficient ATE solution that can concurrently test all serial ports is essential for production. The constant trade-off between performance and integration level results in the separation of SerDes devices into two categories: High-performance-level serial transceivers and high-integration-level gigahertz link macro-cells. The test methods for

32 Test and Test Equipment

each type are selected with cost in mind. The economics of high-performance, long haul, communication related products typically allow a more traditional, instrument based test approach or use of a hybrid tester as discussed earlier. Although reliable DFT features or other low cost test techniques are the ultimate solution for large port count SerDes, there is still a strong desire that the tester can provide at-speed characterization or early production. With accelerating technology improvements, the life cycles for most products are becoming shorter, therefore it may become increasingly difficult to verify and optimize DFT circuitry and yield learning.

4. **Jitter Decomposition and Measurement**—The jitter generated by the transmitter is the key parameter to guarantee transmitter quality. Many serial link standards adopt the concept of separating jitter into deterministic jitter (DJ) and random jitter (RJ). The traditional concept of histogram based peak-to-peak jitter has been replaced by the concept of total jitter (TJ), which is associated with a certain BER for the serial link (typically 10^{-12} or smaller). As the data rate keeps increasing to ~ 10 Gbps and higher, jitter amplification, (a new phenomenon during which high-frequency jitter caused by pulse width shrinkage or duty-cycle distortion (DCD) gets amplified by the lossy channel (e.g., FR4 based)), becomes severe and needs to be bounded. In addition to the conventional DJ, RJ, and TJ components, Pulse width jitter (PWJ) and DCD have become the new jitter components to test at 10 Gbps and higher. Furthermore, in some high-volume and high performance I/Os such as PCIe 3.0, a new criterion of whether the jitter is correlated or uncorrelated (to the data pattern) is established to account for the compensation effects of equalization to the correlated jitter. As such DJ and RJ are further separated to correlated DJ (including data-dependent jitter (DDJ)) and uncorrelated DJ and RJ. Correlated DJ (e.g., DDJ), uncorrelated DJ and RJ, along with PWJ and DCD need to be tested to ensure the interoperability and utilization of the jitter margin provided by equalization circuits. There are also other trends to measure jitter in terms of cycle-to-cycle jitter, peak-to-peak jitter or RMS jitter within a certain numbers of cycles, which could be more meaningful for clocking schemes only allowing very short-term jitter accumulation. Jitter measurement also imposed a very stringent signal integrity requirement. On the other hand, with the source synchronous bus entering the Gigahertz range, it introduces another challenge in jitter measurement – the uncorrelated jitter between the clock path and multiple data paths. This is certainly beyond the traditional serial PHY jitter definition.

To accurately measure the DJ and RJ, the jitter floor instrument or tester needs to be significantly below the DJ and RJ of the DUT to avoid the errors in the measured DJ and RJ. The instrument or tester for jitter measurement had also adopted the DJ and RJ merits for its performance. Most of the high-speed standards specify a ~ 0.3 UI TJ at BER = $1e-12$, with 0.15 UI from DJ and 0.15 UI from RJ. Following this jitter budget allocation distribution, and assuming that a 10% (a relatively loose accuracy target) TJ accuracy for an instrument or tester, then its DJ and RJ accuracy limits can be estimated via dual Dirac model (i.e., DJ distribution is a dual Dirac, and RJ distribution is a Gaussian), and the TJ, DJ, and RJ accuracy limits as a function of data rate are shown in Figure TST10. At 10 Gbps, the TJ, DJ, and RJ accuracy requirements are 3ps, 1.5ps, and 0.11ps respectively according to Figure TST10. These jitter accuracy targets can be met by most leading edge laboratory instruments. If large TJ error is allowed, then DJ and RJ accuracy precision will be relaxed proportionally. It is worth mentioning that this figure intends to give a math and physics based TJ, DJ, and RJ accuracy guideline for a jitter measurement instrument or tester, rather than a specification or hard requirement.

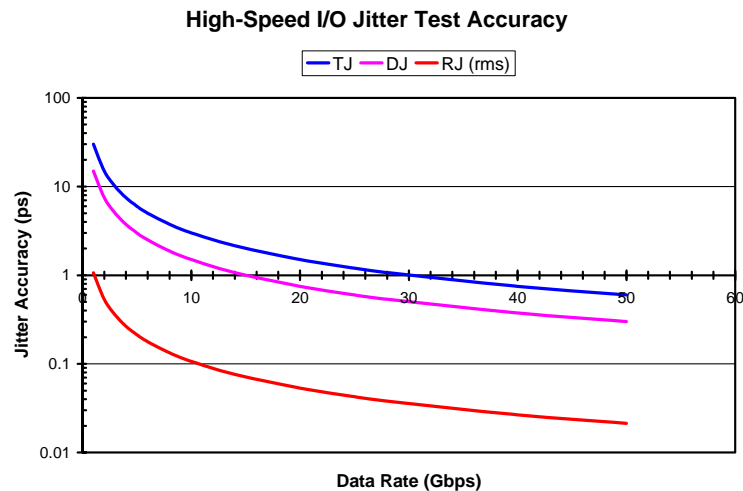


Figure TST10 High Speed I/O Jitter Test Accuracy Requirements Scaling with Frequency

5. **Jitter Tolerance Test**—Jitter Tolerance measures the level of jitter on the input signal that the receiver can tolerate before communication quality, in terms of BER, is degraded. This is a key specification for receiver (Rx) jitter and noise immunity. To conduct a jitter tolerance test, jitter must be deliberately injected into the data stream in a controlled fashion to mimic the worst case jitter and signaling conditions derived from the combinations of transmitter, channel, and reference clock in general. Depend on the architectures of the clocking schemes used; different type of jitter is needed to stress the receiver. Some applications need sinusoidal jitter (SJ) injection, some require a combination of spectrally resolved or frequency-band limited DJ, RJ, DDJ, periodic jitter (PJ) or SJ, and bounded uncorrelated jitter (BUJ) (to mimic the crosstalk induced jitter), and others demand jitter injection in terms of peak-to-peak or RMS jitter within a certain numbers of cycles. An emerging important topic is the receiver equalization test; especially feedback based adaptive equalizations such as decision-feedback equalization (DFE). The challenge comes from the fact that testing observability is intrinsically poor for Rx, with the BER as the only and common observable merit. Without the assistance of on-die or on-chip instrument such as on-die scope, it is very difficult if not impossible to verify DFE tap adaptations to the targeting values. To perform diagnostic or debug test for DFE, on-die instrumentation appears to be a viable solution. Integrated instruments that can inject these kinds of jitter do not exist on previous ATE world, but start to become more available in the lab equipment world, as we have predicted 2 years ago. As of today, some ATE solutions do provide jitter injection capability to stress the receiver and/or clock and data recovery for manufacturing test, which is a big improvement from 2 years ago.
6. **Test Fixture Bandwidth**—The test fixture used to interface the device to the instruments/ATE include a printed circuit board, cable, connector/pogo pin and etc... With frequencies growing and port counts increasing, the ability for delivering high frequency signals to instruments without much loss and distortion becomes a monumental task in test engineering. Once the signal goes beyond 10Gbps, the fixture bandwidth requirement reaches 20GHz or higher to have a minimum ISI jitter from the test fixture. In most cases, the bandwidth is NOT the only problem. The phase response has an even bigger impact on jitter measurements. Although there have been significant improvements in socket bandwidth recently, the new high bandwidth socket solutions are challenged for their limited mechanical specifications. Non-reliable insertions and inconsistent contact are still the most common problems causing test result variation. Because of these test fixture limitations, complicated de-embedding effort became a norm for GHz interfaces in performance tests. This therefore inspired another area of active development – tools to simulate and de-embed the signal path including socket, PCB, cable and connector. However, unlike the narrow band RF de-embedding, the wide band de-embedding challenge requires not only insertion loss but also return loss across a wide frequency spectrum. Some SerDes specs also try to alleviate the de-embedding effort with standardized test fixtures.
7. **Parametric DFT versus logic DFT**—The basic Pseudo-Random-Bit-Stream (PRBS) generator and a BER checker functions have been a norm for years now. In order to cope with adaptive equalizations, more GHz designs now include internal data eye analysis capability. As more analog content is introduced into GHz I/O design, more research and innovation is needed in analog DFT.

34 Test and Test Equipment

8. **Calibration firmware**—Because of the power and area requirements for GHz I/O design, the analog portion of the circuits are under constant pressure to be scaled down. That leads to more digitally assisted analog design implementations, which means more on-chip self calibrations are required. Some adaptive equalization schemes use firmware. This has created a new test requirement – testing the robustness of the calibration firmware.

MEMORY

Memory density will continue to grow over the roadmap period according to Moore's Law. The ITRS ORTC tables show industry agreement on a three-year technology pacing for DRAM. As such, the density trend of all memory devices has slowed compared to the 2005 roadmap. Continued scaling of memory will require further developments in high and low K dielectrics and other materials. Memory I/O data rates will increase over the course of the roadmap with DRAM on the leading edge as compared to other memory types. DRAM will continue to be the I/O performance leader throughout the roadmap while NAND will dominate the density.

Memory test solutions have recently targeted by device (DRAM, NAND, NOR) type, but the recent requirement to easily flex capacity between the various memory types is driving more common solutions, especially in the wafer test functional test environment. The great disparity in I/O data rate across the various memory types will make development of a single low cost solution difficult for performance test.

Table TST8 Memory Test Requirements

DRAM

DRAM bit density is projected to double every three years and will continue to increase in I/O performance. DRAM will reach a data rate of 8.4 Gb/s per I/O in 2022 based upon the DRAM historical model that forecasts data rates thru DDR6. Increasing memory size will drive higher device test time and decreasing manufacturing cell throughput unless offset by DFT or multi-insertion testing is to minimize cost. Failure detection, analysis, and repair are necessary for commodity DRAM. To enhance test productivity, new test-oriented architectures will be required. Multi-bit testing, BIST, and built-in self repair will be essential to maintain the production throughput and yield.

Considerable test parallelism increase from ATE will be required over the roadmap in order to manage test cost. Commodity DRAM will lag the leading edge specialty DRAM in I/O bit rate. In the realm above 2 Gb/s DUT interface signal integrity issues exist with the test socket and probe card. Parallel test exceeding 128 devices per test head will be a challenge due to the interface routing complexity required. Performance DRAM I/O bit rate will be at least two times greater than the HVM production bit rate. High I/O bit rates create a challenge on how to define tester accuracy in the future properly. The traditional I/O bus width of 4, 8, and 16 bits has been supplemented by bus widths of 32bits that are being driven by mobile devices. The bus on high volume commodity DRAM may transition to 32 I/O near the end of the roadmap driven by potential differential bus requirements.

FLASH

NAND will double in density ever year in the short term and slow to a doubling every 1.5 years. The doubling every 1.5 years will be faster than the projected lithography node migration due to the further increase in the number of bits stored in a single memory cell from one and two to four on some cell types. NAND density has generally been 4× the NOR density at any given technology generation and the forecast shows that trend continuing. Use of error correction allows greater uncorrected error rates and enables increased bits per memory cell. NAND bus width has continued to be primarily 8 bit with some products at 16 bit.

NOR memory density is expected to double every three years over the roadmap period. The 2005 and previous roadmaps have consistently shown a 32 I/O NOR product, but 32 I/O for NOR has been removed from the 2007 roadmap. NOR products will be organized as either 8 or 16 bit, but the number of I/Os will vary due to implementations of serial and multiple busses. I/O data rates are forecasted to increase to 533 Mb/s over the roadmap period to keep pace with faster ASIC devices used in embedded applications.

NAND and NOR generally have not had the same test solution due to architectural, usage, and bad block handling differences. However, bus differences between NAND and NOR have blurred over time and the Serial Peripheral Interface (SPI) bus and other serial bus definitions are targeted to support embedded applications. Further proliferation of bus types is expected due to the customization of flash for applications. Bus width is predominantly 8-bit and 16-bit on densities > 4 Mb. The 32 bit bus width NAND has been dropped from the roadmap starting in 2007.

The need for internal voltages that are 3–8 times the external supply requirements is expected to continue, driven by the hot electron and Fowler-Nordheim charge transport mechanisms. Increased absolute accuracy of supply voltages will be required in the future due to the trend toward lower voltages, but percentage accuracy relative to the supply voltage will

remain a constant. I/O voltage decreases are pushing the operation limits of standard tester load circuits; new methods will be required in the future.

Wafer test generally does not require the performance of package test, but error detection, error analysis, and redundancy processing is required. Stacking of various types of Flash and other memory and/or logic components in a single package has become standard and is expected to continue. Multiple die within a package has complicated the package test requirements and has increased pin count and number of DUT power supplies required. Data and clock rates for flash will increase, but there is expected to be a wide variability in the requirements based upon the end application.

EMBEDDED MEMORY

Embedded DRAM bits will not match the density growth rate of commodity DRAM and NAND. The major concern for a merged logic-DRAM design in a dual-gate process will be array noise and sense-amp imbalance. For the 100 nm DRAM half pitch and below, DFT for inline defect detection will be necessary for product development.

Embedded Flash memory bits will grow exponentially in the near term and then double every two years in the later years of the roadmap. More devices will include both DRAM and Flash memory. Oxide reliability, sense-amp imbalance, and oxide-nitride-oxide (ONO) scaling will be the major concerns in flash memories in the future.

To enhance test productivity, new test-oriented architectures will be required. Built-in self-test and built-in self-repair will be essential to test embedded DRAM and embedded Flash memories and to maintain production throughput and yield. The primary test algorithms for Flash memories will continue to be Read-disturb, Program-disturb, and Erase-disturb while March tests with all data backgrounds will be essential for embedded DRAM.

Considerable parallelism in test will be required to maintain test throughput in the face of rising memory densities. In some cases, test is made cost-effective by double insertion of devices rather than testing both logic and embedded memories on the same tester. In double insertion, embedded Flash and DRAM could be tested and repaired on the memory tester, while the logic blocks are tested on the logic tester. Embedded SRAM test requirements are captured in the High Performance Microprocessor section of the chapter.

ANALOG AND MIXED-SIGNAL

The economic benefit of monolithic integration (SoC) and single package integration (SiP) is well established. This integration has combined digital, analog, power management, mixed signal, and RF/microwave circuitry routinely in a single package and often on the same die. This trend has increased the breadth of interface types on a single part, and given rise to test equipment that mirrors this range with a corresponding breadth of instruments.

Another important trend impacting mixed signal and analog testing is the compelling economics of multi-site testing, also called parallel test. To support parallel test, many more instrument channels of each interface type are required to keep test cell throughput high.

The increasing number of interfaces per device and the increasing number of devices tested simultaneously raise the need to process an increasing amount of data in real time. The data from the mixed signal and analog circuitry is non-deterministic and must be processed to determine device quality. This processing must be done in real time or done in parallel with other testing operations to keep test cell throughput high.

Looking forward, the breadth, performance, density, and data processing capability of ATE instrumentation will need to improve significantly to provide the needed economics. The area undergoing the most change is RF/microwave and so it is covered in its own separate section. The digital and high speed serial requirements for mixed signal devices are equivalent to logic and are covered in that section.

This section focuses on analog/mixed-signal test requirements. The Mixed-Signal Test Requirements table focuses on test instruments rather than specific chip applications. The test instrumentation must often cover more than one device market segment to provide sufficient utilization in a single tester configuration, so the requirements for multiple segments are aggregated into a few instrument categories. The analog waveform generation and capture requirements are set in two classes: low frequency—basic/minimum requirements for a mixed-signal ATE and very high frequency high-end requirements. Where appropriate, the mixed-signal instrument requirements are linked to other sections and tables in the roadmap.

There are two important trends. The first is to deliver adequate quality of test. Most analog/mixed-signal testing is done functionally. This requires instrumentation capable of accurately generating and analyzing signals in the bandwidths and resolutions of the device's end market application. Both of these parameters are trending upwards as more information is communicated between devices and/or devices and the physical environment.

36 Test and Test Equipment

The second key trend is to enable the economics of parallel test through instrumentation density and parallel test efficiency, a measure of the overhead in testing multiple parts. The level of parallelism shown in [Table TST2](#) “Multi-site Test for Product Segments” indicates an increase in instrumentation density.

These trends of increasing ATE instrument channel count, complexity, and performance are expected to continue, but at the same time the cost of test must be driven lower (see the areas of concern listed below).

Analog/mixed-signal DFT and BIST techniques are lagging. No proven alternative to performance-based analog testing exists and more research in this area is needed. Analog BIST has been suggested as a possible solution and an area for more research. Fundamental research is needed to identify techniques that enable reduction of test instrument complexity or elimination of the need for external instrumentation.

Table TST9 Mixed-signal Test Requirements

IMPORTANT AREAS OF CONCERN

1. Time-to-market and time-to-revenue issues are driving test to be fully ready at first silicon. The analog/mixed-signal test environment seriously complicates the test fixtures and test methodologies. Noise, crosstalk on signal traces, added circuitry, load board design complexity, and ATE hardware/software issues currently dominate the test development process and schedule. The test development process must become shorter and more automated to keep up with design. In addition, the ability to re-use analog/mixed-signal test IP is needed.
2. Increased use of multi-site parallel and concurrent test of all analog/mixed-signal chips is needed to reduce test time, in order to increase manufacturing cell throughput, and to reduce test cost. All ATE instrument types, including DC, will need multiple channels capable of concurrent/parallel operation and, where appropriate, fast parallel execution of DSP algorithms (FFTs, etc) to process results. In addition, the cost per channel must continue to drop on these instruments.
3. Improvements in analog/mixed-signal DFT and BIST are needed to support item 1 and 2 above.

RADIO FREQUENCY

Four main RF frequency areas are distinguished. The low frequency range up to 3 GHz is dominated today by long distance cell phone communications technology. With wireless client communications protocol (WiMax) this will transfer over time to the higher frequency bands. The 3-6 GHz range is used by satellite TV and 802.11a networking. The 6–45 GHz band is focused on medium distance communications (Bluetooth moving to ultra wide band (UWB)). The high section, 45-94 GHz, is used primarily for short distance radar applications, particularly automotive, but 60GHz will also be used for congested area (short range) WLAN and satellite to satellite communication. CMOS devices are expected to be able to support 60GHz.

The most important movement is the increase in frequency: using higher frequency bands in line with the 802.11 and 802.16 communication standards. For the high frequency ranges (> 12 GHz) classical test techniques (non-modulated) are expected to fulfill the requirements.

An important requirement for test is full synchronization between the power/digital/AC baseband part of the tester and the RF instruments. Error vector magnitude measurements are a prerequisite. An important trend is the move of RF into SoC and SiP solutions. This requires not only test solutions for the RF parameters, but also one in combination with high-end digital and mixed signal requirements. For SiP, the wafer probe capability becomes important. To cope with the economics of RF becoming a real commodity, multi-site will also be mandatory and will increase in this application area. The tooling (load boards, sockets, probe cards) is also critical to ensure signal integrity to and from the DUT. Looking at these challenges, the need for specific design-for-test for RF and finding lower cost alternatives to functionally testing RF devices is expected to increase heavily in the near term.

Table TST10 RF Test Requirements

IMPORTANT AREAS OF CONCERN

1. The increase in performance and frequency in combination with the economics of test will put a strong focus on novel design-for-test and alternative test techniques to be developed in the coming years.
2. RF will much more frequently be embedded into products via SoC or SiP techniques. Combination of RF tests with (high-end) digital and mixed signal will be more common. RF test on wafer level will increase. Next to the test system, there will also be emphasis on the tooling (load boards, sockets, and probe cards) to cope with signal integrity.

3. Source and measurement accuracy for phase noise and signal detectors are adequate today but must improve in the near term. Phase noise at 100 KHz needs to improve from today's -120 dBc/Hz to at least -130 within the next year or two.
4. The visibility of trends in the high RF ranges (> 6 GHz) is evolving.
5. Probe is a challenge over 6 GHz and the timeframe for modulated waveform testing over 6 GHz is uncertain.
6. Impedance standards and calibration methods for high frequency measurement at probe need to be created.
7. The EMI environment of the test development setup may be substantially different than the environment on the production test floor creating yield and correlation issues.

RELIABILITY TECHNOLOGY REQUIREMENTS

Reliability solutions are an optimization of 1) reliability defect density (RDD), 2) learning, reliability screens, and test methods (RS&TM) applications, and 3) design for reliability (DFR). The goal of the reliability solution optimization is to provide the best value for the reliability dollar spent, where value is defined as the ratio of customer satisfaction to customer cost.

In reliability circles, customer satisfaction is measured by the field failure rate or failures in time (FITs). The cost of reliability has two components: manufacturing operations costs and yield. As such, these two components of the reliability cost equation are the primary challenges facing every reliability solution provider. In turn, manufacturing operations costs are also driven by two fundamental components—burn in duration and equipment sophistication. The industry is still searching for a means to accelerate latent defects outside of the traditional elevated voltage and temperature methods. It follows that much progress has been made in detection techniques, but acceleration remains all about applying elevated voltage and temperature.

Applying voltages and temperatures in excess of the application specs most often demand providing solutions to leakage induced power (electrical power delivery and thermal/heat dissipation). The component of reliability cost reduction associated with yield is severely biased towards elimination of “overkill”/“false rejects,” which in many ways are tied to derivatives of the power solution. However, the primary source of false rejects stems back to the stress methodology, through the modeling assumptions, and ultimately finds its root in what one believes about “escapes” from the manufacturing stress process. The argument proceeds like this: the majority of market applications are most concerned with the early life component of the failure rate. Most latent defects that “escape” acceleration will fail early in the product life. In order to guarantee a part received stimulus—and therefore did not “escape” stress—is simply to measure the outputs during stress. Defining terms, measuring outputs is called *in situ stress*, while measuring no outputs is *dynamic stress*. Obviously the “escapes” component is less for *in situ*, and hence the early life failure rate is lower. As anticipated however, this lower failure rate does not come without cost. *In situ* stress requires functionality at stress conditions, which in turn shrinks the performance distribution. Completing the original thread, measuring outputs during stress also introduces a component of yield loss. Due to process variation, some portion of the distribution does not have sufficient margin to function at stress voltages or temperatures, however these same parts operate fine at application conditions. Although these parts contain no reliability defects, *in situ* stress will fail these perfectly functional parts—hence “over-kill.” These same parts with “marginal margin” are the target of the advances in detection techniques mentioned earlier. Achieving reliability requires trade-offs. In most instances performance and yield hang in the balance.

Reliability defect density learning rate is the most cost effective means of achieving the reliability demands of the marketplace. In itself, it is the by-product of the fundamental core practice in achieving profitability in microelectronics, yield learning rate. Defect learning is addressed in the Defect Modeling and Physical Defects section—and although historical data has overwhelmingly supported the premise that the component of defects that are “reliability unique” has been small—recent advances in technology may be changing the picture. The section on defect learning will always be directly applicable to RDD learning; however the high voltages and temperatures of defect acceleration are causing us to peer over the edge of device physics and materials science. Stress conditions are no longer dictated by “technology nominal” specs but by system application conditions. Technology’s recent inability to meet marketplace performance demands at reasonable power has forced systems designers to increase system application conditions (voltage and temperature) to compensate. Shifts in array V_{min} operating range, NBTI-driven performance margin, and gate oxide integrity (time dependent dielectric breakdown (TDDB)) as a result of the application of stress conditions still remain largely unexplained. As such, they dictate compensatory actions and/or reliability failure rate modifications. Even the standard thinking of metal electro migration for C4 and BEOL wiring requires careful scrutiny when confronted with the

38 Test and Test Equipment

radical currents and powers conjured up by stress conditions. The industry's ride on the "Performance Juggernaut" isn't over quite yet.

DFR also has three key components: 1) technology design, 2) chip design (logical and physical), and 3) system design. In each of the three, the DFR work must strive for *defect tolerance*. In the case of technology design, leakage induced power mitigation maintains an edge in importance over defect tolerance. Regarding chip design and DFR, power mitigation and fault tolerance are at par in design priority. Redundant element analysis and power dissipation analysis burn considerable design engineering horsepower. At the system level, defect tolerance exists in the forms of error detection/correction and redundant elements.

In the arena of *reliability screens and test methods*, the literature is rich with techniques and methodologies with champions and supporting/compelling/biased data. Debates vary, depending upon the technology generation, chip/circuit type, design style, performance target, reliability requirements, and defect type. As long as excessive voltage and temperature retain the throne of defect acceleration, RS&TM will challenge the best and brightest minds in power delivery and thermal solutions. One must be able to accelerate defects while avoiding destroying the device—which is a change in precedence. In years past, stress conditions or actions that invoked or even hinted upon wear-out were to be avoided. The adage in the past was "one must be able to accelerate defects while avoiding the onset of wear-out." However this is becoming increasingly more difficult in the face of stretched system applications conditions; sub-10 nm oxides; NBTI; marginal margin (that is, array V_{min}); hundreds of amps and Watts, miles of copper wire, and billions of interconnects.

RS&TM are best categorized by separating them into wafer applications and package (or module) applications and then further segregation into "*detection*" and "*acceleration*" techniques. This tiered structure will help to dilute the perennial argument between test and reliability regarding whether a field return is a test escape or an early life reliability failure.

Regardless of operational process step (wafer or package), acceleration techniques invariably must deal with potent power implications simply because acceleration requires temperature and/or voltage far in excess of application conditions—and leakage varies exponentially with both. The same is not true for detection techniques. In many instances, detection techniques employ conditions that reduce leakage (that is, VLV (very low voltage) or VLT (very low temperature)), and in instances where detection requires application conditions that exacerbate leakage, those conditions typically do not approach the level of acceleration conditions.

BURN-IN REQUIREMENTS

Technical challenges for the burn-in process are driven by increasing device pin count, decreasing package pitch, increasing device functionality and operating frequencies, dramatically increasing leakage current, and eroding voltage/thermal acceleration. Several alternate techniques such as IDDQ, high voltage stress, and wafer mapping are being used to try to improve device reliability, since many reliability failure modes are proving to be resistant to burn-in.

Burn-in system technology must continue to drive down costs, in particular for high power devices. The minimum device core voltage continues to decrease. Scan requires very deep vectors for large memories, while high power requires individual device thermal and power management. The burn-in process (system/driver/burn-in board/socket) will be challenged to meet speeds of the newest technology devices without some form of internally generated clock. Devices without DFT are requiring increasing I/O. The growing need for KGD continues to drive efforts for wafer level burn-in, KGD carriers, or additional stress during probe.

Device power and signal requirements are driving burn-in boards toward higher board layer counts, smaller traces, less space for routing, more complex processes and materials, higher test costs, and board reliability issues. Tight pitch on future devices will require new cost-effective, innovative interfaces between the burn-in sockets and the burn-in boards.

Burn-in sockets are undergoing major design challenges as they must accommodate increasing contact count, decreasing pitch, higher currents, and higher frequencies. At the same time, sockets are a key component of an overall thermal solution designed to prevent high power devices from self-destructing. A major challenge for socket manufacturers is to maintain low costs and short lead times while providing the technology to meet these new demands. Horizontally actuated contact design will be displaced below 0.5 mm pitch ball grid array (BGA) by vertically actuated contacts as pin count increases and existing socket materials fall short of increased mechanical stress requirements. New designs and new materials will be required for higher current carrying capabilities. Socket design will need to accommodate looser packaging specs in areas such as warpage and package dimensions, while coping with increased package size, thinner/more fragile packages, and reduced/non-standard/mixed pitches. Contact design will need to provide greater strength without a loss of electrical/mechanical performance.

Approaches to burn-in include traditional unit level burn-in, system level burn-in, wafer level burn-in, and strip/array burn-in (Figure TST11). On high reliability applications, system level burn-in complements or replaces traditional device level burn-in. Wafer level burn-in technology continues to be developed, but has not been able yet to make significant inroads against traditional packaged level burn-in. The challenge here is to eliminate socketted burn-in and find ways to perform simultaneous multiple wafer level burn-in using scan/logic and memory BIST (MBIST). Strip/array burn-in is becoming more important as more packages are receiving massively parallel test in either strip or array format.

Table TST11 Burn-In Test Requirements

WAFER LEVEL BURN-IN

There is no standard definition of what constitutes wafer level burn-in (WLBI). Some vendors use the term “burn-in” to refer to the application of a simple DC stress that applies opposite voltage potential to the internal nodes of a DRAM. Some say that WLBI requires full wafer contact and the application of high enough temperature over enough time to activate thermal defects, while also applying voltage stress with the device operating in “normal” mode. Some vendors enable the use of WLBI for low-end micro-controllers or SoC through DFT functions such as scan or BIST.

Key challenges are to quantify how to measure the effectiveness of these options and to develop standards that define WLBI and the methods that are used to confirm the effectiveness of this wafer level treatment. The challenge for DRAM in particular, as a device well suited for WLBI, is to provide a burn-in environment for wafers that provides the same functionality, is as effective as package-level burn-in, and yet is no more costly. The concept is to leverage the time spent in burn-in by using the burn-in environment as a massively parallel testing opportunity.

The need for WLBI is increasing. The infant mortality rate is getting worse due to transistor scaling effects and new processing technology / materials for devices. Decreasing operating voltages and margins for devices are reducing the ability to use voltage acceleration / voltage stress testing to guarantee reliability. KGD is becoming a more significant need by the customers due to requirements for chip scale packaging and multi-chip modules. Decreased cycle time and the need for faster feedback of yield / defect information to the wafer fab can be assisted by moving burn-in earlier in the overall semiconductor process. Finally, detection and removal of defective devices prior to the packaging process eliminates packaging scrap costs based on intrinsic device defects.

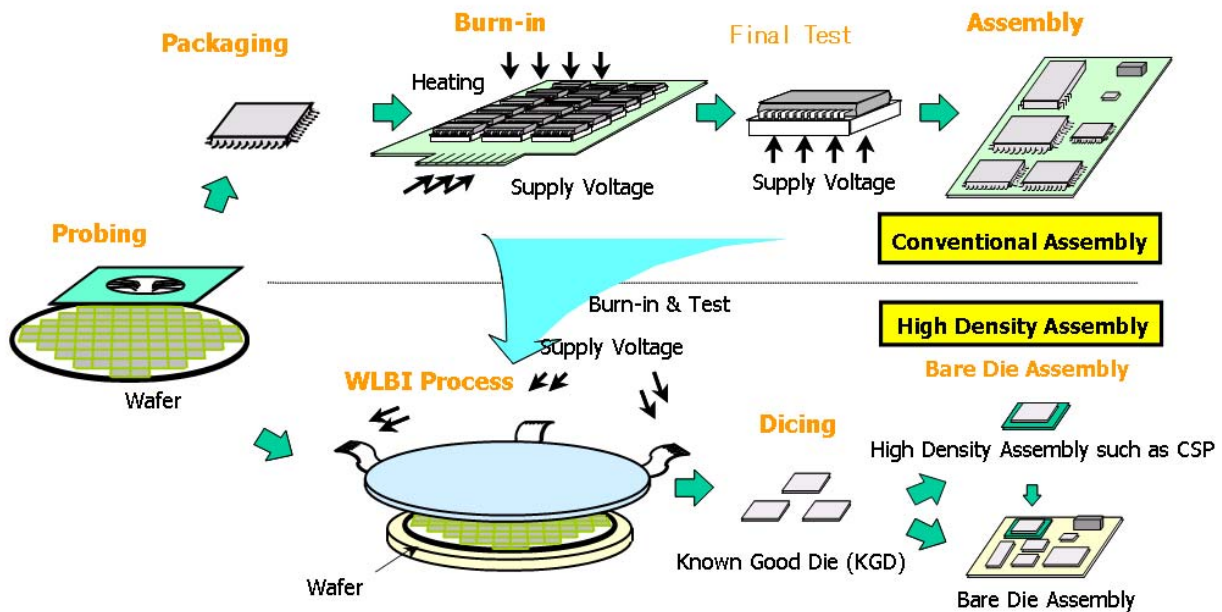


Figure TST11 The Production Process with WLBI Compared with Package Burn-in

PROBING TECHNOLOGY FOR WAFER LEVEL BURN-IN

Contactors for whole wafer contact include TPS probe and micro pogo-pin contactor. TPS probe consists of a substrate board, membrane with bumps, and PCR sheet, where the PCR sheet between two components absorbs the uneven height of bumps to achieve uniform and stable contact. Significant features of this system include the ability to concentrate pressure efficiently at each bump top and the ability to achieve over 20,000 bumps contact with Al pads by control of

40 Test and Test Equipment

bump material and surface condition. Materials with coefficient of thermal expansion (CTE) similar to Si (such as glass and ceramics) are used for the substrate board to prevent CTE mismatch.

A micro pogo-pin contactor consists of a CTE matched probe housing and pogo-pins with moving plungers at both sides. The pogo-pins stand vertically and have enough compliance and independent travel to accommodate height variations between adjacent contacts. The probe pitch is technology dependent.

Other contactors such as spring-functioned material on wafer pads as a wafer level package technology are expected for whole wafer contactor usage. For contactor roadmaps, DRAM is selected as the target application due to its large predominance in general memory burn-in. DFT is considered for system LSI.

OTHER WLBI TECHNOLOGY CONSIDERATIONS

The current consumption of a wafer increases by sub-threshold leakage from shorter transistor channel lengths and an increased number of transistors per unit area. The high temperature of burn-in also increases sub-threshold leakage. Therefore, the burn-in equipment must be capable of supplying over 1000 A of current per wafer. Also, to manage current appropriately, wafer temperature control/uniformity becomes necessary. Finally, the burn-in equipment must be able to accommodate different quality distributions across each wafer.

BIST is capable of decreasing the number of pins under test per device, but die shrinks and tighter pad pitches more than offset this advantage by increasing the total number of die and pads per wafer. The increased number of pins being tested also increases the force required to contact the wafer. In order to enable the use of WLBI through DFT functions such as scan, BIST, and JTAG³, the number of tested pins per device and total cost per device must be decreased and performance of the WLBI technology must be improved.

The probing technology for WLBI is described above. However, probing technology faces several challenges in order to meet the technology trend for the future. When the probe pitch is studied with TPS contactor technology for the devices with LOC and peripheral, a pitch around 70 μm seems feasible. For a pitch less than 70 μm , MEMS technology by use of photolithography is an option. This technology, however, does not yet have a solution for 300 mm wafers. While probing technology for tighter pitches is required, the intelligent use of DFT during pad layout may provide some relief by bypassing every other pad in order to double the probe pitch effectively, as compared to pad pitch. Application to high pin count and low force probing due to low- κ materials will also be required. This will help drive new probing technology.

TEST MECHANICAL HANDLING REQUIREMENTS

Wafer probe and component test handling equipment face significant technical challenges in each market segment. Common issues on both platforms include higher parallelism and increasing capital equipment and interface cost.

Increased parallelism at wafer probe drives a greater span of probes across the wafer surface which ultimately results in full wafer test across a 300 mm wafer. The increased probe count is driving interface complexity to route signals. Prober and probe card architecture will need to evolve to simplify the interface.

Reducing the cost of wafer-level and package-level test in the face of more challenging technology and performance requirements is a constant goal. The demand for higher throughput must be met by either increased parallelism (even with short test times), faster handler speed or process improvements such as asynchronous test or continuous lot processing.

Packages continue to shrink, substrates are getting thinner, and the package areas available for handling are getting smaller at the same time that the lead/ball/pad count is increasing. In the future, handlers will need the capability to very accurately pick and place, small, fragile parts, yet apply similar or increasing insertion force without inducing damage.

Temperature ranges are expanding to meet more stringent end use conditions and there is a need for better control of the junction temperature at the start of test. Power dissipation overall appears to be increasing, but multi-core technology is offering relief in some areas.

It is unlikely that there will be one handler that is all things to all users. Integration of all of the technology to meet thermal, throughput, placement accuracy, parallelism, and special handling needs while being cost effective in a competitive environment is a significant challenge. The 2007 roadmap defined three handler groupings based upon the power requirements of the DUT. High power DUTs consume greater than 10 watts, medium power devices are between 0.5 and 10 watts and low power device have less than 0.5 watts per DUT. For 2009, the high power handler category has been divided into 10-50W and >50W to reflect typical and extreme devices.

³ an IEEE standard 1149 boundary scan

Table TST12 Test Handler and Prober Difficult Challenges

Med-High and High Power Handler	<i>Temperature control and temperature rise control due to high power densities during test</i>
	<i>Continuous lot processing (lot cascading), auto-retest, asynchronous device socketing with low-conversion times</i>
	<i>Better ESD control as products are more sensitive to ESD and on-die protection circuitry increases cost.</i>
	<i>Lower stress socketing, low-cost change kits, higher I/O count for new package technologies</i>
	<i>Package heat lids change thermal characteristics of device and handler</i>
	<i>Multi-site handling capability for short test time devices (1–7 seconds)</i>
Medium Power Handler	<i>Support for stacked die packaging and thin die packaging</i>
	<i>Wide range tri-temperature soak requirements (-45°C to 150°C) increases system complexity</i>
	<i>Continuous lot processing (lot cascading), auto-retest, low conversion times, asynchronous operation</i>
	<i>Shielding issues associated with high frequency testing (>10 GHz)</i>
Low Power Handler	<i>A wide variety of package sizes, thicknesses, and ball pitches requires kitless handlers with thin-die handling capability</i>
	<i>Package ball-to-package edge gap decreases from 0.6 mm to 0 mm require new handling and socketing methods</i>
	<i>Handling package smaller than 1x2 mm for specialist device MEMS or photo sensor with low damage concern</i>
	<i>Parallelism at greater than x128 drives thermal control and alignment challenges</i>
Prober	<i>Consistent and low thermal resistance across chuck is required to improve temperature control of device under test</i>
	<i>Heat dissipation of >100 Watts at > 85°C is a configuration gap in the prober industry</i>
	<i>MEMS wafer stick on prober chuck without damage for chip probing (CP) test requirement</i>
	<i>Advances in probe card technology require a new optical alignment methodology</i>

Table TST13 Prober Requirements

Table TST14 Handler Requirements

DEVICE INTERFACE TECHNOLOGY REQUIREMENTS

As device I/O bandwidth and power demands increase there is a corresponding requirement for high performance power and signal delivery during electrical test. These requirements drive challenges for the assemblies used to interface the test equipment to the device under test. The highest performance interfaces require complete power and signal path modeling from the source instrument to the die, requiring accurate simulation models of the test instrument, path, probe or socket, and die. Shrinking die and package geometries as well as manufacturing productivity further complicate these interfaces with decreasing pitch, increasing pin count and multi-DUT requirements.

‘More than Moore’ improvements are being tackled by innovative heterogeneous architectures such as SiP and Die Stacking. Innovations such as TSV and Proximity communications (non-contact signaling) are being explored to address the density issue for SiP and stacked dies. SiP and staked die (KGD) present major challenges from a yield and test cost point of view. Three dimensional interconnects, whether wired or wireless, create alignment and physical challenges for testing and probing.

Test Insertion into packaging lines is a potential solution for ‘More than Moore’ problems. Traditional Semiconductor fabs have multiple off- line and in-line in process testing. Packaging has traditionally been carried out with a small number of active components. Stacked die and SiP packaging combine multiple active dies and passive components. Testing for packaged parts has traditionally been carried out using post assembly testing.

To achieve ‘More than Moore’ improvements packaging will likely need to adopt semiconductor like practices with respect to test; including in-line test for yield learning and production.

For ‘More than Moore’ the need for higher density of interconnects and KGD both present a challenge to probing. Migration of tester resources onto probe cards or DUTs is continuing thru BIST or enhanced probe cards. A potential solution to parallelism and testing is to use non-contact interconnects as a method of high density wireless probing.

42 Test and Test Equipment

Coordination or test coverage is an ongoing challenge with heterogeneous designs. Solutions may involve the addition of test structures to assemblies or instantiated dies.

PROBE CARDS

Wafer probe technologies face complex electrical and mechanical challenges driven by product specifications, test implementation requirements, test productivity goals, and reduced test cost demands. Across the device spectrum, these challenges include: higher average power demands, higher frequency response (bandwidth), rising pin counts across tighter pitches and smaller pads/bumps, increasing switching currents (di/dt), alternative pad/bump metallurgies and increasing test parallelism. Research and development of new or improved probe technologies is required to meet these challenges to ensure that the basic probing requirement of ensuring reliable, sound, and cost-effective electrical contact to the device(s) under test is achieved. Recent developments in contactless probing technology may help address future density, speed and 3D requirements for probe.

Improvements in passive probe cards are at appear to be approaching mechanical and electrical limits for increased functionality. Intelligent Probe Cards are potentially capable of solving problems of both parallelism and speed. Advances in materials MEMS and heterogeneous integration suggest that intelligent probe cards can be made economically with parallelism and performance to match DUT technologies. There are some extant commercial offerings of intelligent probe cards for use in 3D packaging test.

The tables contained in this section derive trends based on product families similar to the layout of the Test Technology Requirements section above.

TRENDS AFFECTING PROBE CARD TECHNOLOGIES

Along with addressing the key challenges listed below, research and development is urgently required to bring to the market cost-effective probe technologies directed at trends in product offerings and the testing environment.

The continuing volume growth of bumped devices, often with I/O in area arrays, points to the escalating demand for “vertical” style probe card technologies, with a rising need in multi-DUT configurations as well. Multi-row wirebond also supports this vertical style need and is particularly challenging due to tighter pitches.

Some microprocessor products and high end ASIC devices are driving power levels to 500 Watts and 1000 Watts with associated current/probe and thermal issues. Current/needle is also an issue for cantilever and MEMS technology as wire bond devices move into higher technology silicon.

Manufacturing test of devices has moved to parallel test. For some product groups (e.g., memory), wafer probe technologies are available that handle parallel testing of 512 and more devices. Probe technologies capable of full wafer contacting are in use already for 200 mm and 300 mm wafers. Increasing the contacts/DUT for these massively parallel probes is the next challenge.

Innovation in test is required for effective use of new interconnect technologies such as TSV or proximity communications.

Table TST15 Probing Difficult Challenges

Geometry	<ul style="list-style-type: none"> Probe technologies to support peripheral fine pitch probe of 23 μm peripheral staggered pad probes at effective pitches of 20/40, and fine pitch (45 μm) for dual row, non-staggered probing on all four die sides. Fine pitch vertical probe technologies to support 130 μm pitch area array solder bump and 50 μm pitch staggered pad devices. Multi-site pad probing technologies with corner pitch capability below 125 μm. Reduction of pad damage at probe commensurate with pad size reductions (or better). Alternative probe technology for 75 μm on 150 μm pitch dense array (vertical probe; bumped device). Increasing probe array planarity requirements in combination with increasing array size.
Parallel test	<ul style="list-style-type: none"> Need a probe technology to handle the complexity of SoC devices while probing more than one device. Current probe technologies have I/O limitations for bumped device probes.
Probing at temperature	<ul style="list-style-type: none"> Reduce effects on probes for non-ambient testing -50°C to 150°C; especially for fine-pitch devices. For effects on Handlers and Probers, see that section.
Product	<ul style="list-style-type: none"> Probe technologies to direct probe on copper bond pads including various oxidation considerations. Probe technologies for probing over active circuitry (including flip-chip).
Probe force	<ul style="list-style-type: none"> Reduce per pin force required for good contact resistance to lower total load for high pin count and multi DUT probe applications. Evaluation and reduction of probe force requirements to eliminate die damage, including interlayer dielectric damage with lo A chuck motion model is required to minimize probe damage
Probe cleaning	<ul style="list-style-type: none"> Development of high temperature (85°C–150°C) in situ cleaning mediums/methods, particularly for fine pitch, multi-DUT, and non-traditional probes. Reduction of cleaning requirements while maintaining electrical performance to increase lifetime. A self cleaning probe card is required for fine pitch bumped pad devices
Cost and delivery	<ul style="list-style-type: none"> Fine pitch or high pin count probe cards are too expensive and take too long to build. Time and cost to repair fine pitch or high pin count probe cards is very high. The time between chip design completion (“tape-out”) and the availability of wafers to be probed is less than the time required to design and build a probe card in almost every probe technology except traditional cantilever. Space transformer lead times are too long, thus causing some vertical probe technologies to have lengthy lead-times.
Probe metrology	<ul style="list-style-type: none"> Tools are required that support fine pitch probe characterization and pad damage measurements. Metrology correlation is needed for post repair test versus on-floor usage.
High power devices	<ul style="list-style-type: none"> Probe technologies will need to incorporate thermal management features capable of handling device power dissipations approaching 1000 Watts and the higher currents (≥ 1.5 amp) flowing through individual probe points.
Contact resistance	<ul style="list-style-type: none"> Probe technologies that achieve contact resistance $<.5$ Ohms initially and throughout use are needed. A method to measure contact resistance is needed. The traditional continuity test is insufficient to monitor contact resistance.
High frequency probing	<ul style="list-style-type: none"> Traditional probe technologies do not have the necessary electrical bandwidth for higher frequency devices. At the top end are RF devices, requiring up to 40 GHz.

PROBE CARD TECHNOLOGY REQUIREMENTS

Many probe card technology types are available in the marketplace, each with suitability (technical and/or test operations driven) for probing certain device types and limitations that prevent more widespread use. There is no single probe technology capable of addressing the requirements across the entire device spectrum.

44 Test and Test Equipment

This section explores the challenges of probe technologies including those that are independent of the devices being probed. These include the resulting behavior of the probe when/after contacting the wafer, the design of the probe card to realize the productivity benefits of probing multiple die at the same time and the environment that the probe card is expected to operate within.

PITCH AND INTERCONNECT DEFORMATION

I/O density requirements are driving pad/bump sizes to ever-smaller sizes. It is well known that on the leading edge, wirebond pad pitches are under 30 μm (with resulting pad sizes naturally less than that). It is a formidable challenge for traditional probe technologies to scale down continually since with this scaling comes a parallel scaling-down of the permissible probe mark.

The use of cantilever probe cards for probing wirebond technologies, though still today's leading solution, is seen to be reaching practical limits in pitch and scrub within the nearer term horizon. Thus the newer emerging technologies, many using “semiconductor-like” processes (e.g., MEM and membrane structures) offer solutions for reduced pitch scrub requirements.

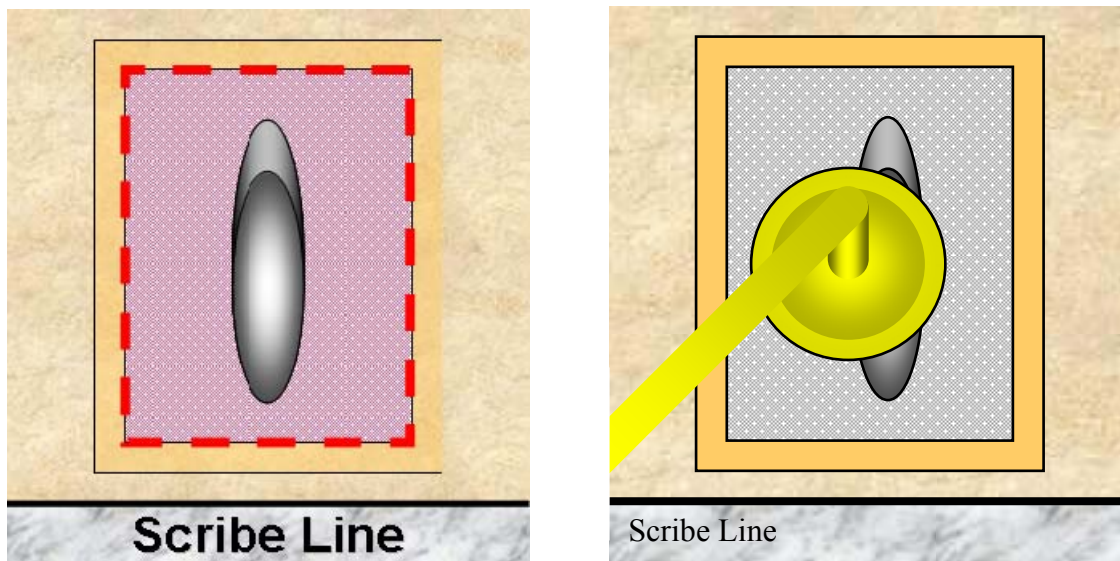


Figure TST12 Probing and Wirebond Contacting a Bond Pad

Area array solder bumps are seeing growing application and driving the commensurate need/demand for vertical probing technologies. As the pitch/bump dimensions get smaller, current vertical technologies, typically an array of guided wires may also see their practical limit, thus requiring development of newer technologies.

MULTI-DUT

Productivity gains are often realized when testing (probing) more than one device in parallel. Memory testing has been a leader in this area, with leading edge approaching 500 devices in parallel. As [Table TST2](#) indicates virtually all memory testing is done in multi-DUT fashion. The move to multiple DUT testing within other product categories is already underway and is accelerating: with the use of DFT and “smart test” techniques, 16, 32, and even 64 DUTs is realizable and up to 4 DUT for high end microprocessors.

Multi-DUT probing requirements drive the need for more and more probe contacts across an ever-growing area. Today some new contact/probe technologies claim full wafer contact capability for 300 mm wafers. Ultimately increasing the contacts/DUT to hundreds will be required.

ELECTRICAL PERFORMANCE

Wafer probe technology, the probe card, provides electrical contact between the device(s) under test on a wafer and the test system electronics. The probe card must faithfully transmit/deliver device under test power and signals from/to the test system.

Within this ITRS document information can be found concerning device operating voltages and AC Characteristics. Additionally, within this Test and Test Equipment chapter tester performance information is provided on a wide range of electrical characteristics that may be helpful in understanding requirements for wafer probing.

There appears to be growth in the current carrying capability of individual probes contacts. At the same time the aggregate total current across the DUT is expected to rise with growing circuit densities and pin counts. Of note is that there are some selected applications that are seeing the need for higher and growing current carrying capability, approaching 1.5 amps and more. Of note is that peak values for transient currents are growing as well.

Contact resistance is always a closely watched probe technology parameter. It is influenced by many factors such as pad/bump metallurgy, contamination from pads/bumps, multi-DUT “off-stepping,” contact force, scrub, cleaning, etc. The values shown in *Table TST16* requirements reflect contact resistance under ‘normal’ usage conditions over the practical lifetime of the probe. Initial and after cleaning requirements are often considerably lower, typically in the 200 milli-Ohm range or lower. There is a growing requirement for lower contact resistance values for longer periods (numbers of touchdowns) before cleaning.

High frequency testing in probe remains a challenge due to the lack of constant impedance structures to the die contact. The roadmap shows digital I/O performance will increase to greater than 20 GHz over the roadmap. Analog pin performance may reach 100 GHz driven by high precision proximity radar and automatic landing and automotive parking systems. High frequency test under High parallelism conditions remains a challenge and significant development is needed to support wafer level KGD (Known Good Die) test.

THERMAL PERFORMANCE

Though stable through the roadmap horizon, the thermal environment for the probe is demanding. With low end chuck set-point requirements well below the freezing point and the upper end past the boiling point, the total range is wide - placing difficult demands on selecting materials that handle the extremes, but possibly more notably to deal with temperature co-efficient of expansion issues and high current demands.

Additionally, handling the heat produced by very high transient current heating effects and/or by high power products may drive the need for active thermal management within probers as well as an improved wafer to chuck thermal interface.

UNIT COST AND COST OF OWNERSHIP

Probe card unit cost and cost of ownership (CoO) trends are not currently covered in this roadmap document. Though individual member companies may have their own approaches to unit cost and cost of ownership measurements and goals, there is a need to develop consistent models that can be used industry wide and cover the wide range of probe card technologies that are in the marketplace.

CLEANING

Generally, online cleaning frequency for cantilever type probes rises slightly through the roadmap horizon, however increasing probe usage (touchdowns) before being taken offline for cleaning is being seen for many of the product families. The goal is better utilization of the test systems and the probe card.

For vertical probes, the rapidly growing number of touchdowns before online cleaning reflects a desire to reduce the vertical technologies’ online cleaning frequency to match/better cantilever technologies more closely. Similar to cantilever technologies, the touchdowns before offline cleaning is increasing but across all product categories.

Notably, in some instances there is a move to eliminate online cleaning for memory products in the outer years of this roadmap’s horizon. This is likely reflective of the design and/or complexity of probes with pin counts approaching full wafer contact.

Table TST16 Wafer Probe Technology Requirements

TEST SOCKETS

The test socket is an electrical and mechanical interface responsible for good electrical connection and transference of high integrity signals between the DUT and the PCB/tester through a mechanical contact mechanism in order to determine the electrical characteristics of DUT. As the semi-conductor designing and manufacturing capabilities have progressed in recent years, the testing process keeps raising the electrical and mechanical requirements of test sockets. Therefore, the socket technologies have been rapidly driven by significantly enhanced electrical and mechanical requirements, both of which are instigated by higher power/voltage/current, reduced package size, tighter pitches, higher pin counts, smaller solder resist opening, and so on. It has been indicated that electrical properties are determined by not only the electrical but also by the mechanical requirements. The multi-physics problems have made socket designs progressively challenging along the higher requirements. Current models show difficulty in making sockets for high ball count devices and achieving I/O bandwidths of > 20GHz.

46 Test and Test Equipment

Table TST17 contains the test socket technology requirements. The requirements have been divided into contacting NAND, DRAM, and SoC devices that are contained in TSOP, BGA, and BGA SoC packages respectively. The TSOP package is assumed to be contacted using a blade; the DRAM BGA is contacted with a spring probe and the SoC BGA is contacted with a 50 Ohm spring probe (Figure TST13). The test socket performance capability is driven by the pitch between balls or leads, so the lead spacing of the assembly and packaging roadmap was used to determine the pitch.

Contact blades are generally used for testing TSOP NAND Flash and contain a spring function in its structure, which is loaded by compressing the DUT into the socket. The structure is very simple and suitable for HVM, however, the contactor blade must be long to maintain the specified contact force, stroke, and achieve a long mechanical lifetime. A weak point is that the blade contactor is not suitable for fine pitch devices due to the need to have isolation walls between adjacent pins. The thickness of the isolation wall must be thinner for finer pitches, which makes fabrication of the isolation wall more difficult. At the same time, the contactor blade thickness needs to be thinner for finer pitch, which complicates achieving the specified contact force, stroke requirement, and mechanical lifetime.

Spring probes are mainly used for testing BGA-DRAM device are formed by use of small-diameter cylindrical parts (probe and socket) and coil springs. Compression of the spring probe creates the contact load. In order to guarantee sufficient mechanical life, the probe diameter should to be large enough to guarantee strength and durability and the length should be long enough to maintain sufficient travel under compression. Negative characteristics of the spring probe are that it can only contact the pad/lead at one or two points and the contact resistance tends to be higher than other types of contactors. However, the spring probe structure is relatively simple and easy to maintain and it is also easy to design a DUT loadboard.

According to the BGA-DRAM roadmap, the spring probe diameter will need to be smaller over time driven by the finer pitch of the package ball roadmap. In addition, the spring probe will need to be shorter to meet the lower inductance values required to support the high frequencies of the roadmap I/O data rate.

Spring 50 Ohm probes are required for BGA-SoC high frequency devices have coaxial structures that can reduce probe length transmission issues through impedance matching. However, advances in the package ball pitch through the roadmap will create restrictions to the coaxial pin arrangement structure (0.5 mm pitch in year 2016). The data rate will increase to 20GT/s in 2016, but the spring 50 Ohm probe will not have good electrical performance due to its multiple parts structure having higher contact resistance than other contactors. To support 50milli-Ohms of contact resistance starting in 2010, advances will be required in materials, plating, and structure.

Conductive Rubber type contactors are used for BGA high frequency SoC devices. Conductive metal particles are aligned vertically in insulating silicone rubber which enables vertical contact and adjacent conductor isolation. Compared to other contacts, it is superior for used with high frequency device test since due to its low inductance and low contact height, but compression travel is limited. Conductive rubber will meet the fine pitch requirement in the road map, but it difficult to reduce contact force without decreasing the compression travel.

Contact blade + Rubber, generally used for testing QFP/QFN high frequency Soc, is a combined structure of a short length metal contact and compression rubber that makes contact thru force and travel. The required compression force can be varied by changing the rubber material, but the cycle life is normally shorter than a Contact Blade type contact.

It is capable to meet 15GT/s I/O data requirement in 2013 by shortening the metal contact, however, it is a challenge to ensure enough compression travel which is required in mass production environment. 40GT/s requirement in 2019 is also a challenge to make the metal contact 0.1nH inductance or less. It will need improvement of the structure, contact material, and plating.

Socket lifetime has not been pursued in this roadmap, but the lifetime problem will become more important in the near future as lead, ball and pad pitch becomes finer and pin counts get higher which drives lower contact force to avoid lead / ball damage. Pb-Free devices require higher contact forces than are required for non Pb-Free packages.

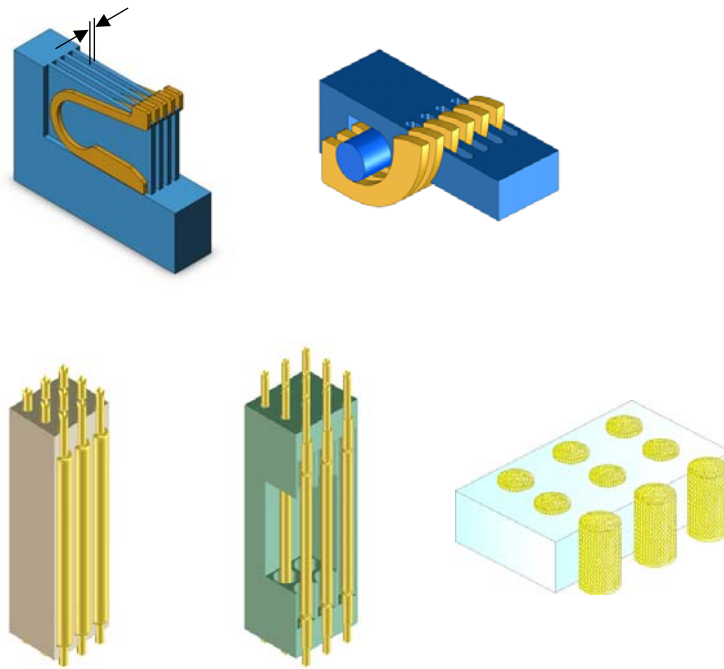


Figure TST13 Contactor Types

ELECTRICAL REQUIREMENTS

Socket electrical requirements include current carrying capacity (CCC) per pin, contact resistance, inductance, impedance, and signal integrity parameters such as insertion loss, return loss, and cross-talk. The higher the power and bandwidth the packages are designed for, the higher the CCC, the lower the resistance, and the better matched the impedance of the pins and/or sockets need to be. Data rate requirements on over the roadmap are expected to exceed 20 GHz, which will greatly challenge impedance matching and the potential signal loss. As package size, solder resist opening, and pitches become smaller and pin counts higher, the smaller pins required to fit within tighter mechanical constraints will greatly increase contact resistance and signal integrity issues. One of the critical parameters to stabilize the electrical contact and insure low contact resistance is the contact force per pin, which generally ranges from 20 ~ 30 grams. As pitches get finer, smaller and more slender pins will be required, which may not be able to sustain a high enough contact force to have reasonable contact resistance. Due to the negative impact of mechanical requirements on electrical properties, it will be necessary to have improved electrical contact technologies or socketing innovations, in which the electrical properties and signal integrity will not be significantly impacted by or will be independent from raised mechanical requirements.

MECHANICAL REQUIREMENTS

The mechanical requirements include mechanical alignment, compliance, and pin reliability. Mechanical alignment has been greatly challenged by higher pin counts and smaller solder resist openings, particularly in land grid array (LGA) applications. Currently, the majority of test sockets use passive alignment control in which the contact accuracy between pin and solder resist opening is determined by the tolerance stack-up of mechanical guiding mechanisms. The limit of passive alignment capability is quickly being reached because the manufacturing tolerance control is approximately a few microns. The employment of active alignment or an optical handling system is one of the options to enable continuous size reduction of package and solder resist opening, smaller pitches, and higher pin counts.

Compliance is considered as the mechanical contact accuracy in the third dimension (Z-dir.), in which the total contact stroke should take into account both the co-planarity of operating pin height and the non-flatness of the DUT pins, in addition to a minimum required pin compression. In general the total stroke of the contact is between 0.3 mm and 0.5 mm. However, as required pin sizes get smaller, it may not be feasible to maintain the same stroke and thus the compression issue may become the bottleneck of electrical contact performance.

48 Test and Test Equipment

Contactor pin reliability and pin tip wear-out have also experienced challenges because tight geometric constraints prevent adding redundant strength to the pins. The testing environment becomes more difficult with higher temperatures, higher currents, smaller pin tip contacts, etc.

Table TST17 Test Socket Technology Requirements

SPECIALTY DEVICES

The test roadmap is not all inclusive so it does not contain test requirements for all devices. Many of the test requirements for some omitted devices fall within the bounds specified for devices within this roadmap. Other devices stretch the bounds specified in this chapter and need to be mentioned for completeness. Devices included in the specialty device section are high volume devices that are generally contained in and driven by the requirements of mobile communication and computing. The intent of this section is to document the challenges of specialty devices. For 2009, LCD display drivers, imaging devices and MEMS accelerometers are featured in this section.

LCD display drivers are unique because of die form factor, which can have a 10:1 aspect ratio (or greater), and by the high number of very narrow pads requiring contact for test. In 2009, LCD display drivers with pads that are $8\ \mu\text{m} \times 150\ \mu\text{m}$ and pitches of $20\ \mu\text{m}$ are planned to be in production. The $8\ \mu\text{m}$ pad width in use in 2009 is less than half of the $25\ \mu\text{m}$ 2022 pad width specified for commodity devices and the pad aspect ratio of 15:1 also exceeds the worst case pad aspect ratio for commodity devices of 1.3:1. Massive parallelism is not needed for these devices due to short test time, but a repeatable economic contacting solution is needed for the probe environment.

Imaging devices are required for every digital camera and have become a standard feature in cell phones and mobile computing devices. Digital cameras with 12 mega pixels of resolution are typical consumer devices in 2009 and will increase in resolution in the future. Automotive industry applications and surveillance need fast frame rates of 60 fps in 2009. Each pixel of the image sensor has a micro lens (Figure TST14) to increase the light intensity and improve the signal to noise ratio and must be tested for consistency, sharpness, and contrast across its detection spectrum. Current production solutions are generally proprietary and composed of a pupil projection system used to project images to the sensor under test over a range of various angles. Back-side illumination BSI is another approach to increase incident light intensity and fill factor of sensor cell of each pixel. Detected image data is processed to ensure picture quality. Image sensors convert the analog information of the picture into digital output, so the test result is non-deterministic. Image sensors coupled with image processing digital logic in a single chip pose significant test challenges when assembled into module form, since typical test access is very limited for module testing, and when embedded in a system such as a cell phone, simple functional tests may be all that is possible.

MEMS accelerometers are increasingly being incorporated into personal electronic devices and automotive application. Some smart phones and personal digital assistants contain accelerometers for user interface control. The cost of accelerometer devices dropped over 10% in 2009, but reduction of test cost is challenged by the difficulty of achieving high parallelism package test and difficulty in enabling probe test data to gain yield improvement.

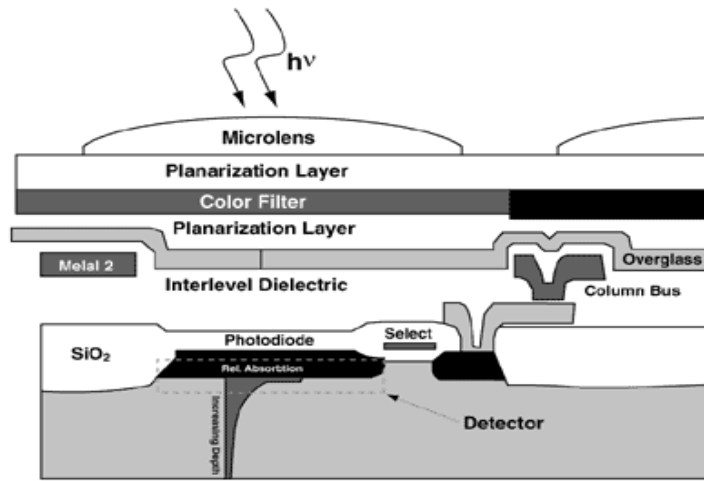


Image sensor structure cross section

Figure TST14 Image Sensor Cell