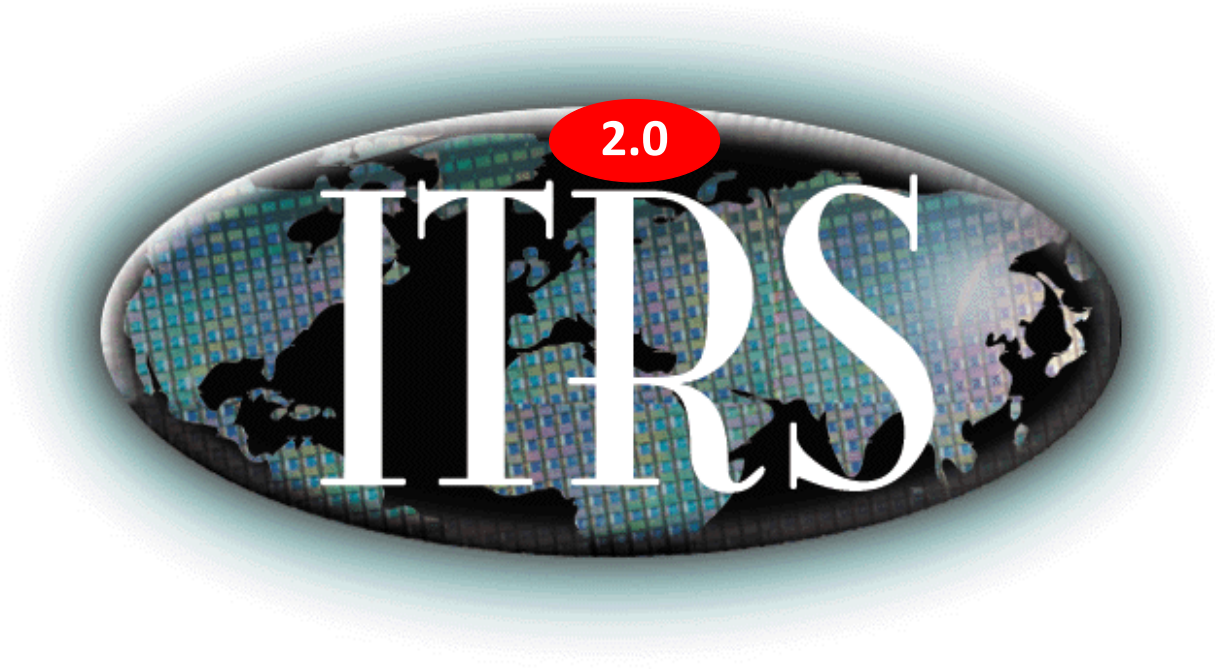


INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS 2.0

2015 EDITION

TEST AND TEST EQUIPMENT

1 INTRODUCTION AND SCOPE OF THE 2015 EDITION 1

2 DRIVERS, CHALLENGES, AND FUTURE OPPORTUNITIES 2

2.1 KEY DRIVERS 3
2.1.1 Increasing Device Integration 4
2.1.2 Process Technology and device architecture 4
2.1.3 Increased device performance 5
2.1.4 Test Process Complexity 5
2.1.5 Cost 5
2.2 DIFFICULT CHALLENGES 6
2.2.1 Test Development As A Gate To Volume Production 6
2.2.2 Detecting System Defects 7
2.2.3 Concurrent Test 7
2.2.4 Wafer level Packaging 7
2.3 FUTURE OPPORTUNITIES 8
2.3.1 Combine Multiple Test Environments 8
2.3.2 Test Environment Simulation 8
2.3.3 Improved Fault Modeling 9

3 TEST AND YIELD LEARNING 10

3.1 ELECTRICAL TEST-BASED DIAGNOSIS 10

4 TEST COST FOCUS TOPIC 11

4.1 COST OF TEST AS A PART OF OVERALL MANUFACTURING COST 12
4.2 TEST COST MODELS AND COST IMPROVEMENT TECHNIQUES 14
4.2.1 Current Top Cost Drivers 14
4.2.2 Future Cost Drivers 14
4.2.3 Currently Deployed Cost Reduction Techniques 14
4.2.4 Cost Reduction techniques that may be Deployed in the Future 15
4.3 INTERFACE COST TREND 15
4.4 MULTI-SITE TREND 15

5 3-DIMENSIONAL DEVICE TESTING 17

5.1 TEST FLOWS, COST, RESOURCES 17
5.2 DFX 18
5.3 TEST ACCESS 18
5.4 HETEROGENEOUS DIE 19
5.5 DEBUG-DIAGNOSIS 20
5.6 POWER 20
5.7 CONCLUSION 20

6 ADAPTIVE TESTING 21

6.1 ADAPTIVE TEST 21
6.1.1 Adaptive Test Definition 21
6.1.2 Adaptive Test Description 21
6.1.3 Example Applications 23
6.1.4 Adaptive Test Architecture / Flows 24
6.1.5 Levels of Adaptation & Data Model 25
6.2 ADAPTIVE TEST INFRASTRUCTURE (DATA EXCHANGE, DATABASES, ETC.) 26
6.3 IMPLICATIONS FOR ATE AND THE TEST CELL 27
6.4 TEST RESULTS DRIVING "ADAPTIVE DESIGNS" 28
6.4.1 Testing resilient features 28
6.4.2 Non-deterministic device behavior: test and run-time availability 28
6.4.3 Testing Adaptive Designs 28
6.5 ADAPTIVE MANUFACTURING 28

6.6 ADAPTIVE TEST FOR CARD/SYSTEM/FIELD 29
6.7 ADAPTIVE TEST CHALLENGES AND DIRECTIONS 29
6.8 SUMMARY 30

7 TEST TECHNOLOGY REQUIREMENTS 32

7.1 INTRODUCTION 32
7.2 SYSTEM INTEGRATION—SOC AND SIP TEST CHALLENGES AND IMPLICATIONS 33
7.2.1 System on a Chip 33
7.2.1.1 Requirements for Logic Cores 33
7.2.1.2 Requirements for Embedded Memory Cores 36
7.2.1.3 Requirements for Integration of SoC 36
7.2.1.4 Concurrent Testing 37
7.2.1.5 DFT for Low-Power Design Components 39
7.2.2 System in a Package 39
7.2.2.1 Stacked Die Testing and Equipment Challenges 40
7.2.2.2 Wafer Testing and Equipment Challenges/Concerns 40
7.2.2.3 Wafer Test for RF Devices 40
7.2.2.4 Reliability Screening at the Wafer or Die Level 41
7.2.2.5 Statistical Processing of Test Data 41
7.2.2.6 Subsequent Processing Affects the Quality of the Die 41
7.3 LOGIC 41
7.3.1 High Volume Microprocessor Trends Drivers 41
7.3.2 System Trends Drivers 42
7.3.3 DFT Trends Drivers 42
7.3.4 ATE Trends Drivers 43
7.3.5 Logic test assumptions and context 43
7.3.6 Logic test calculations 45
7.4 HIGH SPEED INPUT / OUTPUT INTERFACE 46
7.4.1 IMPORTANT AREAS OF CONCERN 49
7.4.1.1 Jitter and Noise Decomposition and Measurement 49
7.4.1.2 Jitter/Interference Tolerance Test 50
7.4.1.3 Test Fixture Bandwidth 51
7.4.1.4 DFT and TFD 51
7.5 MEMORY 51
7.5.1 DRAM 52
7.5.2 Flash 52
7.5.3 Embedded Memory 53
7.6 ANALOG AND MIXED-SIGNAL 53
7.6.1 Important Areas of Concern 54
7.7 RADIO FREQUENCY 54
7.7.1 Important Areas of Concern 54
7.8 RELIABILITY TECHNOLOGY REQUIREMENTS 55
7.8.1 Burn-In Requirements 56
7.8.2 Wafer Level Burn-in 57
7.8.3 Probing Technology for Wafer Level Burn-in 58
7.8.4 Other WLBI Technology Considerations 58
7.9 TEST MECHANICAL HANDLING REQUIREMENTS 58
7.10 DEVICE INTERFACE TECHNOLOGY REQUIREMENTS 61
7.10.1 Probe Cards 61
7.10.2 Trends Affecting Probe Card Technologies 62
7.10.2.1 Probe Card Technology Requirements 64
7.10.2.2 Pitch and Interconnect Deformation 64
7.10.2.3 Multi-DUT 64
7.10.2.4 Electrical Performance 65
7.10.2.5 Thermal Performance 65
7.10.2.6 Unit Cost and Cost of Ownership 65
7.10.2.7 Cleaning 65

7.10.3 Test Sockets 65
7.10.3.1 Electrical Requirements 67
7.10.3.2 Mechanical Requirements 67
7.11 SPECIALTY DEVICES 68
7.11.1 LCD display drivers 68
7.11.2 Imaging devices 68
7.11.3 MEMS sensor device 68
8 ACKNOWLEDGEMENTS 69

List of Figures

Figure TST1 – Test Costs as a percentage of device ASP (Used by permission of VLSI Research) 11
Figure TST2 – Relative cost of Fab, Packaging and Test Equipment 12
Figure TST3 – Cost of Test Reduction realized by traditional cost reduction techniques 13
Figure TST4 – Total Cost of Manufacturing Reduction realized by traditional cost reduction techniques 13
Figure TST 5 – Multi-die Flow 14
Figure TST6 – Importance of Multi-Site Efficiency in Massive Parallel Test 1
Figure TST 7 - Adaptive Test supports feed-forward and feed-back data flows. Adaptive Test provides for data use for test decisions either in-situ (i.e. during a given test step) or post-test. 22
Figure TST 8 - The architecture of Adaptive Test organizes each insertion's test data into one or more databases. A waterfall of manufactured parts may insert, join or query databases for test flow decision-making. 25
Figure TST9 - Organization of Cores for System Integration and Applications 32
Figure TST10 – DFT Compression Factors (Flat with No Compression = 1) 34
Figure TST11 - Scan test view of a core 43
Figure TST12 - SOC with multiple copies of Core_2 45
Figure TST13- Core logic vs. random logic between cores 45
Figure TST14 - The Production Process with WLBI Compared with Package Burn-in. 57
Figure TST15 - Contactor Types 67

List of Tables

Table TST1a- Summary of Key Test Drivers 3
Table TST2b- Summary of Difficult Challenges 6
Table TST3c- Summary of Future Opportunities 8
Spreadsheet Table TST7 – Logic Assumptions 41
Spreadsheet Table TST8 – Logic Test Data Volume 41
Spreadsheet Table TST9 – Logic ATE Requirements 41
Spreadsheet Table TST10 – Memory Test Requirements 52
Spreadsheet Table TST11 - Mixed-Signal Test Requirements 54
Spreadsheet Table TST12 – RF Test Requirements 54
Spreadsheet Table TST13 - Burn-In Test Requirements 57

1 INTRODUCTION AND SCOPE OF THE 2015 EDITION

The 2015 revision of the Test Technology Roadmap identifies many significant changes to the industry as well as a number of significant challenges which the industry is working to overcome.

1. Notable Changes
    a. Cost-of-Test – Reducing the cost-of-test, while always an industry goal, has begun to suffer from diminishing returns when looking at the traditional approach such as capital cost reduction or site count increases. As part complexity increases, test times and thus cost-of-test has also been increasing and driving up the cost-of-test. Additionally, the cost of "consumable" test interface material has come to play a much larger role in cost-of-test. Interestingly, the overall cost-of-test is becoming a lower percentage of the overall IC manufacturing expense while at the same time test is having a greater impact on profitability.
    b. Heterogeneous Integration – As more and more 2.5D and 3D device integrations come on the market it has become clear that additional testing is necessary to avoid scrapping expensive sub-assemblies.
    c. Wafer probing – After reporting big changes in probe pitch and counts in 2013, we once again need to adjust the requirements in 2015. These changes are largely in response to the broad deployment of microbumps and pillar type interconnects which drive up the density and down the pitch.
    d. Logic Table – The product groupings have changes with this publication. The new groupings are:
        i. MPU-HP (Big Data) – High Performance MPU (MPU, Servers, GPU)
        ii. MPU-CP (Mobile) – Mobile Consumer MPU (SOC, APU)
        iii. SOC-CP - (Low power Devices) – Low-Power Consumer (MCU, IoT, MEMs, Analog)

    This product grouping was chosen to provide a better alignment with the ITRS 2.0 documents. An additional benefit of this grouping is it provides a better understanding of the challenges in the testing of low power devices.
2. Areas with Significant Challenges for the Future:
    a. Thermal – Various needs are pushing the performance envelope for high-performance thermal solutions both at wafer probe as well as in a packaged device environment. More sophisticated automotive electronics demand testing to a wider temperature range. In addition, long test times are pushing scan speeds up resulting in a need for better device cooling during test.
    a. Adaptive Device Testing – This technique continues to deliver significant value to the industry. Looking forward this technology is challenged to broadly implement real-time feed forward and feedback systems in an industry which has increasingly complex supply chains. Adaptive test may prevent an increase in test cost.
    b. Parallel Testing – The typical parallel testing site counts for most products are not changing. With the increased device complexity this is one of the driving forces in an increasing cost-of-test. Areas where the industry is working are to increase parallel site counts are RF device testing, LCD testing, and MEMs device testing.
    c. Logic Device Testing – With the broad deployment of FinFET technology the complexity of digital devices continues to increase. Further enhancing this trend is the trend toward heterogeneous integration. While pattern compression techniques help manage the challenge of test data volume it does little to help reduce test times and hence cost of testing.
    d. Optical Interface Testing – In 2015 we are starting to see a trend toward the use of optical interfaces in the high-speed SerDes area. Testing optical device brings with it big challenges for precise fiber alignment as well as new test methods when working with light as a signal source.

Challenges and changes in the other topic areas are less dramatic.

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

.

2 DRIVERS, CHALLENGES, AND FUTURE OPPORTUNITIES

This section provides a top-level view of the general challenges facing Semiconductor test. It is organized into three sections

- Key Drivers: These are the high level technical and economic trends which will drive IC test overall
- Difficult challenges: These difficult challenges define areas where more development or understanding is needed in order to cost effectively meet the semiconductor roadmap.
- Future Opportunities: These are longer-term areas of research and development that would address current or future challenges in the area of semiconductor test.

2.1 KEY DRIVERS

Table TST1a- Summary of Key Test Drivers

Key Drivers (not in any particular order)		Effects & Solutions
Device Integration	*Increasing package-level integration (PoP, SoC, SiP, MCP, 2.5/3D packaging)*	• *Increasing reliance on Application-Level Test* • *Increasing reliance on "Known" good die* • *Full functional test at wafer* • *Integration of non-electrical stimulus into test cells* • *Increased need for standardized DFT in third-party silicon IP*
	Multilayer Silicon integration ("Monolithic 3D")	
	Integration of new RF and sensor functions on CMOS digital die.	
	Integration of non-electrical devices (optical, MEMS, etc)	
	Complex package electrical, mechanical, and thermal characteristics	
	Heterogeneous integration dependent on 3rd party IP test solution readiness	
Process Technology and device architecture	*Process variability that requires device modification post-fabrication (Calibration, Trim, etc.)*	• *More iterative test flows based on measured results* • *Increased need for device reconfiguration (one-time programming)* • *Higher Accuracy and more stable Device Power Supplies*
	Fault Tolerant Architectures and Protocols (redundant components, Field Repair)	
	Lower voltage and more numerous Supply Voltages - Power Integrity (like signal integrity)	
Increased Device functionality	*Increasing device interface bandwidth (# of signals and data rates)*	• *Higher performance instrumentation*
	Complex RF Modulation standards	
Test process complexity	*Feedback data for tuning manufacturing (fab, packaging, etc.)*	• *Unique stimulus per device based on measured results* • *Need to generate and transport large volumes of test data* • *Need to dynamically alter test flow and test limits* • *Ability to program device controllers to facilitate functional test*
	Dynamic setting of pass/fail criteria (outlier detection)	
	Dynamic test flows via Adaptive Test (skip or add tests, change flow based on measured results)	
	Device test after part is made secure	
	Integration of more complex test and device handling functionality in test cells	
	Maintaining Unit level Traceability Through manufacturing process.	
Cost	*Physical and economic limits of test parallelism*	• *Improve test Fixturing to accommodate higher site counts* • *Improve cost models to incorporate overall efficiency and effects of yield* • *Improve instrumentation accuracy to lower guardbands* • *Improve Parallel Test Efficiency* • *Improve handling and interface equipment to support higher site count* • *Increased need for thermal management at test*
	Managing overall test operating costs by balancing tester configuration, device-handling equipment, interface hardware and Overall Equipment Efficiency.	
	Additional test steps driven by customer quality requirements or packaging complexity	
	Increasing effects of device yield on overall production costs.	
	Increasing test times due to scan data volume and thermal limitations of scan shift rates	

ATE—automatic test equipment ATPG—automatic test pattern generation BIST—built-in self test

2.1.1 INCREASING DEVICE INTEGRATION

Device integration happens in two ways:

- SOC integration adds more functionality on the same die
- SIP integration integrated multiple die into the same package. This path is chosen because different functions are implemented in different silicon processes or are produced by different manufacturers, and it is more efficient to combine them at the package level rather than moving functions to the same die. It also allows combining purely electrical devices with mechanical/electronics devices like sensors, MEMs, etc.

These drive two increasingly important test trends

- Before they are shipped to customers, the device must be validated to function in its end application. Most manufacturing test has been migrated to a "superposition" model, where different portions of devices are tested in isolation through structural test, or parametric tests that are performed on isolated analog functions. These techniques leave untested the ability of different portions of the device to function with each other, typically at different operating frequencies with random timing. Increasing, test insertions are being added at which the device runs in "mission mode". Since this adds cost to the overall manufacturing process, techniques must be developed to provide the same test coverage more efficiently as an existing test insertion
- If devices are constructed from multiple die, it is critical that the failure rate of any die be extremely low since having to scrap the final SIP is extremely expensive. Thus these "known" good die must be exhaustively tested before they are integrated into the final package. This will require improved signal performance at probe and the need to perform multiple probe tests without mechanical damage.

In addition, the need for non-electrical stimulus at test will become more prevalent in the test of more highly integrated devices. Non-electrical stimulus and responses may include, but are not limited to:

- Optical communications and contacting
- Pressure including sound
- Rotation
- Gravity
- Chemical, molecular, or protein sensing including Raman spectroscopy, florescence, and other techniques
- Temperature
- Humidity
- Magnetic Field
- Acceleration including vibration and shock
- Fluid flow

2.1.2 PROCESS TECHNOLOGY AND DEVICE ARCHITECTURE

Shrinking dimensions and new structures of silicon processes have evolved to the point where being able to produce devices that function precisely as designed cannot be relied upon. There are two underlying problems:

- Every wafer has enough variations is electrical performance that it is impossible to rely on each device to perform precisely the same. Most high performance devices are designed with this in mind, so mechanism are built into the design that allows the device performance to be adjusted after it is built. This includes "tuning" analog parameters such as voltage levels, frequencies and overall power consumption. This final manufacturing step is now lumped into the "test" process. As a result, the model of a known stimulus causing a known response has evolved to the point where device stimulus must be iteratively adjusted independently on each device based on the electrical characteristics measured at the beginning of the test process. This is a significant change to the model of test as it has historically been done and will accelerate with future fabrication technologies. This will require changes to the operation of test equipment in order to make that process as efficient as possible.
- Defects will always occur in the fabrication process, usually measured in terms of defects per area of the wafer (N defects per square millimeter, for example). As die become smaller, the chances of having some defect on a

given device increases greatly. To compensate for this, increasing amounts of redundant circuitry is added that can be used to replace defective portions of the device, either when the device is manufactured or later when the device is being used in its final application. This has long been a technique used on memory manufacturing, and will become more prevalent in SOC devices over time. Again, this will drive the need for test to very efficiently adapt to the functionality observed during initial testing.

Another artifact of new process technology is that operating voltages are steadily decreasing. This will drive the need to greater accuracy in test equipment at these lower voltages in order to optimize yield.

2.1.3 INCREASED DEVICE PERFORMANCE

Analog performance of devices will continue to improve. A good example of this is the increased performance required of RF modulation/demodulated circuitry as wireless data rates are increased. In concert with cost reduction enabled by Design-For-Test features, the electrical performance of certain test instrumentation will need to continue to match these improvements.

2.1.4 TEST PROCESS COMPLEXITY

In addition to the requirements described above, the test process has evolved in two significant ways based on the capability to better integrate test into the overall production process.

- Test Flows and limits are adjusted based on statistical analysis of results either at the current test insertion or information fed forward from earlier in the production process.
- Test results are fed back to the design and fabrication much more quickly to improve device yield. This correlates to the shorter timelines required to development new device and deploy them to high volume manufacturing.

These developments are driving a rapid increase in the amount of data produced by the test process, and the need to transport and analyze that data quickly and securely. In addition, the collection of that data cannot reduce the throughput of the test cell. This will drive new standards for data file formats and transportation mechanisms to and from the test cell.

2.1.5 COST

Economics are the prime driver for any high volume manufacturing process. Historically, test costs have been reduced by reducing the amount of capital equipment required to produce a given number of devices. This has accomplished by improving parameters such as cost per pin of test equipment, test time and test parallelism (site count or concurrent test).

This approach will have to change for several reasons

- The cost of test has become a very small fraction of the overall cost of producing semiconductors. The return on investment for new avenues of test cost reduction is very limited as compared to investment in other aspects of the manufacturing process such as reducing fabrication costs and improving Time To Market
- Site count as a test cost reduction technique is naturally limited by handling and interface costs, as well as the volume of devices tested. Since the number of devices to be produced is finite, so are the gains from increasing the throughput of a given test cell.
- Improvements in test time that can be achieved are decreasing asymptotically. Equipment overhead has been reduced to the point where it is a very small portion of the overall test time. In addition, average test times are increasing due to increased device complexity, manifested in the form of increased scan test times and overall increases in the number of tests required per device.
- Improving quality of test has is much more beneficial economically than reducing cost of test. Any extra costs caused by test are more than compensated for by the reduction in scrap costs that are achieved by improved yield. Focusing solely on test costs ignores the much greater benefits of reducing fab costs (which are far greater) by investing in better test equipment that would drive an increase in yield. See section 4 for a more detailed discussion on this topic.

2.2 DIFFICULT CHALLENGES

Listed below is selected group of technical challenges for test that will drive new capability.

Table TST2b- Summary of Difficult Challenges

Difficult Challenges		Effects & Solutions
Test Development as a gate to volume production (Time to Market)	*Increasing device complexity driving more complex test development and significantly higher pattern lengths.*	• *Improve software generation tools* • *Improve offline software quality tools* • *Improve tools to manage software from multiple developers* • *Improve handling of larger programs and patterns* • *Reduce complexity of interface hardware*
	Maintaining Test Program quality	
	Complex Test hardware development time	
	Lead time to implement design/coverage changes (new or iterating tests)	
	Accommodating multiple developers working simultaneously	
	IP protection / Security requirements limits collaboration and adds delays to test program development	
	Automatic Code Generation (Analog and Digital)	
Detecting Systemic Defects	*Testing for local non-uniformities, not just hard defects*	• *Correlate ATE-based parametric measurements, structural tests or functional tests to system-level faults*
	Embedded software/firmware flaws corrupting test results	
	Erratic, non-deterministic, and intermittent device behavior in end use configuration	
	Mechanical damage during the testing process	
Concurrent Test	*Implementation Complexity*	• *Refocus efforts on multi-die packages which have inherent partitions*
	Limited number device types which will benefit	
Wafer Level Packaging	*"Known" Good Die*	• *Improve performance of probe interface hardware* • *Development new handling technology for singulated / thinned die* • *Reconstituted wafer testing*
	Singulated die testing	

2.2.1 TEST DEVELOPMENT AS A GATE TO VOLUME PRODUCTION

The desire to reduce the time required to start high volume production of new, complex devices in a constant force in the semiconductor industry. Test development and debug is one of the most "exposed" portions of the process given that it is gated by the device design and fabrication and is the last step before production can begin. There are a number of contributors to this process becoming more complex.

- The sheer number of Lines of Code in test programs is continuing to increase to implement device characterization, as well as the requirements for variable device flows and the implementation of tests to trim various device parameters. In addition, the volume of pattern data continues to grow (See section 7 for a more detailed analysis).
- Multiple test developers are deployed to simultaneously construct different portions of the overall test program. It must be possible to seamlessly combine the work of these developers, and the work of one developer cannot interfere with work of any other. Furthermore, the number of collaborators may be restricted by IP security or similar concerns.
- Operation of the device during test is dependent on embedded software and firmware which may itself be defective.

All of this combines to drive extremely complex code, typically developed by engineers with limited software expertise. *The resultant issues in the quality of the test code are the primary source of overall test quality problems*.

The issue of test program quality must addressed by techniques common to the software industry overall. For example:

- The automated generation of pattern data from devices simulation data is a well-controlled process. This automated process must be extended to the test program itself through the use of standardized code libraries and test definitions. While this has been a goal of the test industry for many years, success to date is not widespread for complex SOC devices.

- Offline software test must be further developed to simulate various program flows, undefined or incorrectly defined parameters, incorrect program flow paths and binning, etc.
- Tools for the integration and management of multiple stand-alone test programs must be improved.

2.2.2 DETECTING SYSTEM DEFECTS

As noted earlier, the desire to eliminate the need for system-level test insertions to guarantee the quality of complex devices is driven by the need to reduce overall production costs. Finding a way to provide test coverages for these types of faults is a continuing challenging. See the discussion in the next section for possible paths to solve this problem.

2.2.3 CONCURRENT TEST

The adoption of concurrent test has been on ongoing challenge, mostly owing to the inability of most devices to utilize it.

2.2.4 WAFER LEVEL PACKAGING

Wafer level packaging will continue to proliferate and will drive the need for more efficient handling of singulated die and need for additional wafer test insertions for temperature test or other reasons.

2.3 FUTURE OPPORTUNITIES

Listed below are several areas with the potential to yield significant improvements in test efficiency and effectiveness

Table TST3c- Summary of Future Opportunities

Future Opportunities (not in any order)	
Combine Multiple Test Environments	*Merge Lab, Silicon Test and System-Level Test environments to reduce costs and improve Time To Market*
Test Environment Simulation	*Integrate various simulation capabilities and methodologies for device and test equipment into one.*
Improved Fault Modeling	*Eliminate redundant testing with more sophisticated modeling to find system-level and other un-modeled faults*

2.3.1 COMBINE MULTIPLE TEST ENVIRONMENTS

There are essentially two environments for device testing.

- The first is used for device bring-up and silicon debug, mostly in a laboratory environment. This environment uses tools and software associated with the design environment and/or end-use programming. It provides the ability to manipulate various device functions via JTAG or some similar path and uses code that is easily ported from EDA tools. End-use customers also use this environment for developing embedded software and firmware to enable applications development for the device.
- The second programming environment is that used for high volume manufacturing test on traditional ATE. The purpose of this environment is to find manufacturing defects and assumes the design of the device has already been validated. The link to the design environment is more indirect, typically just consisting of the ability to convert device simulation data into test patterns.

As noted earlier, devices with multiple heterogeneous cores have manufacturing defects that are only found through functional "mission mode" test of the device, which is almost always developed as BIST functions in the development software environment.

Device bring-up and transfer to high volume manufacturing would be made far more efficient if these two environments were both available on ATE. Benefits would include:

- The ability to easily correlate scan or parametric tests to system-level faults
- The ability to develop BIST functionality that works together with the "external" test capability of the ATE to facilitate system-level test coverage on an existing ATE insertion
- The ability to run tests developed in the engineering development environment on a large population of devices on the ATE hardware utilizing the handling, temperature control and data collection capabilities on production-oriented equipment.
- Shortening the test feedback process to design engineers by having test, DFT and design engineers working on the same equipment.

2.3.2 TEST ENVIRONMENT SIMULATION

Numerous attempts have been made over the years to provide a single simulation environment that encompasses device simulation with simulation of the tester environment. Functional simulation of the tester and device has largely been abandoned due to its complexity and the more effective solution of improving device simulation capability, improving the EDA to test software path, and improving and standardizing test implementation on ATE.

There are still benefits to be gained, however, by focusing on the analog simulation of the signal transmission path from the tester to the DUT itself, especially in the context of improving the design and fabrication of tester interface hardware.

2.3.3 IMPROVED FAULT MODELING

In conjunction with the discussion in 2.2.1, the detection of system-level faults would be facilitate by more comprehensive device fault modeling, including the interaction of heterogeneous cores in the device operation in their end application modes.

3 TEST AND YIELD LEARNING

In addition to the normal sorting function, test provides the essential feedback loop for yield-learning. Specifically, product-based diagnostics are increasingly vital. The need to base diagnostics on actual product hardware is driven by systematic defect mechanisms now being increasingly complex functions of neighboring shapes, local pattern densities, etc. As a result many failure mechanisms may be visible only on product. In addition, product-based diagnostics automatically put focus on key yield-limiting failure mechanisms. Volume-based diagnostics are critical since individual occurrence of any given systematic defect mechanism may be rare. The pooling of data across many failing die are needed to identify true systematic mechanisms.

3.1 ELECTRICAL TEST-BASED DIAGNOSIS

Parametric-related feedback is needed for (1) device and interconnects 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, including within-standard-cell and transistor-layout configurations, it becomes necessary to base learning on product test or test of product-like layout configurations. Embedded, distributed monitor circuits such as thermal and VDD sensors, process-monitoring ring oscillators and critical path proxies are now standard on microprocessor-class ICs and can be used to help diagnose parametric fails and understand variability. Understanding variability includes unraveling the structure of variations into spatial and cross-parameter components (variation in transistor length, Vt, source-drain resistance, etc.) The spatial component includes both die-to-die and within-die components.

Cross-parameter variations, potentially including a spatial component, are important to analog/RF circuits, as well as digital. Methods for understanding/characterizing the manufacturing process and operating environment that are sensitive enough for analog/RF are needed. Moreover, 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.

Top challenges for test-based yield-learning include:

- Better resolution for cell-internal defects. Latest advances in structural testing and scan-based logic/layout-aware diagnosis methods are adequately addressing interconnect and via defects. Statistical approaches built into volume-based diagnostics are able to predict interconnect-related defect modes without an over-dependence on Physical Failure Analysis (PFA). Innovation is required, however, for cell-internal-defect-targeted diagnostics to be able to identify systematic fail modes inside standard cells. Observations of hardware suggest a shift toward a larger percentage of the defect distribution being cell-internal defects, as opposed to interconnect-related. Current best methods for cell-internal defect diagnostics are cell truth-table and gate-exhaustive model-based, with the truth tables established via static SPICE simulations of modeled cell-internal parasitics. These methods suffer from aliasing issues and over-reliance on potentially inaccurate modeling of cell-internal defects used in SPICE simulations. In addition, diagnosis resolution needs to be better due to limitations in the PFA process.
- Managing design data for yield learning. A tremendous amount of design data can be brought to bear for yield learning purposes, but it is often not organized effectively for this purpose. In addition, with hierarchical design and DFT flows, the overall management of this data at most companies today is ad-hoc and limiting its effective use.
- Inadequacy of LEF/DEF as the basis of layout-aware diagnosis. LEF/DEF suffices for the purposes of layout-aware ATPG but is too early in the design cycle to be used effectively for layout-aware diagnosis. LEF/DEF is getting farther and farther away from final mask shapes due to complex OPC, boolean and retargeting steps.
- Yield-Learning in a OSAT/Fabless/Foundry environment. Yield-learning capabilities must be cognizant of the environment that has become the dominant model for our industry. If the technology cannot deal with

the security and logistical concerns of this environment, it cannot be effective. 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.

- Test and data-collection time increases due to longer scan chains. These increases drive a need for focus on LBIST methodologies and scan compression for both test and diagnosis.

4 TEST COST FOCUS TOPIC

Minimizing costs are a key part of the semiconductor manufacturing process. Test is no exception, although steady improvements in efficiencies over the last 15 years have lowered the typical cost of test as a percentage of IC revenue to less than 2-3%. The primary drivers of increased efficiency have been reductions in capital costs per resource and test times , coupled with increases in parallelism and Built-In Self Test (BIST) capability. Most SOC device are tested 2 to 16 at a time, and memory devices can have in excess of 1,000 devices tested at once. Measured as the cost to use capital equipment for test (in terms of cost per hour per device), these decreases in test cost will continue at a relatively consistent rate per year. The figure below shows the historical rate of capital investment in test, interface (consumables) and handling equipment.

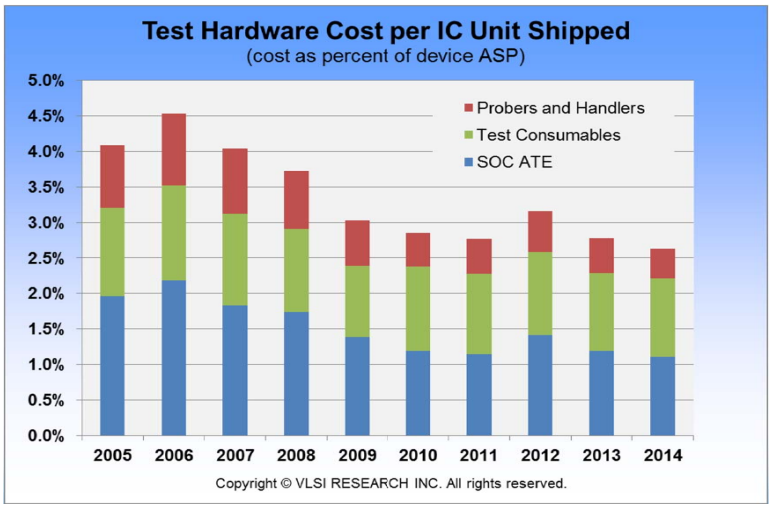


Figure TST1 – Test Costs as a percentage of device ASP (Used by permission of VLSI Research)

Looking forward, there are several trends which will counterbalance equipment efficiency and serve to cause cost increases:

- Increases in transistor count that outstrips compression technology will increase the amount of external data which must be supplied to the Device Under Test (DUT). Coupled with scan shift rates that are limited by power and thermal concerns, the overall effect will be longer test times. This will be addressed primarily with increased parallelism
- Device configuration and one-time programming during test is causing more time to be spent during test to perform initial device calibrations or to reconfigure devices based on defects or electrical performance. As silicon geometries shrink and defect densities drive circuit redundancy, repair functions will also add to test costs
- The eventual drive to multi-die packages will add a requirement for more System Level ("mission mode") testing owing to lack of access to individual die. Without significant Design For Test (DFT) improvements, this

type of testing can take much longer than conventional structural test. This will also drive more exhaustive test processes at wafer probe in order to improve the yield of multi-die packages

Even though continuous improvement in equipment efficiency will be offset by new device test requirements, the overall cost of test will continue to decrease. The major contributors to that cost are described in section 4.2

4.1 COST OF TEST AS A PART OF OVERALL MANUFACTURING COST

While the cost to own and operate test equipment has been reducing, other semiconductor manufacturing costs have been significantly increasing with new silicon technology. Specifically fab costs for leading-edge processes have increased to about 70-80% of the overall cost of producing a large-scale SOC device.

It now costs far more to fab a device than to test it, and that trend will accelerate as new fabrication technologies are deployed..

The figure below represents third-party analysis of the capital and service costs of equipment used in device fabrication, packaging and test.

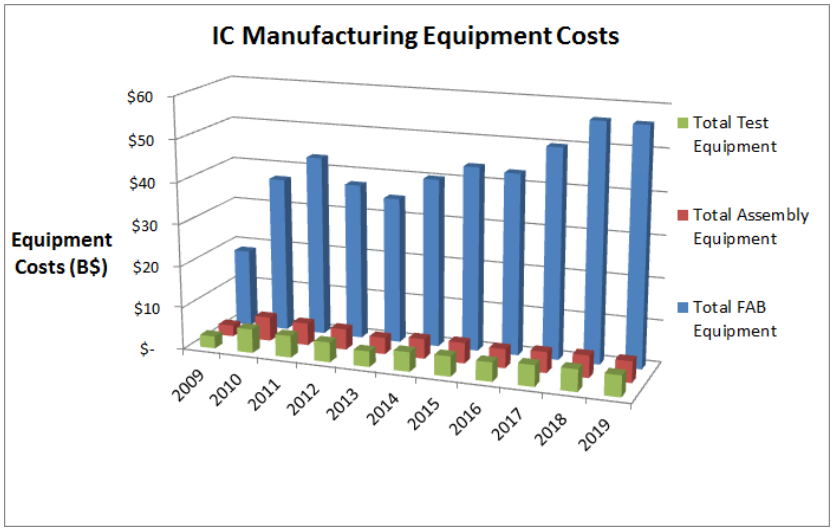


Figure TST2 – Relative cost of Fab, Packaging and Test Equipment

While it is helpful to focus on the cost of test itself, the overall contribution to a manufacturer's profitability from lower test costs will be very small since test is a small part of the device cost overall. The highest avoidable costs in test are devices that are good but are rejected at test for some reason.

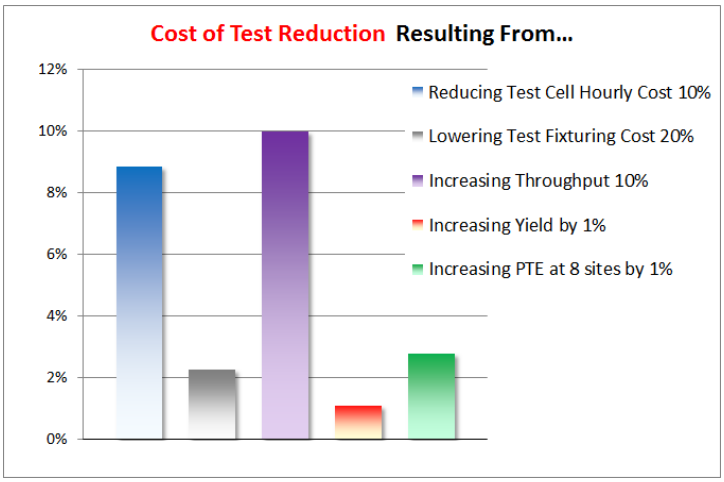
Consider the following, simplified example.

- A device costs $1.00 to manufacture, including Fabrication, packaging, etc.
- Test constitutes 5% of that cost, or $0.05

Reducing the cost of test by 10%, will reduce overall costs by $0.05 X 10% = $0.005 per device
Improving yield by 1% reduces overall cost by $1.00 * 1% = $0.01 per device

While the 10% Cost of Test reduction is good, the yield improvement is better.

The figure below shows the effect on cost of test of traditional cost reduction techniques:



Example Parameters	
Revenue per device	$10.00
Device production costs before test	$3.00
Test Time (Seconds)	10
Sites	8
Parallel Test Efficiency	98.0%
Test Fixturing Cost	$200,000
Test Fixture Depreciation (months)	18
Yield	90.0%
Test Cell Cost/hour	$100

Figure TST3 – Cost of Test Reduction realized by traditional cost reduction techniques

If one considers the effect on total manufacturing costs, including the cost to scrap devices that are actually good, the cost saving due to improved yield becomes far more significant.

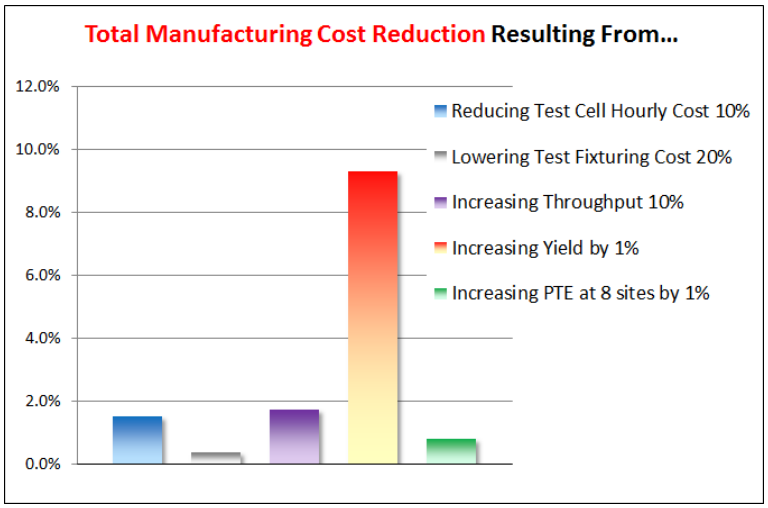


Figure TST4 – Total Cost of Manufacturing Reduction realized by traditional cost reduction techniques

The risk of yield loss is increasing over time for several reasons:

- Trends such as the reduction of power supply voltages and more complex RF modulation standards will drive higher accuracy requirements for test equipment. Test equipment accuracy is typically added as a "guardband" in testing, reducing the range of acceptable measurements. If measured DC and AC values become smaller and there is no improvement in test accuracy, this guardband will cause more marginal (but good) devices to be scrapped.
- As noted earlier, many devices, especially for mobile applications require some sort of calibration or trim during the test process to improve DC and AC accuracy. This dramatically increases both the number of measurements made and the accuracy required of the test equipment. The requirements increase the chance of discarding devices that would otherwise have been good.
- Faster production ramps and short IC product life cycles will reduce the amount of time available to optimize measurements for the majority of devices produced

The remainder of this chapter will examine Costs associated with owning and operating test equipment. It must be stressed that reducing these costs must be done in the context of the overall cost to produce devices and balance reduction in test costs with potential reductions in product yield.

4.2 TEST COST MODELS AND COST IMPROVEMENT TECHNIQUES

The cost of semiconductor test has many drivers as shown in which is further complicated for multi-die SiP precuts as shown in **Error! Reference source not found.**.

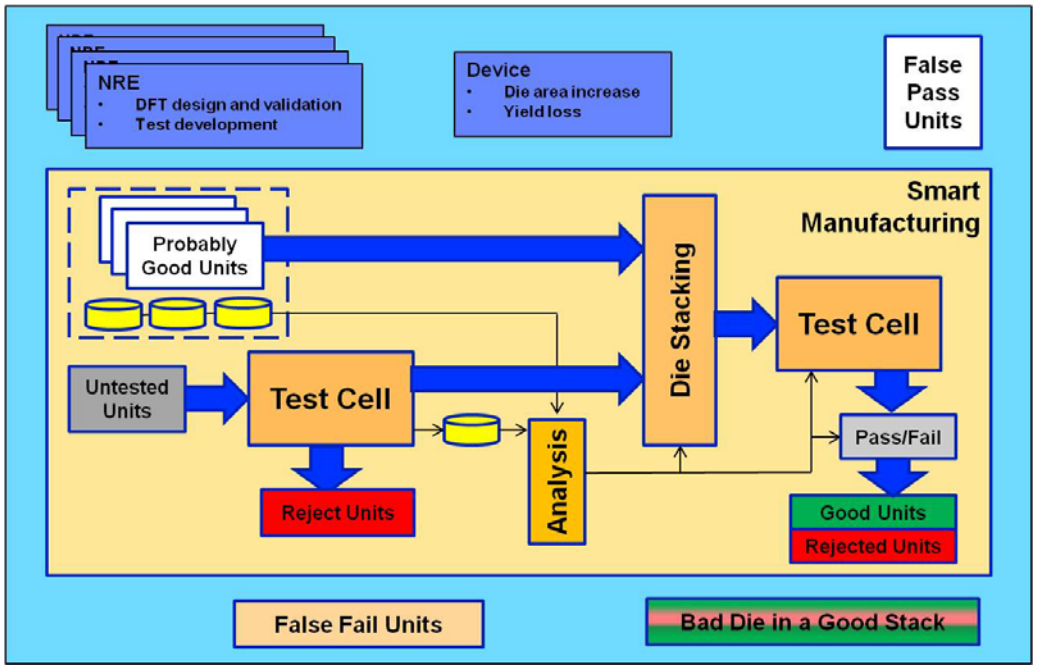


Figure TST 5 – Multi-die Flow

4.2.1 CURRENT TOP COST DRIVERS

The traditional drivers of Test Costs typically include (In rough order of impact to Cost)

- Device Yield
- Test Time, site count and Parallel Test Efficiency (PTE)
- Overall Equipment Utilization
- ATE Capital & Interface Expenditures
- Facility/Labor costs
- Cost of Test Program Development
- Cost of die space used for Test-only functions

4.2.2 FUTURE COST DRIVERS

- Increased test time due to larger scan patterns
- Increased testing at wafer to produce Known Good Die (KGD)
- Addition of system-level testing
- Increased cost of handling equipment to support high site count or singulated die
- Increasing use of device calibration/trimming at test or device repair with redundant components

4.2.3 CURRENTLY DEPLOYED COST REDUCTION TECHNIQUES

- Multi-site & reduced pin-count
- Structural Test & Scan

- Compression/BIST/DFT and BOST
- Yield Learning & Adaptive Test
- Concurrent Test
- Wafer-level at-speed testing

4.2.4 COST REDUCTION TECHNIQUES THAT MAY BE DEPLOYED IN THE FUTURE

- Advanced embedded instruments
- New contacting technologies
- In-system level testing to detect latent defects and potentially repair
- Built in fault-tolerance

4.3 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.

Figure TST6 below shows the relative cost of Test equipment, handling equipment and interface material (consumables) in an SOC test cell

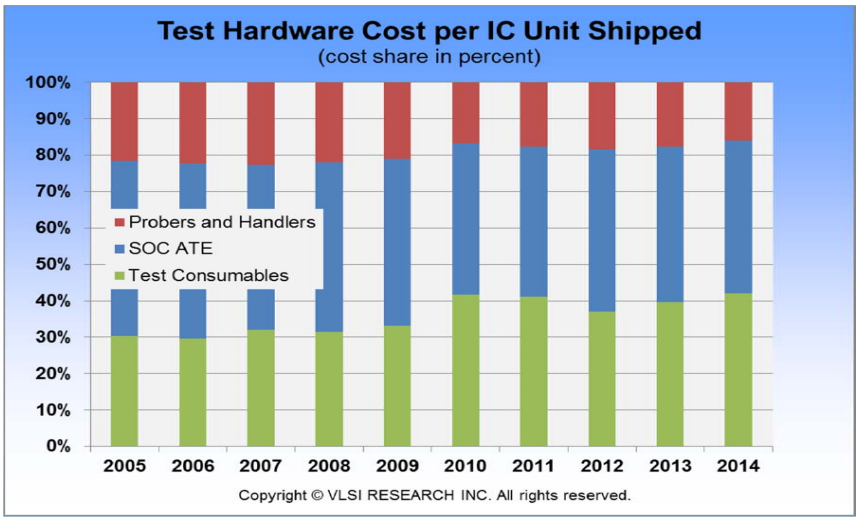


Figure TST6 – Distribution of costs in an average Test Cell

4.4 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%.

Typical site counts for various device types are shown in Spreadsheet Table TST2.

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**Error! Reference source not found.**. There are more efficient ways to reduce overall cost of test than going to the next setup with more sites in certain cases. Especially for high mix, low volume applications, there are many tester utilization challenges. In these setups, frequently, lower degrees of multi-site is preferable because test time improvement of techniques to improve utilization have a higher impact on the overall cost of test

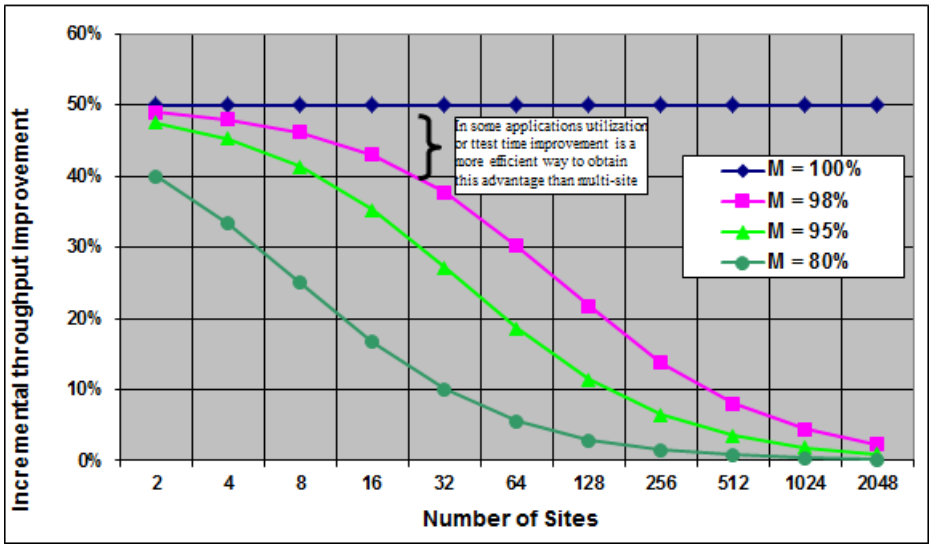


Figure TST6 – Importance of Multi-Site Efficiency in Massive Parallel Test

5 3-DIMENSIONAL DEVICE TESTING

3D/TSV is the next evolution beyond SiP. There has been significant work over the last two years from both academics (research) and industry (standards and working models/test chips) to identify and resolve challenges to testing 3D/TSV devices. In the medium to long term, as TSV-based die stacking becomes more prevalent and more complex/exotic die stacks appear, test challenges will also become more difficult. It is certain that new and additional Design-For-Test features will be needed to mitigate increased tester resource and time requirements, as well as increased test complexity, due to large numbers of different die in the same package. This section will address six key test challenges based on the evolution of 3D from SiP through complex die stacks: test flows, cost and resources; test access; heterogeneous die in a single stack/package; debug and diagnosis of failing stacks/die; DfX (Design for Test, Yield and Cost) and power. It is important to note that 3D/TSV is not yet a mainstream technology, and because of that, it is difficult at this time to make any predictions regarding 3D/TSV test flows. Currently there are two "adjacent" technologies: 2.5D and memory die stacks (Wide I/O, High Bandwidth Memory and Hybrid Memory Cube). Both of these technologies will yield insights to requirements and challenges associated with 3D/TSV. The best that can be gleaned from these technologies at this time is that reliance on BIST and boundary-scan based technologies, and use of fault tolerance with simple configurations tends to produce relatively high yields at the stack level. As these adjacent technologies become more mature and as more 3D/TSV applications emerge, more and better data will enable better predictions and decision making with respect to 3D/TSV test processes.

5.1 TEST FLOWS, COST, RESOURCES

At a high level, the 3D test flow is comprised of four test process steps: (1) pre-bond test – testing the individual die prior to integration into the stack (2) mid-bond test – testing the partially constructed die stack (3) post-bond test – testing the complete die stack assembly (4) final test – testing the packaged assembly. Mid-bond test and Post-bond test are new to the "traditional" test flow. Addition of these test steps will need to consider cost, process complexity and potential for damage with respect to quality of the 3D stack. New defect/fault models will be required to account for new process steps related to 3D, including wafer thinning and die stack assembly. Defect/fault models for TSVs will also be critical to help define TSV test requirements and processes. Mid-bond testing may not be considered based on cost and complexity.

Modeling of test flows for stacked die is already being developed by both academia and industry. Much of this modeling is related to optimizing overall test time; however, additional modeling to optimize resource utilization, cost and yield will also need to be considered. A "cost-weighted yield" model (i.e. which process/test steps have the highest impact on product cost) can incorporate both assembly and test flows to determine the optimal test flow with respect to the stacking process, as well as die level test and yield requirements. More work is needed in "Design for Stack Yield". Redundancy (for die, logic, memory and TSVs) is certainly more attainable in a 3D configuration. Such redundancy may be able to increase pre, mid and post-bond yields by "tolerating" a certain level of defects in either the die or the stack. Several proposals for TSV and die level redundancy and repair have already been published since the last update of the roadmap.

Test costs can be influenced by several factors: test resource requirements, overall test time, and cost-weighted yield. Mid and post-bond testing can add costs beyond the obvious test time increase. Most notably: failure of a single die will most likely compromise the entire stack; inter-die testing may add significant time and complexity; stacked die test resource requirements will be driven by the aggregation of all of the individual die requirements, increasing tester resource requirements and cost; huge test data volume will result from testing multiple die in parallel; data transfer/storage/security/integrity issues will be a consequence of information sharing between stack integrators and die providers.

For mid and post bond testing, the proposed IEEE P1687 (IJTAG) standard may simplify die to die functional test generation on ATE. In addition, boundary scan testing of die-to-die interconnects should be relatively straightforward. However, tools will need to be developed in order to simplify test program generation for partial and/or full die stacks.

Responsibility for some test process steps may also change in a die stack relationship. Reliability testing is one example where responsibility (die provider or stack integrator) is not yet clear. Pre-bond, Die-level, active burn-in (versus passive burn-in, i.e. simply baking) will be difficult given very limited access to the die. However, if burn-in is done post bond there is a risk of significant yield loss. At this point there is no clear direction on where and how burn-in is to be done (if done at all). A significant technology advancement will be required to make burn-in more practical at either the die or stack level, or to eliminate the requirement for burn-in altogether.

5.2 DFX

Design-for-Test/Debug/Yield (DfX) has been a prominent part of each of the previous sections. These embedded resources may enhance controllability, observability and defect/fault tolerance. DfX at the die level may include: a standardized access protocol to the die; built-in test features to enable die level (ATE) test capability for die in the stack (comprehensive logic BIST, memory BIST or "compressed and stored" ATPG vectors); interconnect test capability for all I/Os (boundary-scan and or at speed loopback testing); built-in debug and monitoring features to isolate defects in the stack (including potential capability to measure/monitor TSV continuity and performance); some level of fault tolerance/repair to enable higher yields in the stack (this may include the ability to "partition" given die in the stack). Built-in test and debug features mentioned above will become more prevalent for die in a stack. Partitioning logic or even the die itself can facilitate parallel testing at the die level and may be used to "decommission" logic on the die or a die on the stack (given some level of fault tolerance designed into the die).

DfX at the stack level will evolve over time as knowledge is acquired about defect opportunities during the die stacking process (i.e. back-grinding, wafer-thinning, and laser-drilling) and fault models are developed to accommodate those defects. Stack level integrators may be able to utilize interposers for stack level DfX features. Interposer based DfX features may assist with die to die testing and parallel testing of multiple die. Built-in interposer based test features may also help to reduce the requirement for tester resources driven by heterogeneous die on the stack. Note that early on, fault tolerance may be more important to die yields than DfX.

During the test/debug/yield-analysis process, it is imperative that tests may be conducted on one die alone; with multiple die simultaneously; and even with multiple die interacting with each other. Therefore, the access mechanism must provide two connectivity schemes on a per-die basis: 1) probe pads for bare-die test; and 2) interconnected and scalable TSV connections after stacking. These schemes must accommodate the access requirements mentioned in the Test Access section below.

5.3 TEST ACCESS

This section addresses two significant test challenges for 3D/TSV configurations: pre-bond access (probe access to the die) and mid/post-bond access - access to a single die in the stack. "Access" can be defined as: access to the test logic on the die, and/or access to the die I/O (especially the TSVs). The quality of the die stack is based on the premise of "known good die". That is, all die in the stack have been "fully" tested. Without full access to the logic on the die, test coverage is reduced and the possibility of defects escaping to the stack increases, risking quality and yield. In this case we would refer to the die as "not known bad" (which also infers that the die is "not known good"). The following paragraphs address test access challenges with respect to pre-bond, mid-bond and post-bond testing, including making conscious decisions to move away from "KGD".

The bottom die in a stack (where the "external I/Os" reside) will typically have probe-able pads for wire-bonding or flip-chip bumps. All other dies (middle and top dies) will typically be connected (pwr/gnd, clocks, control, data) through TSVs. A "typical" TSV configuration may be 5μm diameter at 10μm minimum pitch. However, in many cases, TSVs are not directly bonded, but equipped with micro-bumps. Micro-bumps may be 25μm diameter at 40μm pitch.

Current probing technology has not been demonstrated to reliably contact either TSVs or micro-bumps in a "production" environment. There are three issues that impact the ability to reliably contact the die: diameter and pitch of the vias/micro-bumps are too small for current probing technologies; current probing techniques may damage the via/micro-bump (any damage may be too much); potential for ESD damage to the logic on the die. The challenge will be to develop probe technology that can reliably probe on the micro-bumps (not only individually but in "arrays" of probe/contact points). This would require advances in: probe contacting, configuration of "contact arrays", metallurgies, tip cleaning recipes, minimizing damage from the probes. Contactless probing might play a role in this domain. This technique is being examined primarily by academics. Its main benefit is that it does not inflict probe damage. However, it still cannot probe the required sizes/pitches, and power/ground still needs to go through traditional needles. There are promising techniques that have been demonstrated, however, it may take 1 to 2 years for these techniques to be capable of supporting production level volumes.

Probing on 40μm pitch micro-bumps may be feasible in the short term, for limited array sizes. As long as the probe-technology is not up-to-par with the micro-bump/TSV sizes/pitches, we will need additional dedicated probe pads to enable sufficient probing and/or better Design-for-Test features that can access logic without access to the TSV. "Sacrificial pads" may help to mitigate some access issues but will not solve all issues. These pads also come at a price, so ROI needs to be considered. Unfortunately, as micro-bump technology scales down quickly, it will be very challenging for probe technology to keep up with access requirements.

A standardized, test access protocol (both physical and logical) is critical for die-in-stack testing. Test signals will need to be routed vertically (from die to die) and horizontally (within a single die). Without a standardized protocol, routing of test signals, vertically through the stack may be difficult due to test and functional signal density. Programming test features may also be confusing due to potentially different test protocols for individual die in the stack. Currently, test data routing through the stack may be defined by one of two scenarios: 1) the development organization drives the design of all die and all TSVs are physically lined up (test, verification, debug and other requirements are designed and implemented at the stack level as one design effort); and 2) off the shelf die will require either standardized "test access areas", or interposers to support re-routing of test signals to neighboring die in the stack.

Four basic access functions are required for stack-level test access: 1) the ability to provide access to on-die DFx features; 2) the ability to provide a bypass function for skipping-over a die; 3) the ability to provide a turn-around function for terminating the access function at the die; and 4) the ability to provide access to the next die above the current die. JEDEC has already defined some test/access capability in the Wide I/O specification for stackable mobile memories and the High Bandwidth Memory initiative. The Hybrid Memory Cube consortium is also working on a standard that includes some test access capabilities. The IEEE P1838 Working Group is also defining a standardized electrical access (and potentially physical) and test protocol for 3D die stacks. All of the test access mechanisms described above are based on the IEEE 1149.1 standard to some extent. The IEEE P1838 standard is primarily focused on test access while the other initiatives are only considering test access as a small part of the overall standard. For each of these standards, it is contingent on the Working Group to try to release the standard before the stack complexity increases to a point that test access to the die will be a significant challenge.

There will obviously be tester implications based both on the number and location of the test signals, and the protocol to "address" individual die in the stack and to access specific test features on the die. Ideally, the access mechanism should "assemble" itself as the die are stacked – even if the die come from different fabs and are made in different processes. An access mechanism should include a port on the base die, physical TSV definitions, and a communication protocol or control structure to talk to all of the per die DfX in a stack. The access mechanism must allow test of the die before they are stacked (pre-bond test) and must allow test after the die are in a complete or partial stack (mid/post-bond test).

5.4 HETEROGENEOUS DIE

Heterogeneous Die can have several meanings ranging from different functions (including memory, logic, analog and high speed optics/photonics) to different die providers. In any case, the evolution to complex, heterogeneous die stacks will have significant ramifications on test, test access, test application and test accountability. Some of these ramifications have already been discussed in the previous sections. However, the implications of testing die-to-die interactions go well beyond what was described earlier. A stacked die can be analogous to a Printed Circuit Assembly (PCA) from a test perspective. Testing of the 3D stack must account for potential die level, pre-bond test escapes based on: untested, functional die to die interactions; power and signal integrity in the stack (compared to die level testing); yet to be discovered defects/faults based on assembly and interconnect processes (wafer thinning may be a good example) which could expose and/or exacerbate die level defects. One area that will be improved over the PCA environment will be die-to-die latencies. Chip-to-chip latencies were a significant contributor to test escapes. Still, test escapes to the stack will be prevalent. Irrespective of the test time/cost perspective, generating a comprehensive, full stack, functional test can be anywhere between impractical to unperceivable. Depending on when testing occurs (mid-bond "stack and test" versus post-bond "assemble and test") it is possible that there may need to be several versions of functional tests to account for the different variations of die on the stack.

Similar to SiP, comprehensive testing can be accomplished by a combination of Built-in Self-Test (BIST), judicious use of existing and upcoming test standards, such as: IEEE 1149.1 - 2013, IEEE 1500, IEEE P1687, IEEE P1838 and limited functional testing. However, this will require a significant amount of coordination between die providers, stack integrators and the architect/designer. It is imperative that Built-in Self-Test be used extensively, both to test logic on the die and to test die-to-die interactions.

Traceability at the die level will be significant as the number of heterogeneous die on the stack increases. Data sharing between the die providers and the stack integrator will be critical to maintain quality levels of the stack as well as the individual die. Access to die IDs needs to be standardized in some way – either through a standard access protocol or through a standardized description language. In addition, data sharing and analysis tools will need to evolve in order to accommodate data driven process control that extends beyond the die provider to the stack integrator. It should be noted that test data storage requirements will increase dramatically in such a situation.

Ultimately, the testing of heterogeneous die will be a mixture of "component test", where a die is tested (in the stack) against fault models, parametric requirements and yield criteria to ensure the die is "known good"; and "board test" where a die is tested as a part of an integrated stack, covering interconnect, and interaction properties.

5.5 DEBUG-DIAGNOSIS

Similar to a PCA test environment, the major challenges with debug and diagnosis will be the ability to correlate failures at the stack level to defects/faults on the die. Functional tests at the stack level (mid-bond, post-bond and final test) will be very difficult to debug, especially when the failing die is in the middle of the stack (allowing for little to no "debug" access). The problem is compounded by the implication that it will be virtually impossible to remove a die from the stack without significant damage. Failure analysis of "systemic" defects will be costly, time consuming and ineffective unless adequate test/debug/FA resources are available at stack level test. Diagnosis may also be impaired by "environment factors" (thermal and power integrity) and an inability to identify potential TSV defects pre and post assembly. Note, it is imperative that stack level testing be able to discriminate die level defects/failures from failures due to the assembly process.

Significant integration of built-in test and debug features will be required at both the die level and the stack level. These features could include: built in logic analyzers/state capture, oscilloscopes, temperature and power monitors, droop detectors. Data from these debug features should be logged by the stack integrator and provided to the die provider along with the failing, stack level test. Moreover, significant advancements are also needed in Built-in Test and Debug technologies themselves. Areas such as Analog BIST and Functional BIST, as well as current logic and memory BIST require significant breakthroughs in order to facilitate test and debug capabilities in a limited access environment. Use of Built-in Test and Debug features may allow the die provider to recreate both the test environment and the failure at the stack level. Socketing/fixturing technologies will need to be significantly enhanced to allow for partial/full stack configuration on ATE. "Test interposers" may also be used to help identify/debug systemic, assembly-induced defects. Data-driven debug/diagnosis techniques may be another alternative over time.

5.6 POWER

Given that test power requirements can be greater than operational power requirements, power could be a significant test challenge. Power issues may occur at the power domain level, the die level or the stack level. Die and stack level power distribution requirements must be aware of test power consumption. Judicious power monitoring and droop detection throughout the stack will be imperative to guarantee test integrity. Power-aware testing at the die level and the stack level is something that will need to be considered and refined as test power requirements for the stack grow in the future.

High power levels may also contribute to thermal issues. Thermal issues could impact fixture design and performance. In addition, thermal variations may impact not only performance but also the integrity of the die stack itself. There will be a need to develop a "thermally induced, inter-die fault model" to describe the impact of thermal variations across the die and the stack. Thermal modeling and on-chip thermal monitoring may help to identify potential thermal issues in the stack. Validation of individual die and die in the stack, as well as pre-production testing should also consider guard-banding for potential thermal variation.

5.7 CONCLUSION

Significant evolution is required in test equipment, tools, EDA and methodology to address the challenges posed by 3D die integration. As is typical with test, the pace and scope of technology development in 3D die integration will drive test requirements and technology development. Challenges imposed by the manufacturing, assembly and packaging processes will also drive test technology and processes.

6 ADAPTIVE TESTING

Adaptive Test is a set of general principles applied to specific IC manufacturing test methods with the explicit purpose of optimizing the value of test. These principles include the practice of using IC device or system manufacturing test data to change a device or system test binning, future test sequencing or the manufacturing material flow so as to decrease the cost of test, improve performance distributions and/or improve outgoing quality and reliability.
The impact of Adaptive testing is a product value optimization for increased test effectiveness, runtime efficiency and creation of test flows for reconfigurable parts.
To effectively use Adaptive Test, new work is needed in test cell design and circuit design, data systems, and the coordination between IC manufacturers, IC test and assembly. The top challenges facing industry application of Adaptive Testing is listed in Section H.
This whitepaper provides:

1) A description of Adaptive Test and terminology used by its practitioners;
2) Example applications of Adaptive Test as of 2015 and future opportunities and;
3) A list of challenges for the development and deployment of Adaptive Test.

To ensure the industry can fully exploit the benefits of Adaptive Test, this whitepaper describes:

1) The infrastructure requirements for test cells, data systems and device and system designs and;
2) A description of Adaptive Test challenges and the coordination needed between IC manufacturers, OSATs (Outsource Assembly and Test providers) and fabless system integrators.

6.1 ADAPTIVE TEST

The intent of Adaptive Test is to increase the quality and reliability of products (and potentially yield) and/or to improve operation efficiency of testing. Also, Adaptive Testing can help to identify abnormal parts as early in the manufacturing sequence as possible (preferably at wafer test) and to add tests or change test conditions to screen the riskier material. Additional Adaptive Test methods may selectively skip test content on specific material to save costs. In doing so, Adaptive Test must successfully account for the risks of passing bad parts.

6.1.1 ADAPTIVE TEST DEFINITION

Adaptive test comprises a set of methods for automatically changing manufacturing test conditions, manufacturing flow, test content, test limits, or test outcome to reduce test cost, increase outgoing quality and reliability, reconfigure parts, or collect data to further improve test and manufacturing. Adaptive test makes these changes in a manner that does not significantly increase testing time or increase human involvement in test operations. The decisions on when and how to adapt the test are made algorithmically by the tester, other test cell equipment, or an automatic data analysis system in a shorter time than the traditional test improvement cycle involving engineering analysis.

6.1.2 ADAPTIVE TEST DESCRIPTION

Adaptive Test is generally accepted as an advanced test strategy that can be used to achieve quality, yield, and cost goals that might not be reached by normal test methods. Adaptive Test may modify a production test process in any of five ways:

1) **Test Conditions** *(modifying voltage or clock frequency such as VDD)*
2) **Manufacturing Flows** *(adding or deleting test insertions such as burn-in)*
3) **Test Content** *(adding or deleting specific patterns or tests such as transition fault or IDDQ, respectively)*
4) **Test Limits** *(changing the pass/fail limits such as DC power or Vdd-min test specifications)*
5) **Test Outcomes** *(changing the binning of some die based on post-test analysis of the die's test results)*

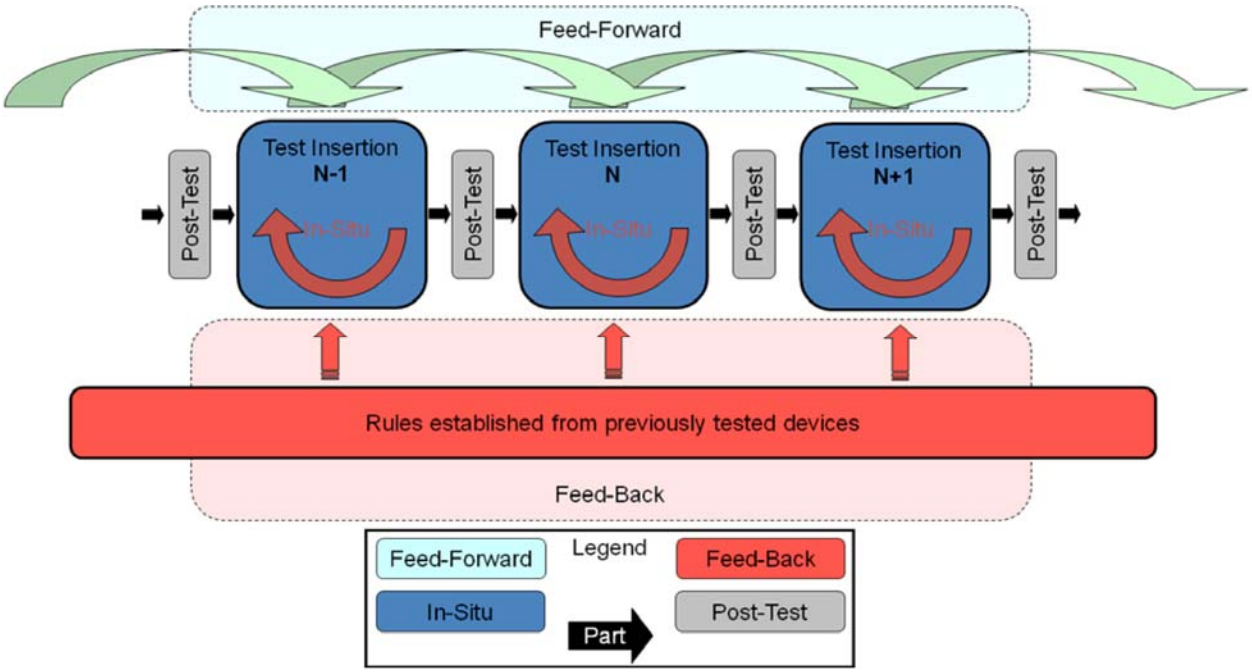


Figure TST 7 - Adaptive Test supports feed-forward and feed-back data flows. Adaptive Test provides for data use for test decisions either in-situ (i.e. during a given test step) or post-test.

6) Adaptive Test applications are organized by when decisions are made to modify the test flow and to which device(s) the modified test flow are applied. The four most common categories are in-situ, feed-forward, feed-back and post-test.

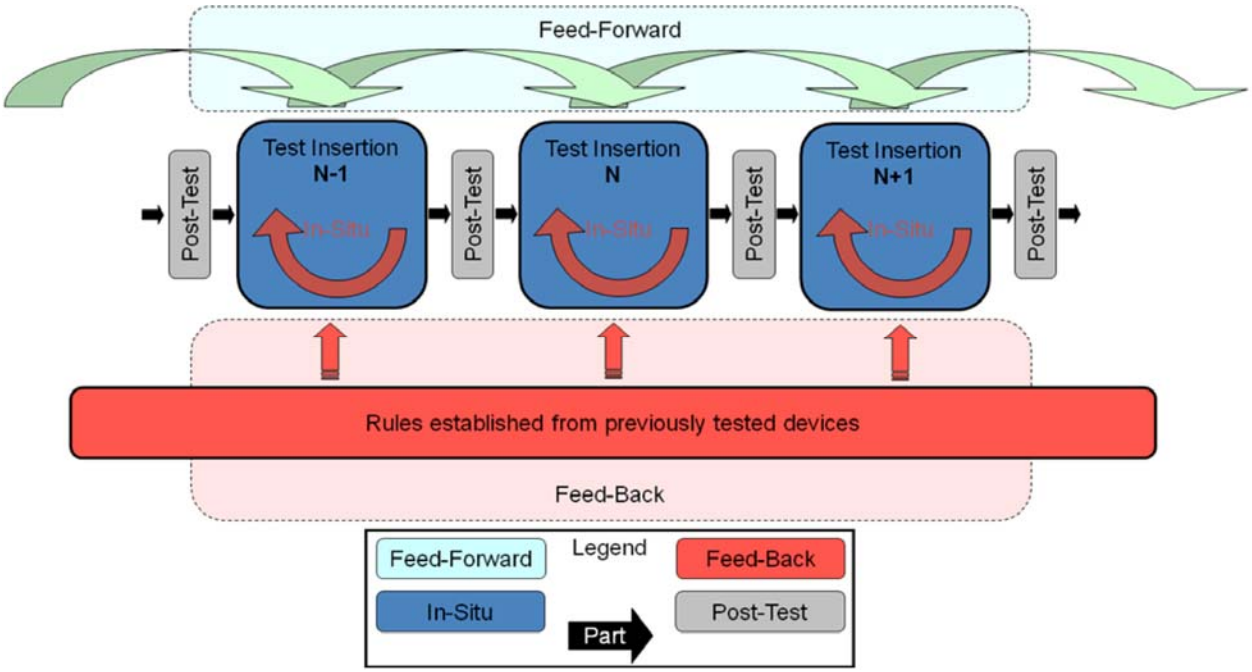


Figure TST 7 - Adaptive Test supports feed-forward and feed-back data flows. Adaptive Test provides for data use for test decisions either in-situ (i.e. during a given test step) or post-test.

is a flow diagram depicting the relationships among these four categories.

1) **In-situ:** *Data collected from the part being tested is used to modify the testing of the same device during the same test insertion. Speed grading is an example of the In-situ category where the data from the device is used to change the conditions of the test plan for the same device. Another example common to analog components would be trim and device calibration.*

2) **Feed-forward:** *Data collected from a previous test step stage (e.g. probe, hot probe, burn-in) is used to change how the same parts are tested at a future stage. An example of the Feed-forward category are statistical methods which identify 'risky' dice or wafers and selects these components (only) for burn-in or "clean" dice that may be candidates for reduced testing.*

3) **Feed-back:** *Data collected from a previous part (or parts) is used to modify the tests or limits of different devices yet to be tested. Skipping some test patterns on high yield wafers, adding more tests to low yield wafers or refining statistical models used for die classification are examples of this category.*

4) **Post-Test:** *Data sample statistics or other analysis is performed between test steps and is used to reclassify certain devices or to change future manufacturing flow and test conditions for these devices. Part Average Testing and outlier identification methods are examples of the Post-Test category.*

6.1.3 EXAMPLE APPLICATIONS

Below is a list of example Adaptive Test applications. Each example is labeled by one or two categories outlined earlier. In addition to clarifying the categories, the examples demonstrate the shift from manual and static methods to automatic methods with little or no human intervention during test execution. Note that the use of electronic die/chip ID (e.g., a die-specific identifier such as wafer/XY coordinate and lot information that is fused on each die) is a key enabler for many of these applications.
There is a list of references that include many example applications at the end of this section.

1) **Dynamic test flow changes (In-situ, Feed-forward):** Die production data is monitored within the test program to add or remove tests, to selectively perform per die characterization for yield learning, or to collect data for later diagnosis. This application supports many common real-time Statistical Process Control methods.

2) **Statistical screening (Post-Test, Feed-forward):** After wafer or lot data collection, identify die which are outliers or mavericks as possible sources of test escapes spikes or reliability failures. Statistical screening is Feed-forward because results can be used to route target dies through test flows different from the main flow.
   - *PAT – Part Average Testing is a statistical technique relating the test result of a device under test to the test result of the remaining dice on the wafer.*
   - *NNR – Nearest Neighbor Residuals is a statistical technique relating a univariate or multivariate test result to a model derived from a local region of the device under test.*

3) **Single-step flow control (Feed-forward):** Data from one test step is used to optimize testing at the next test step to focus subsequent screening on issues observed in the manufactured parts.
   - *For example, inline test modifies wafer test; wafer test modifies package test; burn-in modifies final test; or package test modifies card/system-test.*

4) **Off-tester optimization of test flows (Feed-back):** Off-tester data analysis drives test flow changes for future devices (fully automated).
   - *For example, off-line analysis could optimize test flows, test content and test measurement routines using input from many sources including historical data, test capacity, required turn-around times, DPM (defects per million) requirements, expected yields and parametric data.*

5) **Production monitors and alerts (In-situ, Feed-forward, Feed-back):** Data from multiple sources is merged for statistical analysis to control production test optimization beyond what has historically been possible.
   - *For example, subtle parametric shifts from marginal wafer probe contacting can be automatically identified and action taken during production testing.*

6) **Die matching (Feed-forward, Post-test):** Production data from various sources is used to support the build/test process for multi-chip applications and many of today's board build process to match specific die combinations during assembly.

- *Note die-matching data transfer may require world-wide data sharing, across multiple companies and throughout the entire supply chain.*

7) **On-chip test structures and sensors (In-Situ, Feed-forward, Feed-back):** Data collected from auxiliary on-chip test structures such as ring oscillators, selected critical paths, on-chip voltage and thermal sensors, or on-chip reliability monitors is used to modify the die's test content, test limits or future test flow.
   - *Sensor measurements can be used at all levels of assembly (including system operation) to monitor and adjust functionality.*

8) **On-chip configuration (In-situ, Feed-forward):** Production test data (including test structure data) is used to adjustment features in the design to improve die's performance, power, margin, yield or reliability.
   - *Emerging ICs have more on-chip configuration & adaptability such as clock tuning, partial goods (redundant spare cores), and voltage and frequency adjustments (including per core).*

9) **Card/System configuration and test (Feed-forward, Post-test):** Component test results (such as parametric data, yield characteristics or partial good data) are used to customize the card/system test flow or customize card/system test conditions.
   - *Architectures such as IEEE 11491-2013 standardize the infrastructure that can be used to enable Adaptive Testing applications at the board and system level.*
   - *The feed-forward application enables specific fabrication process and component test parameters (e.g. Vdd-min or wafer x,y location) to be fed-forward and used by the board test program to make decisions on whether to add specific content to test for marginality.*
   - *The feedback and post-test applications enables the creation of a pareto of the board-level failures per Electronic Chip ID of a specified component type is sent to the supplier, enabling analysis to correlate failure types to fab and test parameters. If so, then the supplier can adjust their tests or bins.*
   - *In-situ test where the reading of an on-chip sensor for voltage or temperature enables more or less stressful board conditions to be applied to check for margin and performance. On-chip sensors can also be read during field usage to monitor aging and data can be sent back to the suppliers to adjust their test limits.*

10) **Adaptive Diagnostics (In-situ, Feed-forward):** Test results drive advanced diagnostic data collection.
   - *For example, on-chip BIST (built-in self-test) circuitry can be programmed on-the-fly to localize and characterize specific types of the failures. But these methods must only be selectively applied to ensure reasonable test time/cost.*
   - *Many emerging chips have programmable on-chip test controllers that can interpret test results on-chip – and take action (test, diagnostics, characterization) without requiring extensive data collection being transmitted to/from the test equipment.*

6.1.4 ADAPTIVE TEST ARCHITECTURE / FLOWS

Figure TST 8 - The architecture of Adaptive Test organizes each insertion's test data into one or more databases. A waterfall of manufactured parts may insert, join or query databases for test flow decision-making.

displays a model of the entire End-to-End flow of parts under test and Adaptive Test applications. Note that there are feed-forward, in-situ, feed-back and post-test dispositioning opportunities at each test step. Although

Figure TST 8 - The architecture of Adaptive Test organizes each insertion's test data into one or more databases. A waterfall of manufactured parts may insert, join or query databases for test flow decision-making.

shows a simple view of the database, the actual database structure would probably consistent of 2-3 databases levels each with unique capacity and latency capabilities.

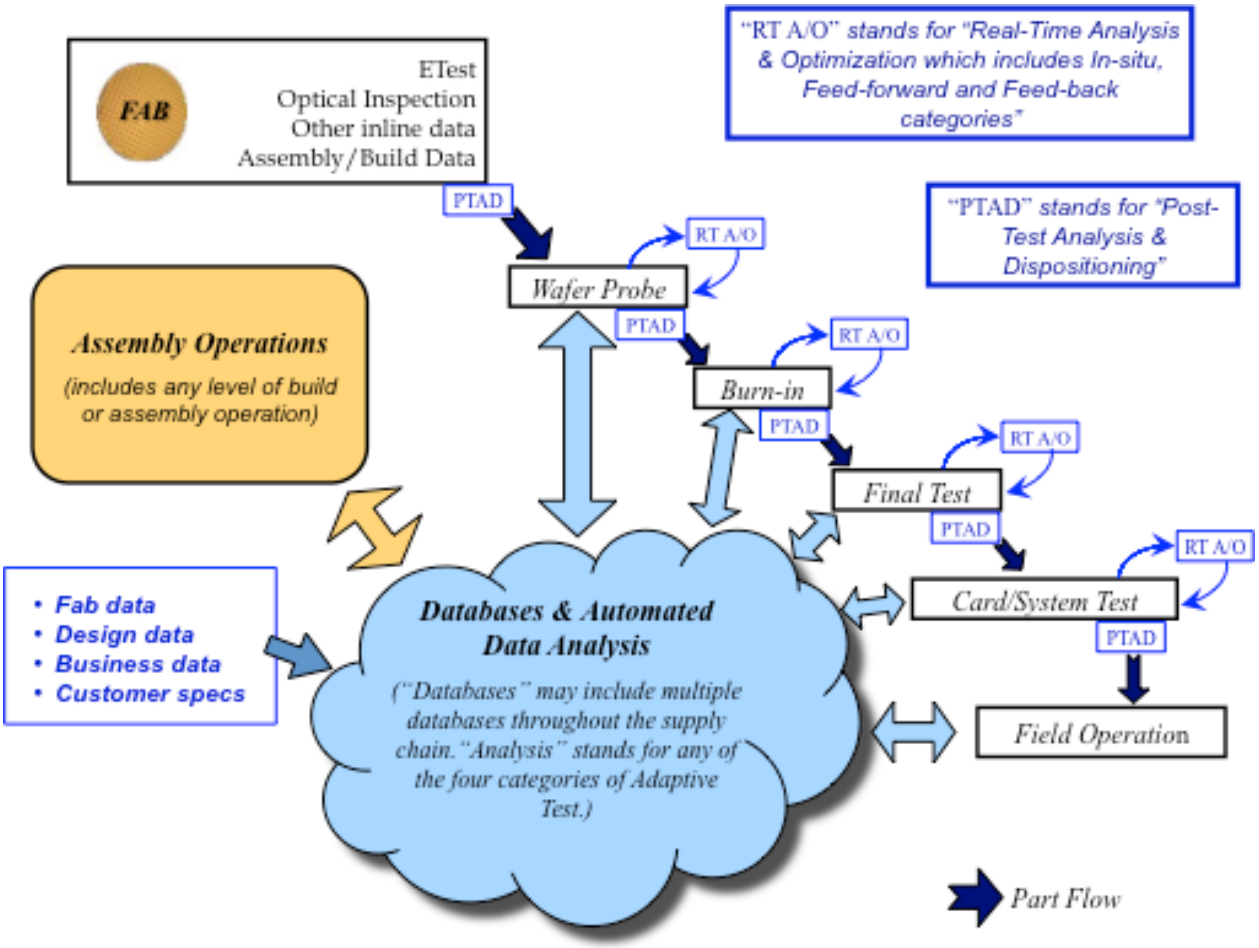


Figure TST 8 - The architecture of Adaptive Test organizes each insertion's test data into one or more databases. A waterfall of manufactured parts may insert, join or query databases for test flow decision-making.

6.1.5 LEVELS OF ADAPTATION & DATA MODEL

Making the decisions to adapt any of the test attributes listed above first involves collecting the right data (which test programs already do well) and then organizing the data into a structured data model so that the right data can be accessed when and where it is needed. At the appropriate time, data of the proper scope, that is data from a particular test run or data from a particular part, wafer, or lot, is accessed from the data model and processed by the applicable decision

algorithms. Similarly, the test variables, such as limits, conditions, flow, content, must be changed at the right time to complete the adaptation decision.
The data model can exist entirely in an off-line database apart from the tester, or be distributed between servers and the tester depending on the latency requirements and convenience. To branch a test flow for a particular part (a real-time decision) latency must be short, i.e. there can be no significant impact to test time. To support low latency requirements, the data needs to either be stored on the tester or be rapidly pulled into the tester. To make an outlier decision such as to re-bin some already tested parts, longer latencies are tolerated such as from the time of test until the time material is shipped. Longer latencies mean an off-line database can be used.
Decisions to adapt a test are often based on comparing the variation observed on the sample of parts in question to a model of the expected variation. In outlier detection, parametric test limits are adapted to track expected variation so as to only discard parts with unexpected variation. Tests can be temporarily dropped from a test flow when their control charts show the manufactured material and test process is in control and within specification. If and when monitoring based on sample testing shows the material or test process has changed and gone out of control, the tests are reinstated on every part. Similarly, diagnostic tests can be added to a test flow when certain types of failures occur more frequently than expected.
Generally, more adaptability means more frequent decision-making in the test flow with the goal of improving the trade-off between defect level from "freeing the guilty" and overkill from "convicting the innocent" or balancing test false negative error rates and test false positive error rates, respectively. Adaptability follows a bottom-up progression from the conventional static limit, to a static parameter variance model (static PAT), to a variance model with variable parameters (dynamic PAT), to choosing variance model equations based upon well-grounded principles. Moving up this progression requires not only more data but also a better understanding of the processes that cause the test response to vary. This progression also means the decision-making generally moves from an off-line human activity to an on-line machine activity.

6.2 ADAPTIVE TEST INFRASTRUCTURE (DATA EXCHANGE, DATABASES, ETC.)

Adaptive Test makes decisions throughout the manufacturing process based upon data from multiple sources and using data of varying detail and completeness. Before actionable decisions can be made with multiply sourced data, new data integration requirements are needed. Some integration requirements are unique to Adaptive Test and are different from data requirements used at any of the originating sources.

Figure TST 8 - The architecture of Adaptive Test organizes each insertion's test data into one or more databases. A waterfall of manufactured parts may insert, join or query databases for test flow decision-making.

highlights Adaptive Test data requirements reach across all test insertions. Changes or new Adaptive Test data processing requirements are anticipated throughout the Silicon manufacturing and assembly. As of today, example databases are coming on-line that merge and centralize Adaptive Test infrastructure.
Data requirements that are different but not unique to Adaptive Test include date and time stamping, test naming, and data recording methods. For example, Adaptive Test data stamps should be consistent across all insertions and between companies. Current date stamping practices are ad-hoc with some companies using different date formats and date references at different test insertions. Database standards exist for date stamping such as Coordinated Universal Time. A time and date stamp requirement policy eases integrating test data when some units are retested and simplifies merging two (or more) data sets in an unambiguous time order. Similar issues arise in recording floating point results of the same test from different insertions with different precisions and formats.
Data requirements unique to Adaptive Test center on the database polices of latency, access and retention period. Latency measures the time between the request for a data item and the availability of the requested item. Access refers to the scope of the user community that can store, retrieve, and act on a data item. Retention period measures the time the data item is electronically available.

- Local processing in the test cell requires low latency. *For example access latency should be in a few milliseconds, if data is to be retrieved on a per device level (Real-Time Analysis & Optimization (RT*

A/O). Post-Test Analysis & Dispositioning (PTAD) applications may have latency requirements of a few seconds to a few minutes. Normally data volumes for these steps would be relatively small.

- Processing in a central database (e.g., "The Cloud") has more relaxed timing constraints (minutes, hours), but typically deal with much larger data volumes
- Since Adaptive Test decisions affect the quality of shipped goods, data retention requirements depend on specific market requirement (which may exceed 10 years in some cases).

Many areas of the IC manufacturing are increasingly more comfortable with using data from the cloud, but a notable exception is the test cell. Test cell integration of Adaptive Test algorithms is one of the most challenging applications. For example, local test cell actions (such as "*clean probe-card now"*) were the sole responsibility of the specific test floor and were designed to guarantee the test cell integrity and test cell-to-test cell correlation. Adaptive Test changes this paradigm in a number of ways:

- Algorithms will be owned by multiple involved parties, including wafer fab, design house and test floor. *Some algorithms may originate from commercial providers, others from the involved parties themselves. They all need be executed smoothly next to each other in a real-time environment.*
- Data collection as well as data access (e.g., to upstream data in case of data feed-forward) becomes a mission-critical task of a test operation as well as the entire supply chain. *This challenges the reliability of the underlying infrastructure, which likely spans multiple companies, geographic areas and cultures.*
- Likely, one wants to simulate the impact of algorithms on historical databases to understand how to maximize the value of Adaptive Test without creating adverse side effects. *This requires the exact same algorithm to be executed in as diverse environments as a cloud database and a test cell measuring real-time data.*

As a consequence, industry needs to develop

- **APIs** to allow algorithms to plug into a diverse set of environments.
- **Data exchange formats** which are flexible, compact and standardized so that only minimal extraction and translation effort is required. A common set of indices is required, such that data remains identifiable and traceable even across heterogeneous supply chains.
- **Recipe management systems** which can handle a diverse set of recipe origins, check for (likely un-intended) interactions and maintain consistency across non-synchronized update cycles from the various origins. Version control systems for these recipes are also required.
- **Execution systems** must be enabled to monitor the health of Adaptive Test algorithms (are basic assumptions met?) and escalate errors to the right entities in an actionable format.

6.3 IMPLICATIONS FOR ATE AND THE TEST CELL

The test cell is expected to deliver a cost-effective means to screen defects for quality, classify devices for performance and collect data for learning. The rate of product complexity is increasing with more clock domains, voltage planes, IOs and configurable fuses followed by the introduction of parallel testing of multiple, dissimilar devices with 2.5D and 3D stack packaging. In parallel, the business demand for higher quality and reduced product cost severely challenges the ability of the test cell to continue to provide an effective test solution and still continue to reduce the cost of test. Adaptive Test methods provide levers to address these additional demands but not without disruptions of their own.

Adaptive test requires the test cell to be able to accept input from external and internal sources, apply device-specific models to determine test conditions and to evaluate results for flow control and device binning. This materially changes the setup, execution and categorization requirements of the test cell and affects both low-level software capability such as firmware as well as high-level software such as the executable test program. Of particular challenge is the relationship of flow control and binning when the test flow becomes a function of non-deterministic evaluations influenced by dynamic test points, limits and device configuration.

Future test cells must support the following:

- *Per-device test flows based on external inputs, the device itself, and dynamic business rules*
- *Zero-cost logging of data for use in the Adaptive Test process in a database-tlike format*
- *A move from being the entire cell controller to a test engine with a standard API*
- *Asynchronous and distributed (multi-insertion) test flows*
- *The ability to perform structural (model-based) testing over functional*

6.4 TEST RESULTS DRIVING "ADAPTIVE DESIGNS"

More and more designs are being reconfigured during testing. Examples include partial goods (on-chip redundancy), VDD/frequency adjustment and local clock tuning. In most cases this product personalization will be based on either test measurements or data feed-forward from other operations. In some cases, this reconfiguration will be based on "application demand".

6.4.1 TESTING RESILIENT FEATURES

Resilient features are on-chip structures that can be used to configure the product to work around hard defects or to tolerate latent defects. These structures span a wide range of circuits and architectures, including fuses, redundant memory elements, and architectures capable of operating on reduced numbers of logic elements like CPU cores or graphics components, error-detection and -retry schemes for hard and soft errors, and the sensing and control circuitry for adaptive designs. Like every other circuit, these structures must be themselves tested and characterized, though these circuits present unique testing challenges beyond standard logic and memory elements, including temporary configurations (for fuses), soft-repair vs. hard-repair validation (for memories), combinatorial explosion of down-cored variants of redundant features, the need for error injection to test recovery circuits, and analog stimulus for sensors (such as voltage or aging monitors).

6.4.2 NON-DETERMINISTIC DEVICE BEHAVIOR: TEST AND RUN-TIME AVAILABILITY

Non-determinism is incompatible with traditional cycle-accurate automated test equipment, but is nonetheless becoming typical on modern SOCs. Several new I/O protocols are non-deterministic, as are the standard methods to avoid metastability in clock-domain crossings (which are commonplace in highly integrated devices). Fine-grained power gating and power-state management can change the configuration of a device and its execution profile during test and normal operation. Adaptive designs take this notion even further with architectural features which can perform state rollback and pipeline retry based on events at arbitrary times during execution. The result is that test patterns, particularly functional patterns which execute in mission mode, must either prevent or be tolerant of non-deterministic response. The former raises coverage questions; the latter pattern and ATE interface challenges.

6.4.3 TESTING ADAPTIVE DESIGNS

Adaptive designs bring the complexity of dealing with advanced power management such as power gating, variable configuration of IP (such as IO and arrays), self-defining performance bucketing and part-specific reconfiguration (such as redundancy, repair and harvesting) to a test environment traditionally characterized by a linear test flow measuring to fixed corners to verify device operability. Instead, on-chip sensors are used to detect the workload, voltage, temperature, and timing margin of the chip as it operates and dynamically adjusts power supplies, clock frequencies, thermal control, and even the instruction stream. The adaptive features of a design make it much harder to define (and thus characterize) both typical and worst-case use models, which in turn makes it more difficult to test appropriately. Additionally, the removal of excess margin represented by traditional guard-banding increases the risk of exposure to subtle defects, necessitating both higher coverage and better correlation between structural and functional test modes.
An emerging direction is to apply Adaptive Test techniques (which modify the parameters or content of a test program based on information collected about the device under test from the current or previous test insertions) to adaptive designs (which modify their own operating point or execution flows based on internally generated feedback). The proclivity of an adaptive design to compensate for the environment in which it is being (functionally) tested will present challenges for data gathering by the Adaptive Test process beyond opening control loops to test at fixed conditions. A means to record and store the conditions to which the device is tested, organized in a manner for ease of retrieval and consumption is required.

6.5 ADAPTIVE MANUFACTURING

An emerging direction is using test results to drive other IC production steps such as packaging multi-chip products. For example, the card/board assembly operation may require that specific dies or types of dies to be used on specific boards based on previous test results. Given the emergence of multi-chip packages (such as 3DICs) and power constraints,

specific bare dies will need to be selected for assembly based on parametric data collected at test such as IDD or power/performance measurements.
Key challenges of End-to-End data feed-forward for assembly operations include:

- *Cross-company data management*
- *Robust data availability*
- *Data security*
- *Full traceability*
- *Data format standardization*

6.6 ADAPTIVE TEST FOR CARD/SYSTEM/FIELD

Adaptive Test methodologies described in this document for IC-level testing can be equally extended and applied to board and system testing and even field usage. While ICs have traditionally been tested standalone in an almost 'noise-free' ATE environment and/or tested with limited structural tests to see whether they perform to their specifications, the board/system environment can be quite different in terms of noise, timing margin, voltage and functional test trigger conditions that structural tests were unable to produce. Improved board yield and IC DPM can be improved significantly where adaptive test that includes the board/system level performance is able drive enhanced screening both at the IC suppliers test and/or board/system manufacturing test.
The four types of Adaptive Test described in Section 2 (In-situ, Feed-forward, Feed-back and Post-test) can all be extended to include the board and system manufacturing.
One of difficulties to extend the chip-level adaptive test to board/system or even in-field test is to track their test trigger conditions and be able to convert between them. For example, chip-level scan-based logic gate test may not be always applicable for board/system/in-field tests due to the difficulties or impossibilities to control the scan chain data, clock pulse, non-stoppable in-field online function executions, etc. Similarly, a functional execution, which can be treated as a functional test may be hard to convert to a chip-level ATE test because the function execution could involve memory contents, their transactions, logic and I/O data flow, etc. Therefore, tracking the test/failure conditions and the capability to convert between them is the key for adaptive test extension to board/system level.
Extending Adaptive Test applications to the board and system level requires extensive data infrastructure, analysis, exchange and security. Companies providing ICs, board design and test need to openly collaborate on a technical and business level to be successful.

6.7 ADAPTIVE TEST CHALLENGES AND DIRECTIONS

This section highlights the key challenges that the industry must address to fully exploit Adaptive Testing across the supply chain.
The color scheme of the table below is:

- *White --Manufacturing solutions exist*
- *Yellow -- Some solutions may be known but not widely accepted or mature.*
- *Red -- Manufacturable solutions are not known -- or all solutions are not industry standard. (i.e., known solutions are based on proprietary solutions)*

Figure TST10 – Adaptive Test Challenges

Challenge	Status	Needs
Recipe *(Input variables, data treatments, output variables/ responses)*	Many good outlier recipes exist. Spatial interpolation over a wafer is becoming popular for sample testing Opportunity to branch test flow for only fault coverage required by current defect rate	Guidance on best measurements to make
		Guidance on which recipes to apply
		Fault or defect coverage metrics
		Higher level of adaptability where best variance model is automatically discovered instead of chosen beforehand

Decision Rule *(Define actions resulting from recipe output)*	Actions defined for gross outliers; loosely defined for less extreme events such as downgrade or reconfigure or escalation/ de-escalation of decisions (such as test sampling)	Where to set outlier thresholds
		How to combine the results of multiple outlier definitions (e.g., develop a metric for Quality or Reliability be developed)
		Criteria for rejecting vs. downgrading or reconfiguring
Infrastructure *(Foundation to enable execution of recipes and decision rules)*	Part traceability enables feed-forward, feed-back but robust environment for data transport, storage and providence is lacking (no commercial solutions currently exist)..	Standard data formats amenable to adaptive test
		Move from working with files to working with databases
		Ability to feed data and decisions both forward (with the parts to future stages) and backward (future parts at given stage) in test & assembly manufacturing flow
		Full traceability of adaptive test parameters for each part: limits, content, flows, decision rules, model parameters
		Full part configuration as tested (e.g., redundant cores, partial goods, on-chip tuning, multiple die such as 2.5D/3D in a package)
		Real-time communication among test cell machines and data storage and analysis engines
Evaluation *(Execution of test cases to prove viability and benefit)*	Receiver-operator curve concept understood by most practitioners but standard methods for experimental definition and ROI interpretation do not exist.	Clear evaluation criteria to build trust in adaptive test methods
		"Gold standard" against which to compare outliers (including all variations of adaptive test flows & settings)
		Good metrics for continuous monitoring of recipe effectiveness.
		Quantification of cost of shipping a bad part
Deployment *(Implementation and release into production use)*	Ad-hoc, company specific implementations utilized but no commercial solution or clearly defined requirements for a turn-key solution are defined.	Commercial adaptive test platform into which methods can be plugged and recipes specified
		Connections to Manufacturing Execution Systems & Product Data Management systems
		Complete visibility across supply chain: fab, test, assembly both internal and external
		Supply chain data integration and processes which automatically detect supply chain issues and implement corrective actions in near real-time

6.8 SUMMARY

Adaptive Testing has the opportunity to improve product quality & reliability, reduce cost and improve product yield beyond today's capabilities. Almost all companies are starting to use some forms of Adaptive Testing, but there is not a sequential roadmap for implementation and many applications are created in an ad-hoc way.

There are a number of challenges that are today limiting the industry's ability to fully exploit Adaptive Testing across the supply chain. (These are highlighted in the table in the previous section.) The industry must collectively address these challenges in the next few years.

7 TEST TECHNOLOGY REQUIREMENTS

7.1 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 TST9 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.

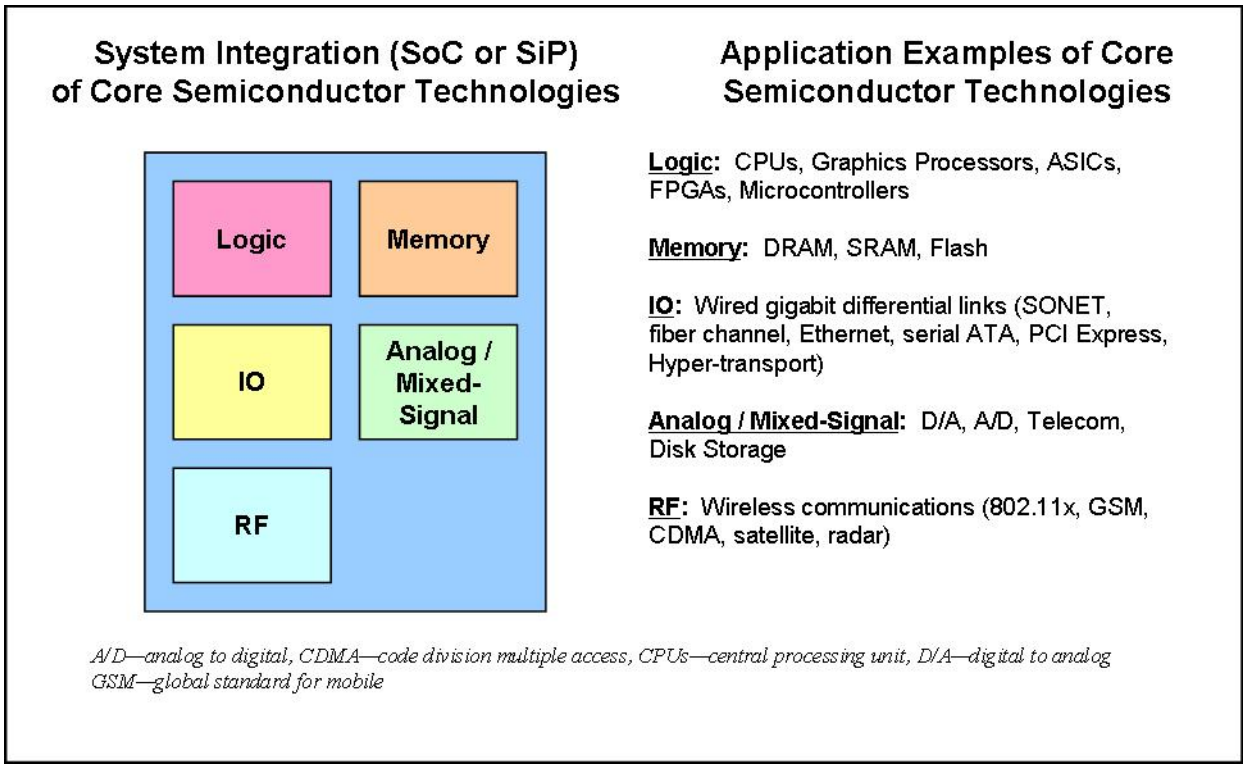


Figure TST9 - Organization of Cores for System Integration and Applications

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.

7.2 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.

7.2.1 SYSTEM ON A CHIP

A SoC design mainly consists of multiple IP cores, each of which is an individual design block and its design, its embedded test solution, and its interface to other IP cores are encapsulated in a design database. There are various types of IP cores (logic, memory, analog, high speed IO interfaces, RF, etc.) using different technologies. This assortment requires a diversity of solutions to test dies of the specific technologies corresponding to these embedded cores. Thus 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 that includes targeting the interfaces between the cores and the top-level glue logic (i.e. a logic not placed within a core) in addition to what is within each core instance. Effective hierarchical or parallel approaches and scan pattern compression techniques will be required to evaluate and adjust the overall quality and cost of the SoC to an acceptable level for customers.

On the other hand, the SoC test technology improvement to handle a progression of design technologies accelerated by the evolving applications is indispensable. The technologies and the requirements of the DFT design (*the design intent*) are addressed in the Design Chapter. The well-organized roadmap and the potential solutions that reflect these *design intents* should be reviewed by the readers. For example, low power design methodologies, which improve the chip performance, are widely adopted in various current SoCs. However, it is not easy to test the SoC without deeply understanding its functional behaviors and physical structures. As a result, the conventional DFT that focuses only on the static logic structure is not enough anymore, and the evolution to tackle this issue is strongly required.

The quantitative trends and requirements of a consumer logic chip are shown in Spreadsheet Table TST4, which is given later in Logic section, compared with a MPU chip. Table TST4 introduces the guideline for DFT design and the requirements for EDA tools.

Spreadsheet Table TST4 – DFT Requirements

7.2.1.1 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, compression logic, core wrappers and test points, which can be kept constant over time by using a hierarchical design approach

Both SoC and MPU devices have an increasing amount of digital logic on the devices. Table TST7, given later, shows a common view of the DFT techniques which are expected to be used moving forward in an effort to cover the most likely faults (as modeled by the EDA systems) while attempting to keep the test costs low by effectively managing the test data volume.

There are four basic approaches in use for scan test generation:

1. The EDA tools can consider the circuit in its entirety and generate what's called a "flat" test without leveraging the hierarchal design elements nor including pattern compression techniques. Virtually no one does this anymore, but it is useful for comparison purposes with the more appropriate approaches briefly described below.
2. The EDA tools can consider the hierarchal design elements to achieve an on-die parallel test setup. Parallel test would be applied to instances of wrapped cores to enable multiple instances to be tested in parallel.

3. The EDA tools can imbed compression and decompression circuitry around the scan chains, allowing for many times more chains internally without increase of ATE scan pin resources, resulting in less data being required to be stored on the ATE for stimulus or output comparison purposes.
4. The EDA tools can implement a combination of 2 and 3 for a compressed hierarchal approach. This would involve cores being wrapped for isolation and including compression within the cores. Further compression may be utilized by testing multiple instances with the same set of scan-in pins, considered scan pin sharing, to allow testing of multiple instances of cores in parallel. The test data/scan outputs from each core instance may be observed independently or further compressed together and sent to a common set of chip scan-out pins, possibly resulting in more chip scan pin sharing.

The approach used to apply tests to embedded cores will have a large impact to test time and perhaps also test data volume. One approach that has been used in the past is to test a core in isolation and route its stimulus and expected responses up to the SoC pins so as to avoid having to do ATPG for the core at the SoC level. This saves CPU time for running ATPG, but fails to help reduce test time for the SoC. A more effective approach that can be applied when using test compression is to test multiple cores in parallel and not put them into complete isolation from other cores. Thus, while test compression may be used inside of cores, it may also be used above the cores to allow the scan stimulus to be sent to multiple cores in parallel and to compact the output from several cores before sending it off chip.

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 reduced over the roadmap. Therefore various approaches, such as test pattern reduction, scan chain length reduction and scalable speed-up of scan shift frequency, should be investigated. However the acceleration of scan shift speed may increase the power consumption during scan shift cycles and so it may make the test power problem more serious. Some DFT and ATPG approaches to reduce the power consumption during scan shift cycles are required. There is also the important issue of excessive power consumption during the capture cycle. Several approaches to relax this issue have been proposed, but most of them cause an increase of test pattern counts and consequently make its impact on test application time intolerable. Some low capture power test approaches to minimize the increase of test pattern counts are also required. The impact on test data volume from these low-power scan sequences is shown with a 20% test data volume premium in the low-power rows. This will be too optimistic for cases where very low (e.g. less than 15%) switching is required since that could easily result in a doubling of the pattern count for the same coverage.

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, high test data compression ratios will be required in the 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 simultaneous test of repeatedly used IP cores in a design that can share a common set of chip scan pins. By broadcasting the same scan-in stimulus to all such core instances, we reduce the bandwidth of data being sent from the ATE onto the chip and need less storage for that data on the ATE. Observing the outputs from each instance independently can aid in diagnosing failures, but by compressing the core instance outputs together and observing them at a common set of chip pins further increases the effective compression that may be within each core.

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 or restrictions on test power may prevent maximum simultaneous test of identical IP cores. The impact of this effect may be investigated for future editions of the roadmap.

The issue of power consumption during test mentioned above is one cause for the increase of test patterns which will increase test data volume. Therefore requirements on test data reduction also take account of this issue.

Year of Production	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Worst Case (Flat) Data Volume (Gb)																
MPU-HP (Big Data) – High performance MPU (MPU, Servers, GPU)	2699	3673	4998	6802	9256	11366	13957	17139	21047	25845	31737	38973	47858	58769	72166	88616
MPU-CP (Mobile) – Mobile consumer MPU (SOC, APU)	1639	2230	3035	4130	5620	6901	8475	10407	12779	15693	19272	23664	29060	35685	43819	53808
SOC-CP - (Low power Devices) -- Low power consumer (MCU, IoT, MEMs, Analog, IoT)	847	1137	1526	2049	2751	3697	4970	6685	9000	12124	16341	22041	29743	40157	54216	73198
Best-Case Test Data Volume (Hierarchal & Compression) (Gb)																
MPU-HP (Big Data) – High performance MPU (MPU, Servers, GPU)	5.7	6.4	7.2	8.1	9.1	9.2	9.2	9.3	9.7	9.8	9.9	10.0	10.4	10.5	10.8	11.0
MPU-CP (Mobile) – Mobile consumer MPU (SOC, APU)	4.8	5.2	5.9	6.6	7.4	7.5	7.5	7.6	7.7	7.8	7.9	8.0	8.1	8.2	8.5	8.6
SOC-CP - (Low power Devices) -- Low power consumer (MCU, IoT, MEMs, Analog, IoT)	5.5	5.7	6.7	8.0	9.4	9.8	11.6	11.7	13.9	16.5	19.6	20.7	24.6	29.2	28.8	34.3
Best-Case Compression Factor (Hierarchal & Compression)																
MPU-HP (Big Data) – High performance MPU (MPU, Servers, GPU)	471	572	694	842	1022	1242	1509	1835	2171	2628	3201	3903	4595	5611	6652	8088
MPU-CP (Mobile) – Mobile consumer MPU (SOC, APU)	342	425	516	625	758	926	1123	1363	1655	2009	2425	2948	3586	4368	5171	6261
SOC-CP - (Low power Devices) -- Low power consumer (MCU, IoT, MEMs, Analog, IoT)	153	200	227	258	292	377	428	570	648	735	835	1065	1209	1373	1881	2136

Figure TST10 – DFT Compression Factors (Flat with No Compression = 1)

Error! Reference source not found. shows the impact of hierarchy and compression DFT techniques on the problem of 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 factors of more than 100X compression are often achieved. However, even a 500x compression won't be enough as shown in Spreadsheet Table TST5 for SoC, therefore, more sophisticated technologies will be required in the future. **Error! Reference source not found.** shows the level of compression anticipated. 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.

Note: the SOC-CP data volume and compression factors do not show smoothly changing values due to the sharing of scan pins done to reduce data volume, but the small number of scan pins available for these devices means that as the percentage of scan pins shared goes up, the number of cores that can be tested in parallel goes up due to rounding more like a step function. When more cores can be tested in parallel, you get lower data volume and better compression factors and with only a few scan pins available, not all pin sharing percentages changes result in a change to the number of cores able to tested in parallel – thus some years see less improvement than other years.

In order to map this anticipated test data volume to tester and test time requirements one must take into account the number of externally available scan chains and the data rate used to clock the test data into and out of the device. Estimates for these important parameters are shown in the SOC and MPU sections of Spreadsheet Table TST4, which is given later. Since these parameters may vary on a part by part basis, the resulting data will need to be adjusted based on the approach taken on one part versus another:

- Designing more scan chains into a device results in more parallel test efficiency and a proportionally faster test time and less memory per pin in the test system. This assumes the scan chain lengths are proportionately reduced.
- Clocking the scan chains at a faster speed also results in a faster test time but doesn't reduce the pattern memory requirements of the ATE.

The other question when looking at the ATE memory requirements is which pattern compression technique is chosen for use on a given device. This question is impacted by many parameters including device size, personal preference and time to market constraints. As such, the analysis in Table TST8 shows the *minimum* patterns per pin necessary to test the most complex devices. Thanks to the usage of more elaborate pattern generation techniques the data suggests that the minimum pattern requirement will only grow by 2x to 3x over the roadmap period.

The test time needed to drive and receive this much test data volume is impacted by the data speed in use. As cost effective higher speed tester capabilities are deployed we feel that these data speeds will be used to speed the tests and reduce the test time per device. The analysis calculates this impact and suggests that test times will be dropping over time due to these faster scan shifting speeds. 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.

Concurrent parallel test in the core hierarchy has the potential of reducing test time. ATPG/DFT level reduction technologies should be developed in the 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 is some research regarding this 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 the 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. It is also that the leak current of test circuit itself should be also considered as a part of power consumption.

The discussion so far in this section has focused on the automatically generated scan based testing requirements. Functional test techniques continue to be broadly deployed in order to enhance the scan based testing techniques in an attempt to confirm the device's suitability for the desired end-use application. Additionally, more and more memory arrays are getting embedded inside of both MPU and SOC devices.

7.2.1.2 REQUIREMENTS FOR EMBEDDED MEMORY CORES

As process technology advances, and due to some special application needs, both the number of memory instances and the total capacity of memory bits increases and will cause an increase in area investment for BIST, repair and diagnostic circuitry for memories. As the density and operating frequency of memory cores grow, memory DFT technologies as follow are implemented on SOCs and be factors of area investment increase:

- To cover new types of defects that appear in the advanced process technologies, dedicated optimal algorithms must be applied for a given memory design and defect set. In some cases, a highly programmable BIST that enables flexible composition of the testing algorithms is adopted.
- Practical embedded repair technologies, such as built-in redundancy allocation (BIRA) which analyzes the BIST results and allocate redundancy elements, and built-in self-repair (BISR) which performs the actual reconfiguration (hard-repairing) on-chip, are implemented for yield improvement.
- On-line acquisition of failure information is essential for yield learning. A built-in self-diagnostic (BISD) technology distinguishes failure types such as bit, row, and column failures or combinations of them on-chip without dumping a large quantity of test results and pass the results to ATE to utilize them for the yield learning. The testing algorithm programmability mentioned above has to be more sophisticated to contribute the diagnostics resolution enhancement. It must have a flexible capability to combine algorithm and test data/condition, and a memory diagnostic-only test pattern generation capability which is not used in the volume production testing.
- All the above features need to be implemented in a compact size, and operate at the system frequency.

The embedded memory test, repair and diagnostic logic size was estimated to be up to 35 K gates per million bits in 2013. 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 and a low-speed portion might reduce the area investment and turn-around-time for timing closure work. 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. 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.

7.2.1.3 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, IEEE1500[1]); and this format must be preserved and extended to other core types, such as analog cores. The DFT-ATE interface is going to be standardized (for example, IEEE1450[2]), and it should include not only test vectors but also parametric factors. Automatic design and test development environment is required to construct SOC level test logic structure and generate tester patterns from the test design information and test data of each IP core. This environment should realize concurrent testing described below.

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 in the 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

[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).*

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.

7.2.1.4 CONCURRENT TESTING

For SoC test time reduction, concurrent testing, which performs tests of plural IP (non-identical) cores concurrently, is a promising technology. For instance, the long test time of high speed IO can be mitigated if others tests could be performed at the same time, which would decrease the total test time drastically. To realize the concurrent testing concept, there are items that must be carefully considered in the product design process. These items include the number of test pins, power consumption during test, and restrictions of the test process. These items are classified as either DFT or ATE required features in Figures TST13 and TST14. IP cores should have a concurrent test capability that reduces the number of test pins (Reduced Pin Count Test: RPCT) without a test time increase and a DFT methodology which enables concurrent testing for various types of cores. As these requirements differ corresponding to the core types on a chip, a standardized integration method of RF, MEMS and optical devices into a single SoC with conventional CMOS devices can be developed. It includes unification and standardization of test specification which are used as interfaces by IP vendors, designers, DFT engineers and ATE engineers that can be combined with breakthrough analog-mixed signal / RF DFT methodologies. (e.g. Integrated efficient interfaces for test of core itself and core test access, or wide adoption of IEEE Std 1500, and its extension to analog etc.)

DFT and ATE must cooperatively consider concurrent testing requirements and restrictions. This may not be an easy task as there are multiple challenges to enable concurrent testing. For instance, ATE software needs to be able to perform concurrent test scheduling after analyzing the power and noise expected during testing based upon design and test information specified for each IP core and chip architecture by the designer.

Figure TST13 – Required Concurrent Testing DFT Features

Features	Contents
External test pin sharing	Each JTAG enabled IP core must use the 5 JTAG interface (TRST, TMS, TCK, TDI, TDO). Cores that have non-JTAG interfaces must be able to share external test pins with other cores.
Design for concurrent testing	The test structure of an IP core must be operationally independent from that of all other IP cores
Identification of concurrent test restrictions	The presence of any test restrictions for each IP core must be identified to the scheduler. (e.g. Some IP cores are not testable at the same due to noise, measurement precision, etc.).
Dynamic test configuration	Test structures/engines that can change the order of test and the combination of the simultaneous test to each IP core
Test Data Volume	The test data volume of all IP cores must be able to be stored in the tester memory
Test scheduling	Critical information on each IP must be available to the test scheduler a) Test time of each IP core. b) Peak current and average power consumption for each IP core c) Test Frequency for each IP core
Common core interface	The test access interface of IP cores must be common among all IP cores (e.g. IJTAG)
Defective IP identification	There must be a mechanism to identify defective IP cores prior to and during test

Figure TST14 – Required Concurrent Testing ATE Features

Features	Contents
Numerous Tester Channels with Frequency Flexibility	A large Number of Test channels that cover a wide range of frequencies will enable efficient concurrent testing. Test channels must provide test data such as clocks, resets, data, or control signals to many corresponding IP blocks. Testing can be more flexible if channels assignments are dynamically changeable
Mixed Data type support	Capability of loading / unloading test data that is a mixed combination of digital, analog, and high speed I/O data is required.
IP block measurement accuracy	Measuring accuracy of testing (e.g. high-speed I/O test) should be preserved in concurrent testing to match the specifications.
Test Data Handling Efficiency	Test data loadable to each divided test channel should closely match memory usage efficiency as that of non-concurrent test.
Power supply capability	A Large number of capable power supplies pins will enable large number of IP blocks to be simultaneously tested
Multi-Site Testing capability	Capability to perform both multi-site testing and IP-level concurrent testing at a time will enable efficient testing
Capable Software	Automated test scheduling software that can decide test scheduling configurations while considering many constraints is required.

Multi-site testing is another approach to reduce effective test time per a die or a chip. Effect of cost reduction in each approach depends mainly on the number of test pins and production volume as shown in Figure TST15 – Comparison between Multisite and Concurrent. Larger number of test pins will make the number of multi-sites smaller, and higher production volume will get larger profit by the cost reduction. To estimate an accurate profit, cost of jigs and expense on engineering and designing should be also considered.

Figure TST15 – Comparison between Multisite and Concurrent

Pin Count	Production Volume	Efficiency	
		Multisite Testing	**Concurrent Testing**
Many	Large	Medium	High
	Small	Low	High
Few	Large	High	Medium
	Small	Low	Low
Consideration		Cost of Jig (Initial Cost) - Probe Card, Test Board etc. Cost of Tester - Pin Count, Power Supply etc.	Reduction of Test Pins (RPCT) Cost of Chip - Impact on area etc. Cost of Design

7.2.1.5 DFT FOR LOW-POWER DESIGN COMPONENTS

Low-power design is indispensable for battery-powered devices to enhance system performance and reliability. The design includes multiple power domains which are independently controlled by PMU (Power Management Unit), and some special cells used for controlling power at physical level, such as level shifters, isolators, power switches, state retention registers, and low power SRAMs.

However, the design raises new requirements in testing, which are some dedicated functions in test. For example, an isolator should be tested in both active and inactive mode to fully test its functionality and a state retention cell requires a specific sequence of control signals to check if it satisfies a specification. Please refer to the Figure TST16 -- Low-power Cell Test for more low-power cells. Some of the defects on the special low power cells may be accidentally detected in an ordinary test flow but it is usually not enough to ensure entire low power features of a design. These functions have not been treated in the historical scan test that only focuses on the structure of circuits. Therefore a full support of these dedicated test functions for special low power cells is strongly required.

Figure TST16 – Low-Power Cell Test

#	Component	Test Contents
1	Isolator	Generate patterns controlling power-on/off of the power domain
2	Level Shifter	Include the cell faults in the ATPG fault list
3	Retention F/F	Generate patterns to confirm saved data after RESTORE operation
4	LP SRAM	Generate patterns which activates peripheral circuit inside the macro during the sleep mode and confirm the cell data retention
5	Power Switch	Generate patterns to measure IDDQ with domains power on/off

7.2.2 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.

7.2.2.1 STACKED DIE TESTING AND EQUIPMENT CHALLENGES

Stacked Die (SiP and TSV) products can present many unique challenges to backend manufacturing flows because these products can 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

Stacked Die 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 Stacked Die situation, 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.

7.2.2.2 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.

7.2.2.3 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.

7.2.2.4 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.

7.2.2.5 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.

7.2.2.6 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.

7.3 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.

Spreadsheet Table TST4 – Logic Assumptions

Spreadsheet Table TST5 – Logic Test Data Volume

Spreadsheet Table TST6 – Logic ATE Requirements

7.3.1 HIGH VOLUME MICROPROCESSOR TRENDS DRIVERS

The trends in Spreadsheet Table TST4 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 Spreadsheet Spreadsheet Table TST4 show the trends of functions per chip (number of transistors) and chip size at production. Chip size in terms of area 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 - the diminishing returns of clock frequency increases. There is a trend to

greatly increase the number of cores with each process generation, and these will include multiple cores of a particular instruction set, but also include other types of cores, such as graphics units (GPUs), specialized I/O units (e.g. USB) and various other cores not necessarily specific to that microprocessor.

7.3.2 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 (including 3-D chip stacking) 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 fully optimize 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. Refer to the High Speed Input/Output Interface section for more detailed discussion on the test requirements and challenges for testing these interfaces.

One manifestation of increasing data bandwidth will likely be the use of 3-D stacked memory with high-density / low-power through-silicon via (TSV) interfaces. This presents a number of test challenges, including the economic imperative to identify Known Good Die (KGD) prior to stacking, the inability to physically probe the TSVs, the associated loss of access to pins which are often re-purposed as scan inputs/outputs, and the increasing test cost associated with post-bond testing of a stacked system.

Because of the huge amount of transistors able to be placed on current and future chips, producing a new chip design in a reasonable amount of time practically requires the use of an IP/SoC design methodology based on embedded cores. Such a core-based or IP/SoC design methodology allows some chip components to be designed by different groups or companies and then integrated into a single SoC. Using multiple instances of the same core design helps reduce the designer's time and effort to apply the available transistors for meaningful benefit in the final product. Test can also exploit the use of cores by applying a hierarchical test methodology that isolates cores from surrounding logic using, for example, the IEEE 1500 standard for Embedded Core Testing. Much of the logic test analysis is based on the assumption of the use of hierarchical and core-based testing for both microprocessor and SoC devices now and going forward. Specifically, the calculations for test time and test data volume assume that all embedded cores are wrapped and tested with a two-pass approach: first, the individual cores are tested with the wrappers in "inward-facing" (or INTEST) mode; then the connections and glue logic between the cores is tested with the wrappers in "outward-facing" (or EXTEST) mode.

7.3.3 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:

- When multiple identical cores are designed into a device the trend for these cores to be tested in parallel while sharing the same scan data stream will continue. We assumed here that there is an ongoing requirement for

visibility into any faults on a per-core basis that will require individual scan output data access to all cores – even identical copies of the same logic.

- 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; in many cases it is expected that core designs will include such self-test and/or test data compression functions to be utilized when isolated from their surrounding logic during core INTEST modes of die testing. The use of scan compression on a per-core basis is assumed in test cost projections.
- 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 (such as internal loopback BIST and eye diagram mapping circuits) 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.

7.3.4 ATE TRENDS DRIVERS

Automated Test Equipment (ATE) must continue to provide the necessary operating environment and interfaces to meet the test requirements of future SOCs. Drivers include both device parameters like I/O speeds, chip power and thermal environment, and test length, as well as test floor requirements like multi-site testing and tester footprint and cost.

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 speeds and protocols of the chip interfaces. Requirements for cycle-accurate tester determinism may be relaxed with the advent of protocol-aware testers. Support for a wide variety of different I/O types, including on the same device, will be necessary through at least the near term horizon

There is a trend toward low-power design to serve the mobile client and dense server markets, and even the traditional high-performance microprocessor designs appear to be peaking at a power consumption during test of 400 W through the end of the roadmap (though it is likely that there will always be a segment of the microprocessor market which pushes the power envelope). The low-power design trends include device states with specified power envelopes, necessitating accurate power measurement instrumentation which can be applied during various test conditions.

Test equipment vector memory requirements are expected to grow modestly over time, but are at risk if on-chip scan compression ratios hit a plateau. The following section explains the context which drives the test data volume calculations.

7.3.5 LOGIC TEST ASSUMPTIONS AND CONTEXT

The scan test context for an individual core is shown in Figure TST117.

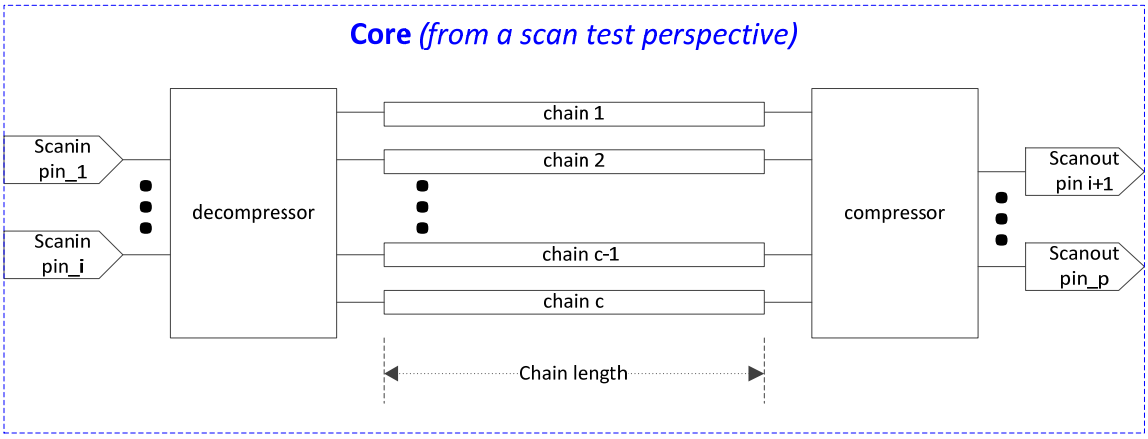


Figure TST11 - Scan test view of a core

Each scan-testable core is assumed to include scan compression hardware (a decompressor and a compressor, referred to collectively as a "codec") which expands a small number of scan input channels into a large number of scan chains (each

with a uniform length), then compacts those chains into a small number of scan output channels. The core includes scan input pins and scan output pins for these scan channels. If there are pipeline flops added within the codecs to support a desired scan shift frequency, these are simply added as overhead to the scan chain length.

An SOC consists of a number of cores, as shown in Figure TST8.

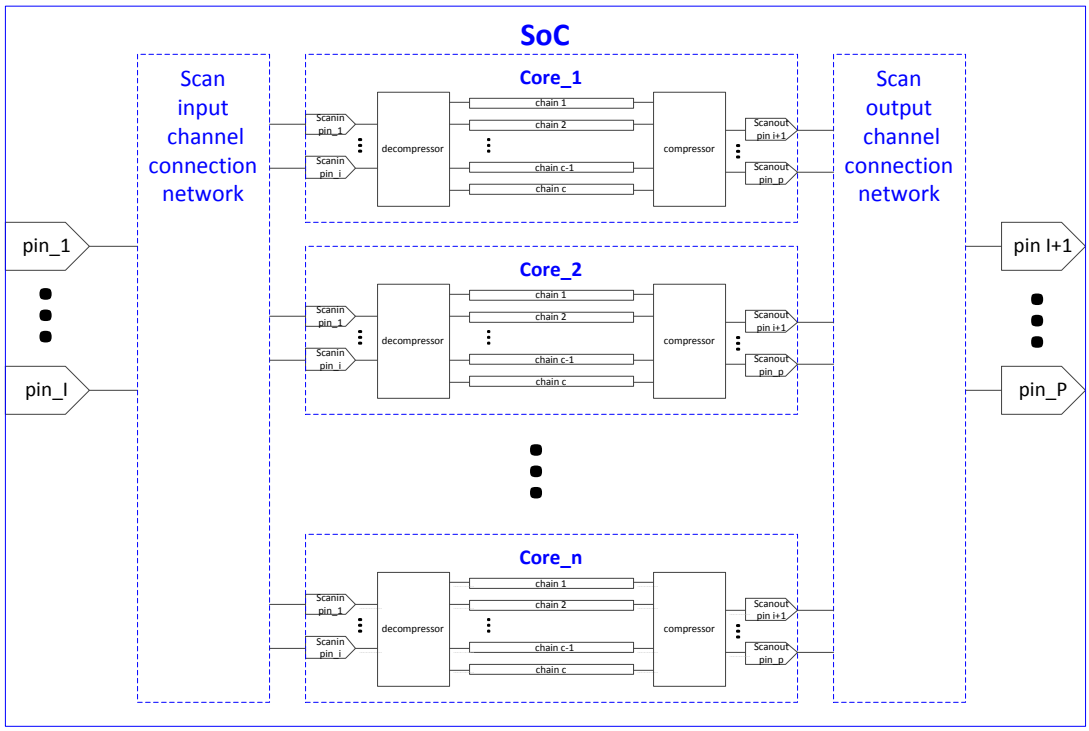


Figure TST18- SOC containing many scan-testable cores

The SOC consists of a number of scan-testable cores. In order to provide the stimulus to the cores and observe responses from them, the chip pins must be connected to the core scan-in and scan-out ports. There are a number of techniques used for this purpose, including the use of dedicated test pins, the re-use of functional chip pins in test mode, or combinations thereof. Figure TST8 shows networks between the pins and the cores (on both input and output sides) which are responsible for mapping chip pins to core scan ports. If there are enough chip pins to connect to all the core scan ports, then these networks are trivial (i.e. wires). However, it is often the case that there are insufficient chip pins, so the networks may implement schemes which share, multiplex, sequence, combine, or otherwise map the pins to the cores. The approach chosen has an impact on the test application time and test data volume. However, for the purposes of keeping the projections independent of this network architecture, the calculations have simply modeled the situation by assuming that the number of chip pins is sufficient to connect to all the scan ports of all the cores. Compensation for the size of the chips is done by allowing the length of the scan chains to increase. While this isn't physically accurate – cores come with chains of a pre-determined length – it does result in the correct calculation of test data volume without nearly as much complexity in the model.

One important variant of this architecture is shown in Figure TST129, where there are multiple identical copies of Core_2.

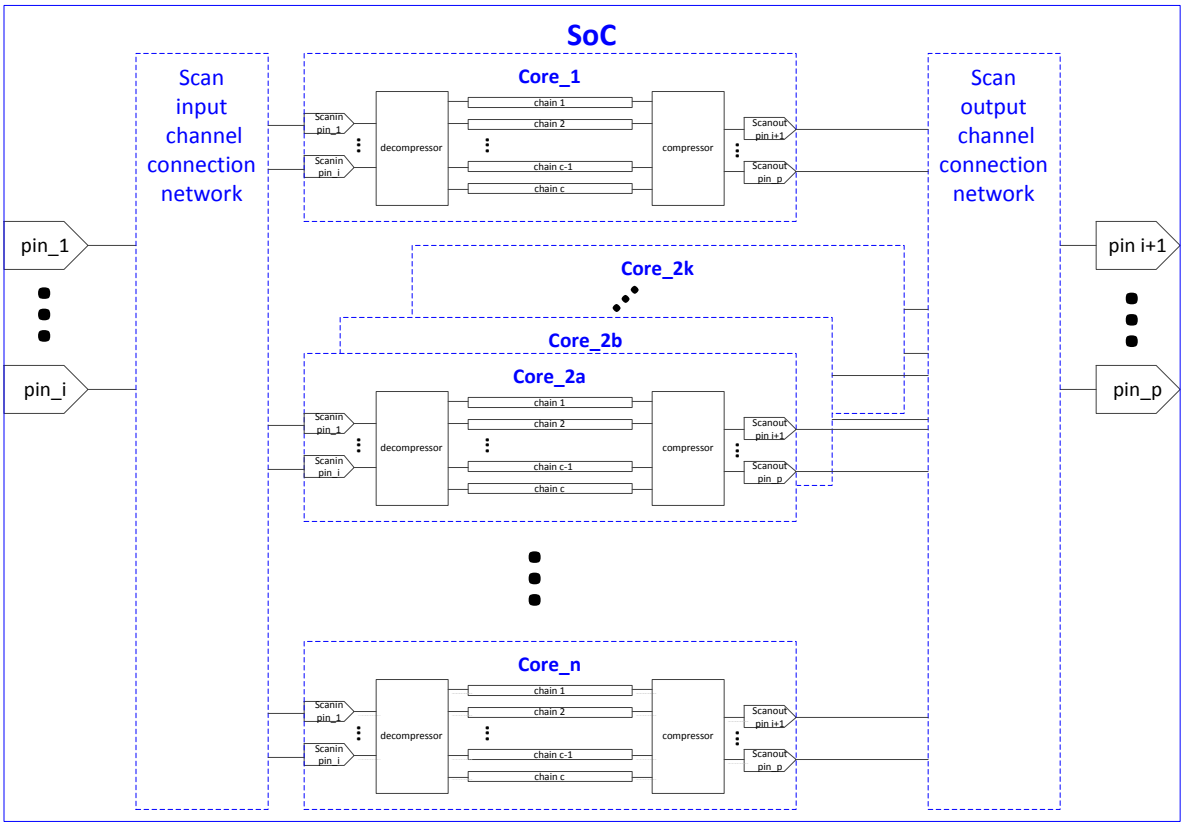


Figure TST12 - SOC with multiple copies of Core_2

Multiple identical cores offer an opportunity for significant test cost savings since the copies may all be tested in parallel. There are a number of methods to accomplish this, the most common of which is illustrated Figure TST129: the inputs to all identical copies of the core are broadcast, while the outputs from each core are uniquely observed. This can lead to the need for a large number of chip output pins to perform all tests at once. If the number of SoC pins is less than the number of core pins, the tests for the cores must be broken into a number of sessions. However, as noted above, this is handled in the calculations via the trick of simply pretending that the pins are there but that the scan chains are longer.

The final concept to set the context for the calculations is the separation of core logic from random logic, as illustrated in Figure TST13.

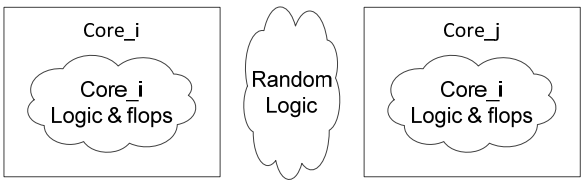


Figure TST13- Core logic vs. random logic between cores

In addition to the core-level scan tests (sometimes called "inward-facing" or "intest" modes), there are also tests for the random (or "glue") logic between cores. These tests utilize the wrapper flops inside the cores in their "outward-facing" or "extest" mode to exercise the random logic. The calculations separate core logic from random logic.

7.3.6 LOGIC TEST CALCULATIONS

The assumptions used for test data volume calculations are shown in Spreadsheet Spreadsheet Table TST4.

The test data volume projections are shown in Spreadsheet Spreadsheet Table TST5.

New this year we have published the equations behind the various logic table calculations in Spreadsheet Table TST22.

Graphs of some of the key conclusions from this work are shown in Spreadsheet Table TST21.

One of the key results of the projections is shown in the graph of test data volume in Figure TST2121.

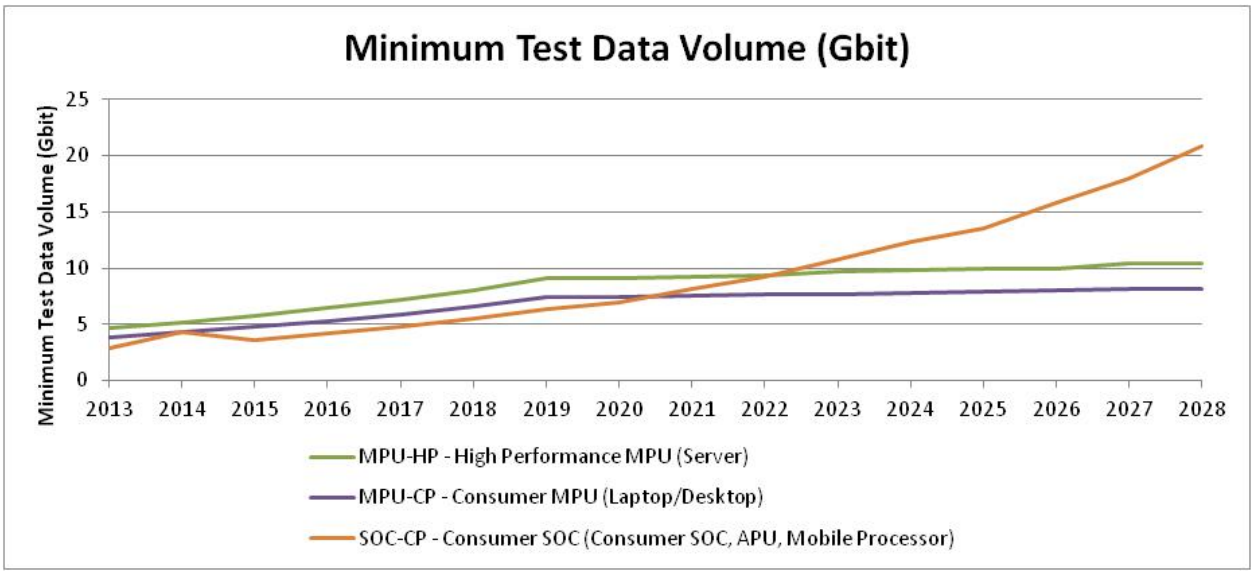


Figure TST21 - Test data volume trend over time

Rather than the historically steady increase in test data volume, the coming trend towards the use of many copies of identical cores (and the associated opportunity for concurrent testing of those cores) will significantly flatten out the growth over time. This trend is more apparent in highly homogeneous server MPU devices and less impactful on SOC-CP devices which tend to be more heterogeneous.

The ATE needs projections in terms of pattern depth and test-times are shown in Spreadsheet Spreadsheet Table TST6.

7.4 HIGH SPEED INPUT / OUTPUT INTERFACE

High frequency multiple Gbps/GHz I/O technology continues to show significant growth in speed and port count in computing, networking, and consumer applications, beyond its true serial communication origin. It is now a very typical IP block found in microprocessors, peripheral interface chips, and inside peripheral devices. Gbps interface standards such as PCIe, QPI, HMC, Thunderbolt, GDDR, DisplayPort, DDR, USB, Infiniband, SATA, SAS, Fiber channel, Gigabit Ethernet, XAUI, SONET, OTU, and OIF/CEI 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 a very different cost structure than the high-end low-volume devices and therefore have 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 demands different test considerations. Because of the above noted reasons, the technology trends are illustrated as a scattered plot in **Error! Reference source not found.**.

The booming telecomm industry of 1998~2001 drove the standard serial communication technology to 13Gbps, and towards 40Gbps using more exotic processes. Multi-Gbps/GHz interface speed experienced an exponential increase (as shown in **Error! Reference source not found.** 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, used in high speed interfaces, the adoption was far from simple. For example, most of the Gbps/GHz 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/mW or pico Joule/bit, which is becoming a major combined performance metric 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, Gbps/GHz 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 Gbps/GHz plateau and it will require a serious undertaking in technology to go beyond the somewhat established 13 Gbps range. As of 2009~2011, shrinking process technology is starting to support growth on data rates beyond 13 Gbps; such as 14 Gbps FC and 16 Gbps PCIe. As of 2015, emerging technology reached beyond 56Gbps with leading CMOS technologies. Unlike the long haul telecomm applications, some of the Gbps/GHz 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 100G/400G Ethernet IEEE 802.3ba standard). And the increase level of SOC integration also moved some of the demanding bandwidth needed from chip-to-chip into the same IC, such as the CPU and GPU interface. Future data rate progression may be slower than that seen in the booming telecomm industry around 2000 but still requires doubling every few years to absorb the performance scaling with better power efficiency in terms of mW/Gbps. However, this slower but still fast ramp in data rate increase could be disrupted by new technologies. There are three technologies on the horizon that can rekindle the data rate ramp and pose significantly new challenges for testing.

The first long term test challenge is on silicon photonics. Because of the higher cost associated with today's discrete optical components needed (e.g. VCSEL, DFB, and lasers), the optical interface was mainly used for communications with reach distances of > 10 m communication links (e.g., LAN, Metro, and WAN networks). The recent silicon photonics advancements could potentially offer an alternative to the electrical interface for distances down to < 1 m. This is largely driven by the fact that the loss for the commonly used FR4 copper material becomes significant (e.g., ~ 5dB/in at 25 GHz), and the maximum reach distance would be limited to < 10 inch (or ~ 25.4 cm), significantly shorter than backplane distance required (e.g., ~ 40 in or 1 m). One may argue to use other relatively low-loss copper materials, but its higher cost (several times higher than FR4) diminishes the cost advantages for copper over fiber, even for reach distances in the range of 1m or shorter. Due to the loss, power, and cost characteristics between copper and fiber media, it is speculated that at 50~100 Gbps, many chip-to-chip and chip-to-module interfaces will use optical signaling and optical fiber as the media. The integration of some optical components on to silicon can potentially lower the cost barrier for using optical media instead of electrical. The challenge for integrated photonic devices in production ATE environment presents a new challenge in the future. In 2015, many of these OE integrations are using embedded waveguide and optics on MCM substrate, which demands a thorough package level testing. In case of true 3D stacking for the OE integration, the wafer level testing involve optical testing. In some cases, where the light signal comes out of under the a flip chip package, it is exceptionally difficult. However, we are glad to see some pilot programs in the ATE industry to facilitate integrated optical test solutions, beyond just a lab instrument based add-on. Test hardware design and signal integrity solutions will add a new dimension in the mixed optical and electrical environment.

The second long term test challenge is on multi-level encoded/modulated I/Os. In 2015, we started to see PAM4 based SerDes became the primary solution for 50Gbps and above. With the CMOS technology advances to support further growth in data rate, new technology may be developed to maximize efficient usage of the electrical media bandwidth. The first round of these technologies was focused on analog and digital equalization to offset media loss with the non-return to zero (NRZ) or pulse-amplitude modulation-2 levels (PAM-2). It is now a common feature to have adaptive equalization (e.g., decision feedback equalization (DFE)) for Gbps/GHz interfaces with data rates in the 6~25 Gbps range. Multilevel encoding scheme (e.g., PAM-M) was already used to reach Gbps data rate for media with much lower bandwidth for a few Ethernet standards (e.g., 1000Base-T, PAM-5). The test challenge is obviously high for multilevel encoding while most of the current tester instruments are binary in nature. New test solutions will need to be developed to align with the Gbps/GHz interface technology if it chooses to adopt PAM-M signaling. The transmitter test figure of merit will change to analog SNDR type of measurement, which calls for real-time scope over sampling capability. For receiver testing, multilevel signal generation and noise injection techniques is a major deviation from the binary instrument. Presently, even lab instrument capable of 50G+ PAM4 testing has very limited offering and extremely expensive.

Maximizing the reach of signal at these high data rate introduced a lot of the complexities too. To further expand the tolerance in bandwidth limiting copper links, the adaptive analog equalization (AEQ) and digital feedback equalization (DFE) are already widely used. The equalization adaptation loop response has been a challenge for design and test. On top of that, some industry standards introduced Forward Error Correction (FEC), which is accomplished by adding the redundancy to the transmitted data stream and correcting the errors after channel and at the receiver FEC decoder. FEC can be decoded via hard decision or soft-decision algorithms, with soft-decision has relatively better performance at the cost of more complex code and decode algorithms, which corresponds to more gates or silicon area, as well as power

consumption when implemented with ICs. FEC reduces BER performance down to 10^{-15} or lower for long channel reach (e.g., backplane channels) telecomm and datacom applications. FEC can also be used to mitigate the media loss in chip-to-chip (medium reach) and back planes (long reaches). Moreover, FECs are also used commonly in the optical datacomm and telecomm, to improve the optical link BER performance for a fixed reach, or extend the reach for a same BER, or the combination of both. FEC will introduce other challenges such as even faster speed from coding overhead by over-clocking. On the other hand, over-clocking could be avoided by introducing a transcoding/decoding before and after FEC encoding/decoding, this is the case for the 100G Ethernet where a 256b/257b transcoding is added after the 64b/66b encoding and before FEC. If very complex iterative decoding is used in the future, the power transient can be aggravate the Power Integrity challenge for testing.

There are many types of FEC, providing different coding gains (signal to noise ratio (SNR) gain for a same BER) and latency. For Ethernet, Reed–Solomon (RS) code was chosen for 100G-Based and 400G-Based Ethernets, due to its capability of correcting both random and burst errors. To maintain a low latency (e.g., 100 ns or less for 100G Ethernet), FEC is striped over multiple lances, and re-aligned via alignment marker (AM) at the receiver FEC decode. Moreover, FEC is mandatory. The testing implication is that final link test needs to include FEC, and is based on per link over multiple lances. This could pose challenges to today's high-speed testing method and infrastructure which is predominantly physical layer focused.

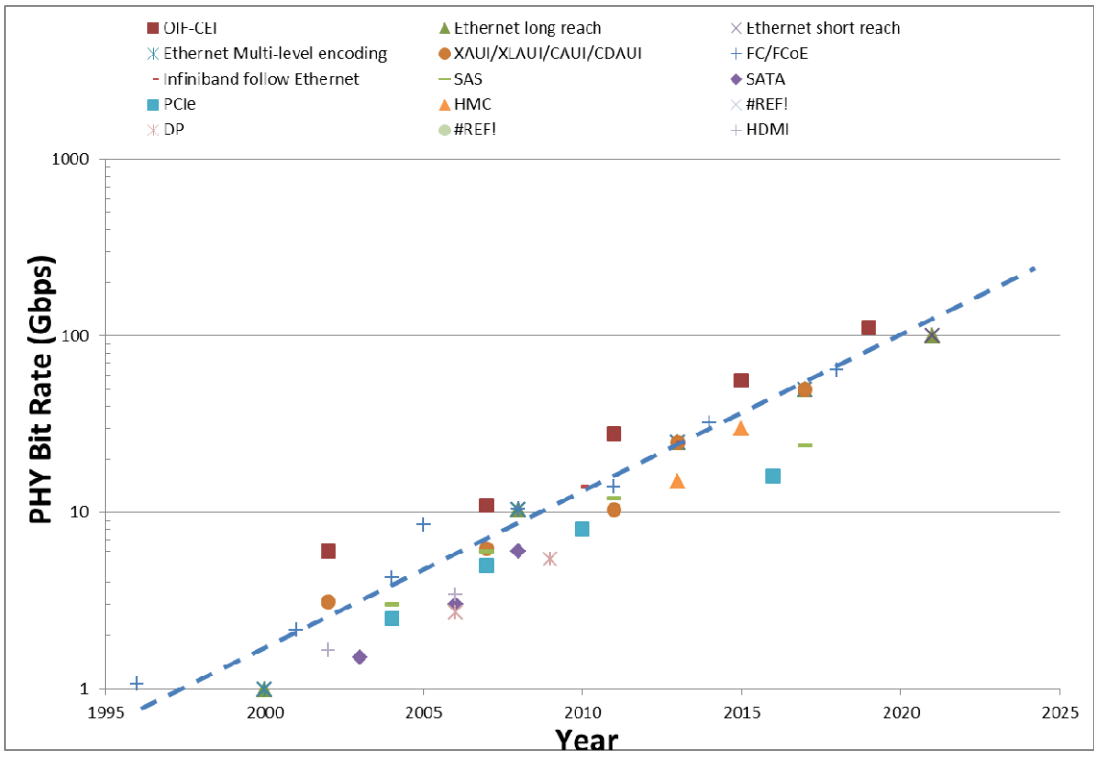


Figure TST22 - High Speed Interface Trend

In the past few years, the ATE industry have been providing instruments for the 8~10Gbps range for the higher volume product in the computing space. However, for the datacom and telecom application space, the ATE suppliers have been discouraged by the Return of Investment (ROI) from pursuing ATE a test instrument. The highest speed integrated ATE test solution in 2015 is at 16Gbps, which is not adequate for the 17~28Gbps new design as of 2015. Many unique features and performance demands related to specific standards made a generic ATE test solutions harder to achieve. Commercial high speed ATE test instruments are traditionally set to target high volume industry standards at the ramping phase to provide necessary test coverage for early adoptions, but lag behind the leading edge high speed interface serving enterprise needs. 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 tradeoff 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.

The near term challenge to high frequency IO testing stay the same in general except that we keep the threshold of pain to higher and higher frequencies. As in 2015, the 6G and below space became mature with well understood in design margin and defect mechanism with years of HVM test accumulated data. Even the 8G~10G space is ramping into very HVM with test solutions determined. There are still a lot of silicon learning and refinement effort in the 12~25Gbps range for these technology to become mainstream. There are a lot of challenge to construct customized test solutions to thoroughly testing these new designs in terms of defect isolation, performance and design margin optimization. With limited ATE instrument offering at this speed range, a lot of application based test module, add-on instrument, and analog DFT/BIST are filling the gaps. We anticipate that will be the near term challenge before the multilevel encoding based SerDes take the center stage around the 50Gbps range in the long run.

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. With the per transistor silicon cost scaling down, and high speed test instrument and interface hardware cost staying flat, this has encouraged more focus on DFT and BIST. Test hardware design in 25 to 100 Gbps/GHz needs a lot of innovation to keep the off-chip test solution relevant in high volume productions. However, with the design margin coming with a price in power and silicon area, 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 can vary from application to application. The overall goal is to minimize manufacturing test cost and efficiently test high port count devices.

7.4.1 IMPORTANT AREAS OF CONCERN

7.4.1.1 JITTER AND NOISE 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 Bit Error Rate (BER) for the serial link (typically 10^{-12} or even 10^{-15}). 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 (PWS) 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. Ideal data edges are commonly used as the references to determine the correlated DDJ, and clock edges of the recovered clock are used as references for determining uncorrelated DJ and RJ, PWJ and DCD are determined from data edge to data edge differences. 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 number of cycles, which could be more meaningful for clocking schemes only allowing very short-term jitter accumulation. Jitter measurement also imposes a very stringent signal integrity requirement. On the other hand, with the source synchronous bus entering the Gbps/GHz 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 noise floor of the instrument or tester needs to be significantly below the DJ and RJ of the DUT to avoid errors in the measured DJ and RJ. Instrumentation and testers for jitter measurement have also adopted the DJ and RJ merits for 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 **Error! Reference source not found.**. At 10 Gbps, the TJ, DJ, and RJ accuracy requirements are 3ps, 1.5ps, and 0.11ps. 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 Figure TST15 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.

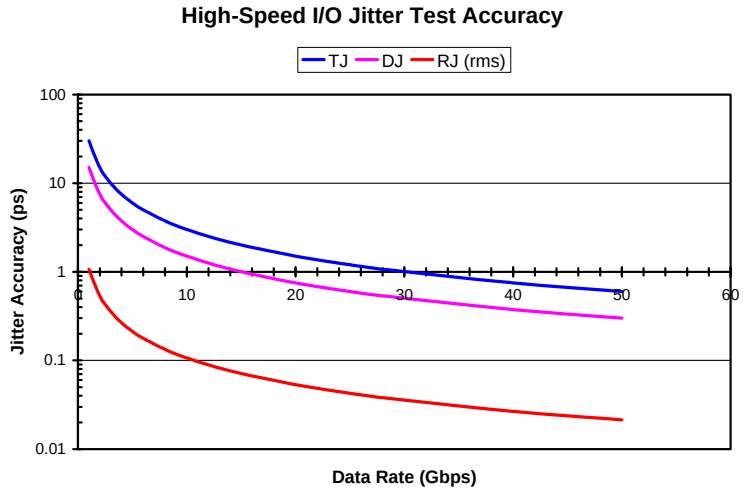


Figure TST23 - High Speed I/O Jitter Test Accuracy Requirements Scaling with Frequency

The long term challenge will be related to the multilevel encoded SerDes such as PAM4. Jitter can be viewed as 1-dimensional meteorology in time domain. PAM4 level transition related DDJ is a new effect for PAM4 signaling. With the PAM-4 case, the extra voltage level modulation margin introduces another dimension to the statistical based margining (i.e. noise). Furthermore, for PAM4, the signal noise ratio (SNR) is 9 dB worse compared with that of a NRZ. As such, PAM4 signaling is challenged by both jitter and noise.

The basic jitter theory applies to noise well. Noise can be separated to deterministic noise (DN) and random noise (RN), Total noise (TN) is the noise counterpart for timing TJ. Under DN, it includes data-dependent noise (DDN), bounded uncorrelated noise (BUN). For PAM4, there will be three stacked eye-diagram, corresponding to 3 pairs of TJ and TN, versus only one pair of TJ and TN, as shown in Figure TST24.

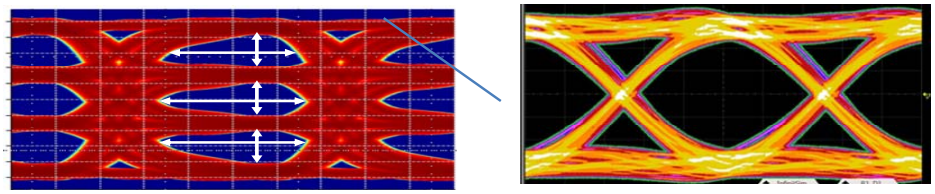


Figure TST24 Three stacked eye-diagram for PAM4 v.s. the traditional binary eye.

Noise for instrument and tester need to be significantly below that for a DUT. It typically needs to be at < 1 mv. Most of the 25-50 Gbps oscilloscopes have noise floor < 1 mv.

7.4.1.2 JITTER/INTERFERENCE 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. Depending on the architectures of the clocking schemes used, different types of jitter are 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 number of cycles. An emerging important topic is the receiver equalization test, especially feedback based adaptive equalizations such as DFE. The challenge comes from the fact that testing observability is intrinsically poor for Rx, with the BER as the only

common observable merit. Without the assistance of on-die or on-chip instruments such as an 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 in the previous ATE world, but are starting to become more available in the lab equipment world.

Worst case or cases "signal and noise test" in addition to jitter tolerance test is another new requirement for the SERDES tests, especially for backplane capable or PAM4 SERDES receiver test, where the channel loss can be very high (e.g., 25-35 dB at Nyquist), and the eye-diagram at the input of the receiver is closed, even when the transmitter equalizer such as feed-forward equalizer (FFE) is utilized. The goal for receiver signal and noise tolerance test is to insure that receiver can tolerant the worst combinations of transmitter and channel in terms of signal and noise, and is able to recover the signal at certain BER level (e.g., 1e-12, or 1e-15). All the receiver sub-circuit blocks (e.g., continuous-time linear equalizer (CTLE), DFE, CDR) is stressed and tested simultaneously. Insuring and calibrating the input signal and noise to the required worst conditions (e.g., worst transmitter signal swing, return-loss, rise/fall time, jitter and noise, channel insertion loss, return loss, crosstalk, and insertion loss deviation (ILD)) is a key for receiver signal and noise tolerance test, which can require a sophistical signal, jitter, and noise generator with a good precision and programmability, and a oscilloscope to measure and calibrate. The calibration may even include a simulation tool that warrants the worst signal and noise conditions are met.

7.4.1.3 TEST FIXTURE BANDWIDTH

The test fixture used to interface the device to the instruments/ATE includes a printed circuit board, cable, connector/pogo pin, etc… With frequencies and port counts increasing, the ability to deliver high frequency signals to instruments without much loss and distortion becomes a monumental task in test engineering. The bandwidth requirement is commonly set to be at 3-5th harmonic of the signal under test, depending on its rise/fall time, to have minimum ISI jitter from the test fixture. Bandwidth of 15-25 GHz for a 10 Gbps signal, and 37.5-62.5 GHz for 25 Gbps, are commonly required. 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 mechanic 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. Test hardware design for 25~28Gbps and above is still challenging for volume productions, with 50~56Gbps in the horizon.

7.4.1.4 DFT AND TFD

The basic Pseudo-Random-Binary-Stream (PRBS) generator and BER checker functions have been a norm for years now. In order to cope with adaptive equalizations, more Gbps/GHz designs now include internal data eye analysis capability. As more analog content is introduced into Gbps/GHz I/O design, more research and innovation is needed in analog DFT.

Because of the power and area requirements for GHz I/O design, the analog portion of the circuits is under constant pressure to be scaled down. As a process variation removal technique, the calibration and trimming become widely adopted. That leads to more digitally-assisted analog design implementations, which means more "on-chip self-calibrations" or "trimming based on test" are required. This has created a new test requirement – testing the robustness of the calibration firmware. The performance related to trimming also positioned the testing charter from defect detection into a new dimension of performance guarantee process. In other words, it is not only "design for testability" but also "test for design performance".

7.5 MEMORY

There have been significant changes in the Memory environment in recent years. Up until 2003, DRAM bits comprised approximately 90% of bits shipped per month. By 2009, NAND had become the dominant form of memory and comprised 85% of monthly bits shipped. During the same period, NOR Flash has largely migrated from parallel interface devices to serial interface devices with extremely small form factors in order to reduce PCB size, complexity and power. With the introduction of mobile smart phones in 2007 and tablets in 2010, the traditional dominance of PC

DRAM has been eroded by lower power LPDRAM which is required for longer battery life. New generation memory types such as PRAM, RRAM, STT RAM, and CBRAM along with 3D multilayer instantiations of NAND will further change the memory environment over the next decade. The continuing shift to a battery powered wireless environment will continue to drive changes in memory usage. Existing memory types will likely not be replaced, but will be used in joint solutions with newer memory types.

Memory density has kept pace with Moore's Law since the first DRAM was manufactured in 1969, but lithography cycles as well as the performance roadmap are expected to push out for litho nodes less than 20nm, so the typical 2 year technology pace is forecasted to stretch to initially 3 years and then 5 years later in the roadmap. Multi-layer 3D NAND technologies allowed the use of longer gate lengths, so the 2 year technology pace should continue for the near term.

From a test perspective, most memory will be structurally tested at wafer test utilizing with low pin count interfaces of 4 to 10 pins per device. Structural test, when used, will likely include a self-test performance validation.

Spreadsheet Table TST7 – Memory Test Requirements

7.5.1 DRAM

Historical trends show that PC DRAM has doubled its performance every 5 years and will reach a data rate of 8.4 Gb/s per I/O in 2022 with DDR6. However, DDR4 appears to be the end of the DDRx era as there are no further enhancements of the DDRx architecture beyond DDR4 in definition or development. DDR4 itself has severe PCB design restrictions in order to meet the I/O performance requirements, so an enhancement of the current DDRx single ended interface will present additional challenges and constraints. The Wide IO memory architecture is targeted to mobile applications such as phones and tablets and is an evolution of the DDRx that decreases I/O bit rates while expanding the number of I/Os up to 512. Hybrid Memory Cube (HMC) is targeted to servers where it provides very high performance over a SERDES interface however at increased cost. High Bandwidth Memory (HBM) is similarly targeted to graphics video cards and applications. Based on existing roadmaps, DDR3/4 will continue to serve the PC market for the foreseeable future. LPDDRx and GDDRx DRAM families will continue to be a driver for the near future.

DRAM will become increasingly difficult to scale in sub-20nm nodes and transistor wear out will increase the frequency of errors. On-chip error correction and memory management will likely become a requirement before 2020. Dynamic failure detection, analysis, and repair will become necessary over the product life. To enhance test productivity, new test-oriented architectures will be required. On-chip correction may also change the DRAM test paradigm.

Maintaining high ATE test parallelism is required over the roadmap period to manage test cost. However, probe card performance test at high parallelism may be a challenge at high GT/s due to the interface routing complexity required. These challenges will ultimately drive the need for die self-test.

7.5.2 FLASH

NAND will double in density ever year in the short term and slow to a doubling every 2 years. The doubling every 2 years will be faster than the projected lithography node migration due to the increase in the number of bits stored in a single memory cell from one and two to four on some cell types. Use of error correction allows greater uncorrected error rates and enables increased bits per memory cell. NAND bus width has continued to be dominantly 8 bit with a decreasing number of products at 16 bit I/O. As large amounts of NAND are being consumed in Solid State Drives (SSD), a new NAND interface may emerge that is more optimized for SSD use.

The need for internal voltages that are 3–8 times the external supply requirements is expected to continue in the test process, 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.

NOR memory density is expected to increase slowly over the roadmap period and remain flat toward the end of the roadmap. NOR has been transitioning from a parallel to a serial interfaces since 2007 to reduce package size, and power. Further increases in NOR performance along with increasing test requirements are not expected.

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.

7.5.3 EMBEDDED MEMORY

Embedded memory consumes greater than 80% of transistors in many MPU and SoC designs and will scale with the increase of transistors in these devices. Embedded Flash and DRAM bits will not match the density of standard DRAM and NAND. Newer memory types such as RRAM or STT RAM may become embedded over the course of the roadmap.

To enhance test productivity, new test-oriented architectures and/or interfaces will be required. Built-in self-test (BIST) and built-in self-repair (BISR) will be essential to test embedded DRAM and Flash memories cost effectively. 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 are tested and repaired on a memory tester, while the logic blocks are tested on a logic tester.

7.6 ANALOG AND MIXED-SIGNAL

The economic benefit of monolithic integration (SoC) and system in package (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. Now this trend has again escalated with the emergence of through silicon via (TSV) packaging technology driving the challenge in a 3rd dimension.

Another important trend impacting mixed signal and analog testing is the compelling economics of multi-site testing for devices manufactured in extremely high volumes, also called parallel test. To support parallel test, many more instrument channels of each interface type are required to keep test cell throughput and Parallel Test Efficiency (PTE) AKA Multi-Site Efficiency (MSE) high is of increasing importance to avoid severely impacting Units Per Hour (UPH).

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. In fact, as site count increases, overall throughput can decrease if good PTE is not maintained.

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 categorized by 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 three classes: low frequency—basic/minimum requirements for a mixed-signal ATE, high frequency 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.

The second key trend is to enable the economics of test through instrumentation density and parallel device testing. The level of parallelism shown in **Error! Reference source not found.** indicates the expected 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, partial BIST or elimination of the need for external instrumentation altogether.

Spreadsheet Table TST8 - Mixed-Signal Test Requirements

7.6.1 IMPORTANT AREAS OF CONCERN

- 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 debug 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.
- 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, require 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 as the density continues to increase in support of parallel test drivers.
- Improvements in analog/mixed-signal DFT and BIST are needed to support the items above.

7.7 RADIO FREQUENCY

Two main RF frequency areas are are distinguished apart from each other. These are the Mobility space and the infrastructure/automotive RADAR/Industrial space

The mobility area of the table is a dynamically evolving topic. In years past, WiMax was thought to be the next 4G digital communication standard and UWB would wirelessly connect all consumer devices was the novel trend touted by analyst and experts. However, LTE and LTE-advance have emerged as the world-wide 4G standard for all mobility devices and UWB was never commercially adopted. WiGig looks promising to take the place of UWB. The market was driven by evolving existing standards to larger bandwidths like WCDMA to LTE-Advance and 802.11a/b/g/n to 802.11ac WiFi to drive faster data rates. The primary drivers of consumer mobility devices are smartphones and tablets that use LTE/LTE-advance and WiFi as predominant standards.

The Mobility portion of table has been reworked to reflect those changes. RF SSB BW remains constant at 80MHz (driven by 802.11ac). The WiGig is in place in the near future requiring a 1760 MHz modulation as reflected in the tables.

The committee feels that system sensitivity driven by better phase noise and less spectral growth from non-linear distortion drives adoption of higher density modulation standards to carry forward faster data rates is the trend. The RF port count per device increases to 64 over time because frequency allocations are non-standard across the globe and the need for backwards compatibility to the many existing digital communication standards need to be fulfilled.

The popularity and acceptance of collision avoidance radar detection systems and point-to-point backhaul is driving the 2nd area of the table – Infrastructure/automotive RADAR/Industrial. That coupled with emerging point-to-point connectivity with a non-world-wide standard for frequency allocation is driving the 100GHz challenge. and the need for backwards compatibility to the many existing digital communication standards need to be fulfilled.

Internet of Things (IoT) is a market that is high volume cost driven more than a test technology driver. It primarily consists of huge numbers of RF sensors that transfer data into a global data distribution center which will then communicate to the outside world and internet

Spreadsheet Table TST9 – RF Test Requirements

7.7.1 IMPORTANT AREAS OF CONCERN

- 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.
- 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.

- The visibility of trends in the high RF ranges (> 8 GHz) is evolving.
- Probe is a challenge over 8 GHz and the timeframe for modulated waveform testing over 8 GHz is uncertain.
- Impedance standards and calibration methods for high frequency measurement at probe need to be created.
- 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.
- Most SOC test requirements are still trending from RF-to-BB and vice-versa. RF-to-Dig or line to line is limited but widely discussed.
- OEM is conceptualizing RF BIST/Loop back test methodology but functional and parametric tests still dominate the market.

7.8 RELIABILITY TECHNOLOGY REQUIREMENTS

The objective of reliability solutions and burn-in is to eliminate latent defects in ICs that will cause early-life failures and screen them out before the product is shipped to the customer. Reliability screens are critical to achieving the low failure rates required by high-reliability applications, such as automobiles. The visibility of automotive applications is increasing dramatically, as the electronics in cars includes not only engine and brake control, but also also communications, internet access, entertainment, GPS, collision avoidance, and many other nascent applications. Reliability requirements in other mobile and mass storage applications is also increasing in importance, as the number of ICs and their transistor density increases. Latent defects are typically removed by process improvements, design improvements and accelerated stress methods during the test process. 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 meet the reliability specification and the needs of the end customer while providing the best value for the reliability dollar spent.

In reliability circles, customer satisfaction is measured by the field failure rate or failures in time (FITs). The cost of reliability screening 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 three fundamental components—burn in duration, BIB/socket cost 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.

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 "escapes" from the manufacturing stress process.

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. The best way 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 and functional test at stress conditions 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 may operate fine at application conditions. Although these parts may contain no reliability defects, in situ stress will fail these perfectly functional parts—hence "over-kill." Determining the proper test method and interpretation of the test results are key ingredients of a successful in-situ burn-in strategy to ensure that the latent defects are identified and overkill is minimized. 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 Vmin 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 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 Vmin); 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.

7.8.1 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. In addition to burn-in, several alternate techniques such as IDDQ, high voltage stress, and wafer mapping are being used to try to improve device reliability.

Burn-in system technology must continue to evolve with device technology. 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. Without continued innovation by the burn-in system manufacturers in cooperation with the IC manufacturers, all these trends tend to increase the cost of burn-in systems and sockets.

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 (**Error! Reference source not found.**). In certain applications, system level burn-in complements or replaces traditional device level burn-in, but this typically involves a significantly increased cost, since the burn-in system, socketing solution and burn-in time tend to increase. Wafer level burn-in technology continues to be developed, but has made only limited inroads against traditional packaged level burn-in. The challenge here is to use techniques such as scan/logic and memory BIST (MBIST) to improve the technical feasibility of wafer level burn-in.

Spreadsheet Table TST10 - Burn-In Test Requirements

7.8.2 WAFER LEVEL BURN-IN

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 just 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, especially stacked die in mobile and mass storage applications. Reliability failures of the packaged part increase exponentially with the number of die in a multi-die package, so the need for reliable die before packaging is increasing substantially in importance. 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.

There are two methods of performing a wafer-level burn-in. 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. This is typically referred to as wafer-level stress, as it applies a stress voltage for a short time, but does not apply the high temperature of burn-in. Actual 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. DFT functions such as scan or BIST are enablers for WLBI.

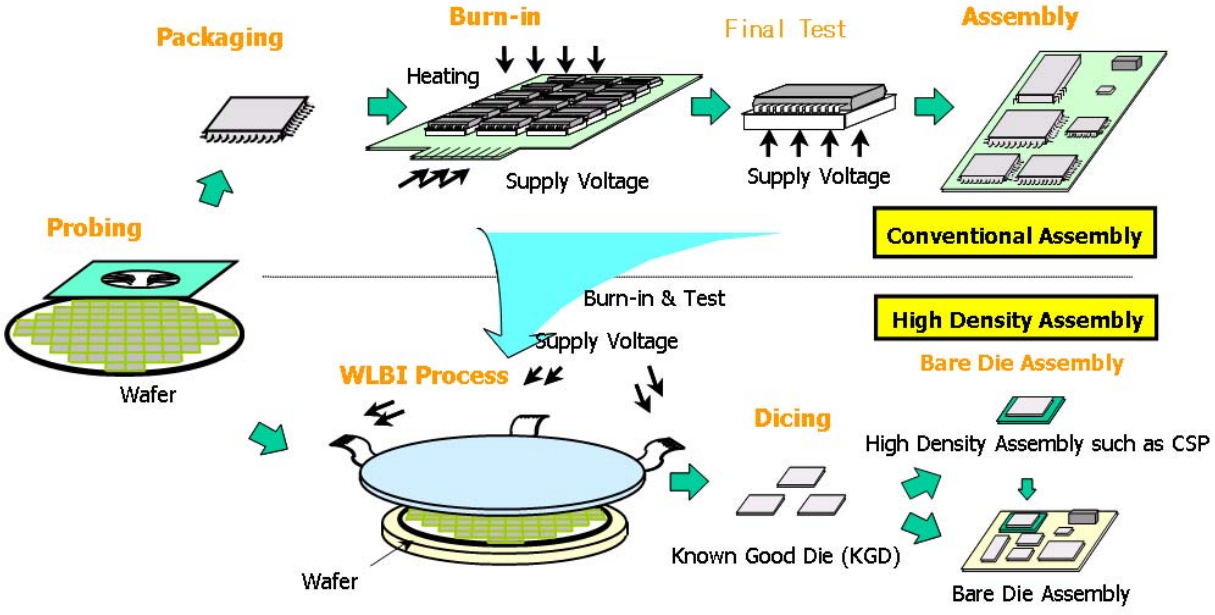


Figure TST14 - The Production Process with WLBI Compared with Package Burn-in.

The challenge for DRAM for example, 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 does not increase the cost of the final part. Leveraging the time spent in burn-in by using the burn-in environment as a massively parallel testing opportunity can effectively lower the overall cost-of-test

7.8.3 PROBING TECHNOLOGY FOR WAFER LEVEL BURN-IN

Full-wafer probing is a significant challenge, both technically and economically. The cost of a full-wafer probe tends to increase as the number of pads on the wafer increase and the pitch of the pads decreases. Contacting all the pads on a state-of-the-art wafer can require contacting in excess of 250,000 pads across a 300 mm wafer at a pitch of 60 microns or less over a wide temperature range. Intelligent use of DFT and pad placement rules by the semiconductor manufacturer can make this challenge less daunting. A WLBI 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. Other vertical pin contactors operate in a similar manner. The probe pitch is technology dependent.

For a pitch less than 70 μm, MEMS technology by use of photolithography is an option. This technology, however, is very challenging 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.

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.

7.8.4 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 in certain applications. 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 can 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 JTAG[3], the number of tested pins per device and total cost per device must be decreased and performance of the WLBI technology must be improved.

7.9 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. A better thermal solution is a very important parameter along with performance for better yield management. Memory applications are increasing the total power across a 300mm wafer and wafer probe needs to dissipate this total power to sustain the set-temperature during test. Power density per DUT is increasing and it's very challenging to manage a stable wafer level temperature test. 3D integration technology requires very precise probing technology in X, Y and Z, as micro-bumps are easy to get damaged during the probing process. MEMS applications require a variety of testing environments such as pressure, magnetic, and vacuum environment, also wafer shape and package style is becoming very unique depending on the application type.

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. 3D integration technology requires the new contact technology for the mid end process which will be added between

[3] *an IEEE standard 1149 boundary scan*

conventional front end process and back end process. New contact technology to probe on the diced die's micro-bump or C4 bump after the die is mounted on the interposer is needed. For the handler, the main tasks are the alignment accuracy to enable fine pitch contact, die level handling without damaging the die, and the tray design that supplies/receive the die.

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, immediate heat control technology, and temperature control to enable the stable DUT 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 wide temperature range, high temperature accuracy, high 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. The 2011 roadmap classified the handler types by the power dissipation of the device. In the 2013 roadmap, the classification is made based on the application type. Gravity feed, turret, and strip handlers have been added to the table while retaining the pick and place type handler. Gravity feed handler is used on SOP, QFN, and DIP packages. A Turret handler is widely used on discrete type QFN devices. Strip handlers are used on the frame before dicing. Strip test enables high parallelism with fewer interface resources which enables cheaper test cost. These additional three types of handlers are widely used on relatively low end or low cost devices. Evolution of these handlers are quite different but important for various type of LSI.

Figure TST26 - Test Handler and Prober Difficult Challenges

Pick and Place Handlers (High Performance)	Temperature control and temperature rise control due to high power densities
	Continuous lot processing (lot cascading), auto-retest, asynchronous device socketing with low-conversion times
	Better ESD controls as products are more sensitive to ESD. On-die protection circuitry increases cost.
	Lower stress socketing, low-cost change kits, higher I/O count for new package technologies
	Package heat lids change thermal characteristics of device and hander
	Multi-site handling capability for short test time devices (1–7 seconds)
	Force balancing control for System in Package and Multi-Chip Module
Pick and Place Handlers (Consumer SoC/Automotive)	Support for stacked die packaging and thin die packaging
	Wide range tri-temperature soak requirements (-55ºC to 175ºC) increases system complexity for automotive devices
	Device junction temperature control and temperature accuracy +/-1.0 deg C
	Fine Pitch top and bottom side one shot contact for Package on Package
	Continuous lot processing (lot cascading), auto-retest, low conversion times, asynchronous operation
Pick and Place Handlers (Memory)	Thin die capable kitless handlers for a wide variety of package sizes, thicknesses, and ball pitches < 0.3mm
	Package ball-to-package edge gap decreases from 0.6 mm to 0 mm require new handling and socketing methods
	Parallelism at greater than x128 drives thermal control +/-1.0 deg C accuracy and alignment challenges <0.30mm pin pitch
Prober	Consistent and low thermal resistance across the chuck is required to improve temperature control of the device under test. There is a new requirement of active/dynamic thermal control, which can control junction temperature(⊿T) during test Heat dissipation of >200 Watts at > 85ºC 3DI and MEMS application require very thin, thick and special shape wafer testing (handling) technology, but no industry standard exists Probing on micro-bump is technically proven but there are many challenges "parallelism/multi-site", "Thermal conduction" and "bump damages/reliability" Advances in probe card technology require a new optical alignment methodology.
	Dicing frame probers can cover a wide temperature range, but a dicing sheet cannot cover the full range. Greater parallelism/multi-site, and higher pin counts require higher chuck rigidity and a robust Probe Card changer. Power Device applications require very thin wafer which drive need for 'Taiko Wafer' and 'Ring attached wafer' handling technology
	No standard cassette / FOUP exists for thin wafer / warp wafer
	Enhanced Probe Z control is needed to prevent damage to pads
	Probers for wafer sizes greater than 300mm may drive changes for probe floor layout, operation, and building design
Gravity Feed Handlers	Thinner packages and wafer will require a reduction in the impact load to prevent device damage
	Test head size increase due to higher test parallelism may alter handler roadmap
	Reduction of static electricity friction and surface tension moisture friction on very small

	packages (<1 x 1 mm)
Turret Handlers	Test contactor support for > 100A current forcing on power devices
	Kelvin contact support (2 probes) to very small area (0.2 x 0.2mm) contacts on small signal devices
Strip L/F Handlers	Testing process infrastructure configuration
	Accuracy of the contact position for high temperature testing environment

Spreadsheet Table TST15 – Prober Requirements

Spreadsheet Table TST16 – Handler Requirements

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

7.10.1 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 t 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.

7.10.2 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 and300 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.

Figure TST27 – Probing Difficult Challenges

Geometry	Pad size
	Moving to Memory pad sizes below 45 x 45 um for full wafer contactors on Dual temperature probe cards remains a significant challenge in both probe card design and use environments to manage the expansion and contraction of the probe card. Also for Full wafer contactor probe card overall prober accuracy in both X, Y, Z and theta are critical for smaller pad sizes especially dual temperature probe cards.
	Pad pitch
	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.
	Pad / Bump pitch vs. Space transformation from probe pitch to Probe card PCBA pitch
	As pad pitch and especially array bump pitch are reduced the challenge of building space transformers that permit full functional test remains an difficult challenge as array size and depth increase.
Parallel test	Increasing probe array planarity requirements in combination with increasing array size.
	Different varieties of cleaning materials are necessary to maintain appropriate contact resistance and scrub performance. Probers need to store and handle multiple cleaning materials. There are materials which are difficult to focus on its surface to detect height, new technology requires applying appropriate over travel during cleaning to not decrease probe card life time.
Probing at temperature	High temp test and low temp test requires appropriate pre-heat/sort time to settle probing position changes caused by heating or cooling. More pre-heat/soak time causes less system utilization. Also probers and user environments will need to manage the thermal movement of probe cards during wafer and lot changes to minimize time lost due to re-soaking.
	Probing high power devices at high parallelism at temperature will require probers that can remove heat from chucks generated by the wafer.
Probe force	Rigid stiffener is necessary to support higher probe force. This brings cost increase and difficulty to handle heavier probe card.
Probe cleaning	Different cleaning materials are necessary to maintain appropriate contact resistance and scrub performance. Probers need to store and handle multiple cleaning materials. There are materials which are difficult to focus on its surface to detect height, new technology requires applying appropriate over travel during cleaning to not decrease probe card life time.
	A self- cleaning probe card is required for fine pitch bumped pad devices
Lead time	Development of high temperature (85°C–150°C) in situ cleaning mediums/methods, particularly for fine pitch, multi-DUT, and non-traditional probes.
	First article probe cards within the FAB cycle time from Tape out to wafer out remains a significant challenge, especially as the probe card complexity increases
	Space transformer lead times are too long, thus causing some vertical probe technologies to have lengthy lead-times.
Contact resistance Optical interface probing	Repair lead times need to be reduced
	Achieving low contact resistance with lower and lower probe force necessary for small MicroBumps as well as very high probe count full wafer contactor probe cards
	A method to measure contact resistance is needed. The traditional continuity test is insufficient to monitor contact resistance.
	CMOS Image sensors - increasing parallelism to full wafer contactor remains a challenge in the light source

TSV Probing	Solutions for probing communications optical transmitter / receivers need to be developed
	Probing on TSV micro-bumps down to 20um will only be viable if limited pad damage can be achieved.

7.10.2.1 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.

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.

7.10.2.2 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.

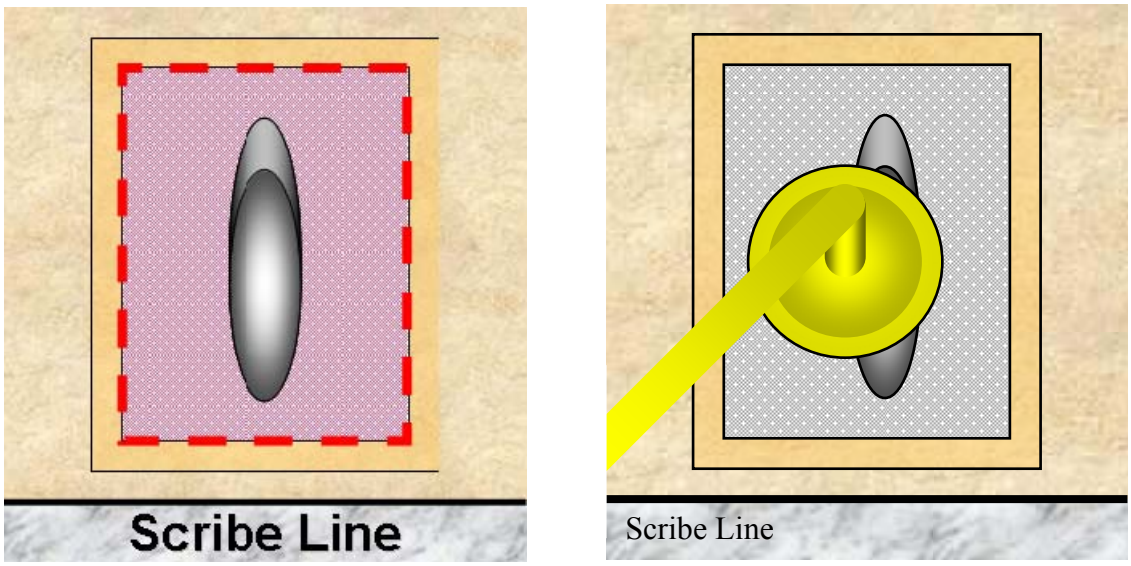


Figure TST28 - Probing and Wirebond Contacting a Bond Pad

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.

7.10.2.3 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 **Error! Reference source not found.** 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 for SoCs 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.

7.10.2.4 ELECTRICAL PERFORMANCE

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 probe 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 Spreadsheet Spreadsheet Table TST18 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.

7.10.2.5 THERMAL PERFORMANCE

Though stable through the roadmap horizon, the thermal environment for 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.

7.10.2.6 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.

7.10.2.7 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.

Spreadsheet Table TST18 – Wafer Probe Technology Requirements

7.10.3 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 semiconductor design 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.

Spreadsheet Table TST19 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. 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 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 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 2016, 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 uses with high frequency device test 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 is 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 life cycle is normally shorter than a Contact Blade type contact.

It is capable of meeting 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 a mass production environment. The 40GT/s requirement in 2019 also is 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.

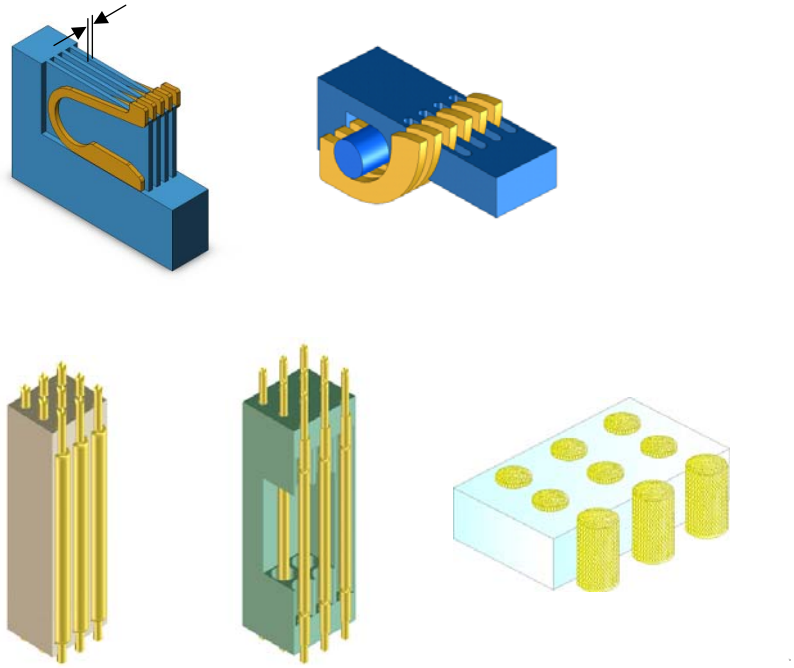


Figure TST15 - Contactor Types

7.10.3.1 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. To handle these high frequency signals, the user has to carefully consider the signal integrity of overall test system including board design/components/socket.

7.10.3.2 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.

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.

Spreadsheet Table TST19 – Test Socket Technology Requirements

7.11 SPECIALTY DEVICES

The test roadmap is not all inclusive as it does not contain test requirements for all devices. Many of the test requirements for some omitted devices will fall within the bounds specified for devices contained 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 2015, LCD display drivers, imaging devices and MEMS multimode sensor devices are featured in this section.

7.11.1 LCD DISPLAY DRIVERS

LCD display drivers are unique because of the die form factor, which can have larger than a 10:1 aspect ratio, and thousands of very narrow gold bump pads requiring contact for test. In 2015, LCD display drivers probing pad pitch was already down to 18μm. Technical difficulties in the gold bump process have slowed down pad pitch shrinkage and the alternative solution has been to shorten the pad length and reduce bump height. Right now the cantilever probe card till is major cost effective solutions to achieve probing LCD driver such narrow and fine pitch gold bump in mass production. The new feature of LCD display driver is high speed I/O protocol interface such as CalDriCon up to 2.5Gbps in 2015 and predict rising to 5Gbps in 2019. The potential solution of coaxial cantilever probing structure could provide 2.5Gbps test requirement in 2015. Foreseeing the probing challenges are becoming more severe and new technology is needed to economically solve the probing problems.

7.11.2 IMAGING DEVICES

Imaging devices are widespread on different applications. The digital cameras became a standard feature in cell phones and mobile computing devices. Most of DSLR camera started to change design to use CMOS Image sensor for increase resolution up to 24 Mega pixels in 2015 and foresee 50 Mega pixels in 2016. Automotive industry applications and surveillance need fast frame rates over 120 fps to 1000 fps in2015. The other new innovative technology of imaging device is IR sensor for thermal or 3D imaging application. Beside application, the innovation of design and manufacture process also have breakthrough in this years. Such as Back-side illumination BSI which increase incident light intensity and fill factor of sensor cell of each pixel instead of Front-side illumination when image pixel size smaller than 1.4μm and sensors coupled with image processing digital logic along with memory in a single chip implemented by a 3DS IC. The test challenges of probing technology became more severe for the trend of increasing test parallelism “data rate“, functionality, integration and optical systems. The process of WLCSP and WLCCM (wafer level camera module) is getting mature and blooming. How to achieve high parallelism testing after dicing on wafer form for reducing test cost to fulfill consumer application will be another big test challengers need to overcome in next decade.

7.11.3 MEMS SENSOR DEVICE

MEMS sensor devices were increasingly being incorporated into personal electronic devices and automotive application in last decade. The new application of HealthCare and IOT continually drive MEMS technology toward new area and market growing. The MEMS sensor devices for portable or wearable device need smaller die size with higher functionalities integration for example multi-mode MEMS sensor devices which integrated functionalities of accelerometers, gyro, magnetic, optical and pressure sensor with MCU in one package. The complexities of multi-stimulus testing input increase the challenges of test cost. The innovation of MEMS DFT technology become more and more importance. The novel 3DSIC process of MEMS sensor will benefit BIST development and increase test parallelism during wafer test for yield enhancement and cost reduction.

8 ACKNOWLEDGEMENTS

This document and the associated set of tables were prepared by the ITRS Test Working Group consisting of Mike Alfano, Kenichi Anzou, Takenori Aoki, Jeff Arasmith, Dave Armstrong, Roger Barth, Don Blair, Ben Brown, Carl Buck, Yi Cai, Wendy Chen, Steve Comen, Zoe Conroy, Bill Eklow, Ira Feldman, Shawn Fetterolf, Derek Floyd, Anne Gattiker, Shuji Hamada, Kazumi Hatayama, Frank Hensel, Hirokazu Hirayama, Hisao Horibe, Marc Hunter, Hideki Ichikawa, Shuichi Ito, Arasmith Jeff, Brion Keller, Satoru Kitagawa, Hidehiro Kiyofuji, Masahito Kondo, Ken Lanier, Lenny Leon, Marc Loranger, Adriano Mancosu, Phil Nigh, Katsuhiko Okada, Jayson Park, Mike Peng Li, Jeff Rearick, Mike Rogers, Tomonori Sasaki, Keno Sato, Masayuki Sato, Yasuo Sato, Ulrich Schoettmer, Hideo Shimada, Ryuji Shimizu, Yoichi Shimizu, Fumio Sonoda, Hiroyoshi Suzuki, Katsumi Suzuki, Tetsuo Tada, Masatomo Takahashi, Satoru Takeda, Tomoaki Tamura, Masanori Terasawa, Steven Tilden, Yuichi Watanabe, Tian Xia