2023 Edition

Chapter 17: Test Technology

For updates, visit http://eps.ieee.org/hir

The HIR is devised and intended for technology assessment only and is without regard to any commercial considerations pertaining to individual products or equipment.

We acknowledge with gratitude the use of material and figures in this Roadmap that are excerpted from original sources.
Figures & tables should be re-used only with the permission of the original source.

Table of Contents

Chapter 1: Heterogeneous Integration Roadmap: Driving Force and Enabling Technology for Systems of the Future

Chapter 2: High Performance Computing and Data Centers

Chapter 3: Heterogeneous Integration for the Internet of Things (IoT)

Chapter 4: Medical, Health and Wearables

Chapter 5: Automotive

Chapter 6: Aerospace and Defense

Chapter 7: Mobile

Chapter 8: Single Chip and Multi Chip Integration

Chapter 9: Integrated Photonics

Chapter 10: Integrated Power Electronics

Chapter 11: MEMS and Sensor Integration

Chapter 12: 5G, RF and Analog Mixed Signal

Chapter 13: Co-Design

Chapter 14: Modeling and Simulation

Chapter 15: Materials and Emerging Research Materials

Chapter 16: Emerging Research Devices

Chapter 17: Test Technology

- Executive Summary and Scope 1
- Section 1: RF Test 2
- Section 2: Test of Photonic Devices 4
- Section 3: Logic Testing 9
- Section 4: Specialty Device Testing 16
- Section 5: Memory Test 19
- Section 6: Analog and Mixed Signal Test 23
- Section 7: Wafer Probe and Device Handling 29
- Section 8: System Level Test 39
- Section 9: Data Analytics 40
- Section 10: 2.5D & 3D Device Testing 46
- Section 11: Key Drivers and Test Costs 49

Chapter 18: Supply Chain

Chapter 19: Cyber Security

Chapter 20: Thermal

Chapter 21: SiP and Module

Chapter 22: Interconnects for 2D and 3D Architectures

Chapter 23: Wafer-Level Packaging, Fan-in and Fan-out

Chapter 24: Reliability

Chapter 17: Test Technology

Executive Summary and Scope

> Send corrections, comments and suggested updates to the TWG chair, using our HIR SmartSheet:
>
> **https://rebrand.ly/HIR-feedback**

The 2023 revision of the HIR test chapter has aimed at focusing on test trends resulting from semiconductor market and technology inflection points and emerging use cases, and less on providing extensive market context commentaries.

Sections Updated in 2023 Revision

Below we provide a high-level summary of the key test challenges and needs for each of the device types addressed in the test sections that were updated in this 2023 test chapter revision.

RF Test: Need 1) Non-frequency-gapped ATE RF test capability in the 0-100 GHz frequency range, either for characterization, quality assurance, and/or high-volume production testing; 2) Higher ATE RF bandwidth production test capability up to 400 MHz for Wi-Fi 7 (with EVM in the 48+ dB range) and satellite; and up to 2 GHz to support 5G mmWave, UWB, and 6G THz; and 3) High-volume over-the-air (OTA) handler-based testing for mmWave and THz, and possibly automotive radar, will become increasingly relevant as DIB cabling for increased site count becomes cost-prohibitive.

Photonics Test: Need 1) Novel test approaches for testing optics in co-packaged heterogeneous devices in high volume; and 2) Emphasis on test time containment and test time reduction as the number of lanes and wavelengths per fiber increase.

Logic Test: Need 1) New test methods for testing chiplet devices with mixed technologies (for example, need for retargetable test IP for next level of integration into SIP or system); 2) test methodologies using Silent Data Corruption (SDC) logic testing methods; and 3) Standardized test interfaces and methods for chiplets that can be used by both chip foundries and packaging integrators (such as OSATs).

Specialty Test: Need 1) Higher test parallelism to reduce cost of test; and 2) multi-functional and cost-effective test capabilities as specialty devices become part of heterogeneous packages.

Memory Test: Need 1) Test capabilities for addressing higher interface speed, power, and thermal management requirements; 2) Test capabilities for overcoming the challenges of electro-mechanical interface capability of wafer and component test as NAND memory density increases due to vertical scaling; and 3) Testing of higher DRAM bandwidth requirements.

Analog/Mixed Signal Test: Need 1) High speed instrumentation that can accept, force, and tolerate higher voltages and currents, driven by wide bandgap materials; 2) DC accuracy below 50 uV over the entire temperature range; 3) Closed-loop temperature forcing test capability at final test; 4) Test capabilities for A/MS devices housed in heterogenous packages; 5) Novel test solutions for overcoming the inherent physics of high voltage test at very high multisite testing; 6) High density floating resources with high accuracy, medium current capability, and large isolation voltages; and 7) Need for fully floating low-speed digital instrumentation for testing chip-to-chip communications devices which are shifted by tens to hundreds of volts above or below system ground.

System Level Test: Need 1) Flexible DFT architectures for both structural and functional test content; 2) Effective SW/HW system failure diagnosis methods; and 3) Deep component parametric data extraction to data analytics.

Data Analytics: Need 1) For advanced and comprehensive data analytics solutions that take full advantage of data from across the entire value chain; 2) Significant improvements in the development and adoption of key enablers such as communications infrastructure, data interchange formats, traceability, data security, and advanced data analytics algorithms; 3) Efficient methods for accessing, curating, managing, and analyzing data from on-chip sensors IP, equipment sensors, and test results.

2.5/3D Test: Need 1) Known-good-die DFT test methods that enable high quality wafer probe test – thus reducing fallout at final test; 2) Faster die-to-die communication standards that enable thorough testing at final test; 3) Standardized test and repair methodologies that consider new trends in 3D interconnects; 4) Yield prediction and analysis methods that ensure fallout at all levels of testing are understood; and 5) End-to-end data analytics capability that applies to all dies on the package.

Test Cost: Need 1) New probing technology which allows testing of singulated die; 2) New PCB and interposer technology to lower the cost and complexity of consumable materials; 3) Improvements in the test process by increased use of data analysis and machine learning based on measured data; and 4) Cost reduction of system-level testing.

Test Technology Working Group Leadership Team

Co-Chairs:	Ken Butler Jeorge Hurtarte		
RF Test:	Jeorge Hurtarte	**Analog/Mixed Signal Test:**	Rich Dumene
Photonics Test:	Dave Armstrong	**System Level Test:**	Harry Chen
Logic Test:	Marc Hutner	**Data Analytics**:	Ira Leventhal
Specialty Test:	Wendy Chen	**Test Cost:**	Ken Lanier
Memory Test:	Jerry McBride		
2**.5D/3D Test:**	Morten Jensen and Boris Vaisband		

Section 1: RF Test

In the mobile wireless sector, history shows that there is a new "G" every 8-10 years. Thus, while we saw the emergence of 5G in both the sub-8 GHz and mmWave (24-53 GHz) during 2018-2022, we can expect 6G (THz) product prototypes to start emerging in the 2027-2030 timeframe.[1]

In the Wi-Fi connectivity wireless sector, Wi-Fi 7 (802.11be) in the 2.4, 5, and 6 GHz spectrum bands (the latter extending up to 7.125 GHz with up to 320 MHz of bandwidth) will see initial volume production in the 2024-2025 timeframe.[2] We can also expect that micro positioning capability will be added into Wi-Fi 7 at 320 MHz (802.11bk), in addition to 802.15.4z UWB (up to 11 GHz).

In addition to mobile and connectivity, automotive radar applications in the 76-81 GHz frequency range will continue adoption, while satellite connectivity in the Ku-band (12-18 GHz), and Ka-band (26.5-40 GHz) will see higher volume in the 2025-2030 timeframe as an effort to reach remote areas.

These wireless market segment technologies trends translate into the following high level ATE requirements up to 2030, which are further discussed below.

- Non-frequency gapped ATE RF test capability in the 0-100 GHz frequency range, either for characterization, quality assurance, or high-volume production testing. It is also likely that "IF" frequencies for 6G THz will fall within this 0-100 GHz range.
- High-volume over-the-air (OTA) handler-based testing for mmWave and THz, and possibly automotive radar, will become increasingly necessary in the 2025-2030 timeframe as DIB cabling for increased site count becomes cost-prohibitive.
- Higher ATE RF bandwidth production test capability up to 400 MHz for Wi-Fi 7 (with EVM in the 48+ dB range) and satellite; and up to 2 GHz to support 5G mmWave, UWB, and 6G THz.

For mobile devices, we will see the expansion of 5G millimeter wave into the 71 GHz range with the adoption of 3GPP Release 17.[3] In addition, with the advent of 6G, we can expect RF frequencies beyond 100 GHz into the THz range.[4] These two trends will require non-gapped frequency test capabilities from "0-100 GHz" as customers will not want to have multiple instruments to test different frequency ranges. While it is not yet clear that GHz and THz devices will be 100% tested at those frequencies in production, such capabilities need to be present in the tester for characterization and quality assurance purposes (for example, for analyzing field failures).

Millimeter wave and THz will require novel and cost effective over-the-air testing (OTA) methodologies, which started to appear around 2022 from companies such as Teradyne and Advantest, but these will require more maturity

[1] https://www.testconx.org/premium/wp-content/uploads/2021/TestConXMesa2021s1p1Hurtarte_9106.pdf
[2] https://www.ieee802.org/11/IEEE%20802-11-Overview-and-Amendments-Under-Development.pptx
[3] https://www.qualcomm.com/documents/download-our-5g-nr-rel-17-presentation
[4] https://www.qualcomm.com/content/dam/qcomm-martech/dm-assets/documents/Qualcomm-Whitepaper-Vision-market-drivers-and-research-directions-on-the-path-to-6G.pdf

to achieve high-volume-handler-ready solutions.[5] [6] OTA test techniques will compete with other more cost-effective methods, yet may not be as reliable for performance testing, such as “leakback” and “radiateback” test techniques. Such alternative solutions will push the limits of the device interface boards (DIB) wiring and cabling for multisite device testing, and thus the need for cost-efficient and high-performance handler-based OTA test techniques.

Table 1 shows an increased bandwidth requirement, as a minimum, for characterization testing of various millimeter wave devices, most notably in the 2GHz bandwidth range for higher volume use cases (e.g., 5G FR2-2).

Table 1: RF Frequency and Bandwidth Requirements for 2020-2030

RF Frequency and Bandwith Requirements for 2020-2030				
Wireless Standard Technology	**Min Frequency (GHz)**	**Max Frequency (GHz)**	**CC Bandwidth (MHz)[1]**	**Description**
3GPP TS 38.101-1	0.4	7.125	100	5G FR1
802.11ax	0.4	7.125	160	Wi-Fi 6E
802.11be	0.4	7.125	320	Wi-Fi 7
802.15.4z	1	11	1300	UWB
Satellite	12	18	250	Ku VSATs
ETSI TR 101 982	21	27	200	24 GHz SSR Auto Radar SS
Backhaul	18	38	60	BTS Backhaul
Satellite	26	40	250	Ka VSATs
3GPP TS 38.101-2	24.25	52.6	400	5G FR2-1 mmW
3GPP Rel. 17	52.6	71	2000	5G FR2-2 mmW
Backhaul	57	66	4000	BTS Backhaul
802.15.3c	57	66	5500	Motion sense / Hand gesture
3GPP TR 38.806	52.6	71	1000	5G FFS mmW
802.11ay	55	76	4000	WiGig
Backhaul	71	76	4000	BTS Backhaul
ETSI TR 101 983	76	77	1000	77 GHz LRR Auto Radar FMCW
ETSI TR 101 263	77	81	4000	79 GHz SRR Auto Radar FMCW
Backhaul	81	86	4000	BTS Backhaul
4D Imaging Radar	77	86	4000	SRR 4D Imaging Radar
Backhaul	92	95	4000	BTS Backhaul
U-SRR	120	140	> 4000	cm radar
6G (THz)	95	3000	> 4000	6th Generation Mobile Networks
Note 1: CC = Component Carrier				

IEEE 802.11 continues to work on new connectivity Wi-Fi standards such as 802.11be (aka Wi-Fi 7) with a maximum channel bandwidth of 320 MHz and 4k QAM modulation.[7] Thus, the key test requirements for Wi-Fi 7 are the capabilities to test waveforms with 320 MHz bandwidth in a single measurement at EVM of greater than 48 dB. The more stringent EVM requirement stems from the 4K QAM (Quadrature Amplitude Modulation) which enables each signal to more densely embed greater amounts of data compared to the 1K QAM with Wi-Fi 6/6E. For high order modulations such as 4096-QAM, which require stringent transmitter accuracy, selecting test equipment with a low EVM floor is critical, otherwise the error uncertainty contributed by the test equipment reduces the confidence in the final measurement.[8]

UWB (Ultra-Wideband) is defined in the IEEE standard 802.15.4 for micro positioning applications. Test requirements will continue to be imposed for testing Time of Flight (ToF), Two Way Ranging (TWR), and Angle of Arrival (AoA), at full spectrum bandwidth (see Table 1).[9] In addition to the 802.15.4 UWB standard, a new IEEE

[5] https://www.teradyne.com/2022/08/17/the-future-of-wireless-test-is-over-the-air/
[6] https://www.testconx.org/premium/wp-content/uploads/2021/TestConXMesa2021s1p2Semancik_2948.pdf
[7] https://www.intel.com/content/www/us/en/products/docs/wireless/wi-fi-7.html
[8] https://www.litepoint.com/blog/error-vector-magnitude-why-it-matters-and-how-its-measured/
[9] https://www.litepoint.com/uwb/

802.11bk standard is emerging for micro positioning at 320 MHz bandwidth and thus such testing capabilities will need to be added when this standard becomes available in late 2024.[10] [11]

Beyond mobile and connectivity device test requirements, Table 1 also shows various other RF wireless applications requiring test capabilities in the millimeter wave range, such as Ka/Ku VSATs for satellite rural internet deployments[12], automotive radar in the 77 GHz and 79 GHz frequency bands for SAE levels L4-L5 autonomous driving, and other applications such as base transceiver station (BTS) backhaul, and hand gesture/motion detection applications.[13] These miscellaneous millimeter wave use cases are likely to require similar test capabilities as explained above for 5G FR2-1 and FR2-2 mobile devices.

Section 2: Test of Photonic Devices

Executive Summary

In the electronic integrated circuit (EIC) industry, testing has become a mature process supported by practices and equipment that have been heavily optimized to drive down the cost and time spent on IC testing. In contrast, development of similar methods and tools for the photonic integrated circuit (PIC) community is still at an early stage, and the extra complexity that arises from having to measure both in the optical and the electrical domain poses many challenges. In this section. we define a number of key areas where development is needed, and in each of these areas we strive to leverage as much as possible the existing knowledge, practices and infrastructure from the EIC industry.

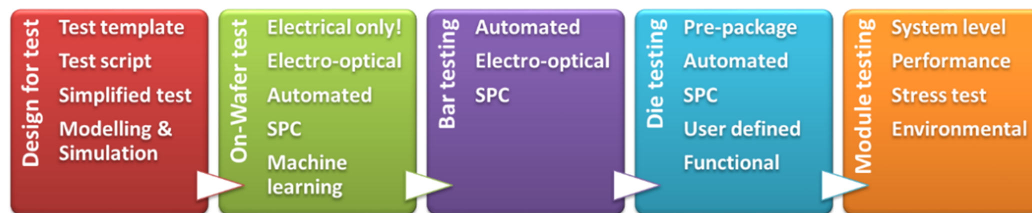


Figure 1. Overview of the test processes across the manufacturing chain of photonic integrated circuit based modules. Statistical process controls (SPC) require adequate test methods and data collection plans which should be accounted for already at the design phase.

A summary of photonic device test methods is available **at this link**. Based on that information, we see three key development areas:

- Standardization of test metrics
- Consolidation of design and test workflows
- Test time reduction

INTRODUCTION

This Test section focuses on unique attributes of testing optical devices, concentrating primarily on testing data communications products.

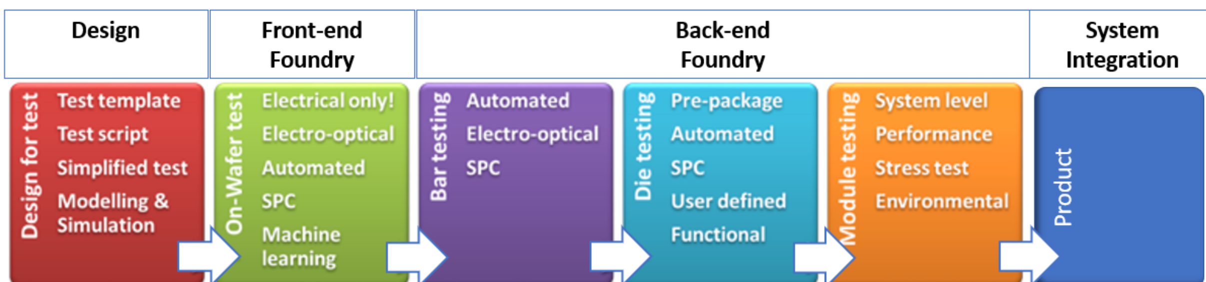


Figure 3. An overview of the PIC production chain for test.

[10] https://mentor.ieee.org/802.11/dcn/22/11-22-1353-02-00az-11bk-320mhz-ftm-csd.docx
[11] https://www.ieee802.org/11/Reports/802.11_Timelines.htm
[12] https://www.prnewswire.com/news-releases/satellite-internet-roll-out-to-gain-momentum-in-rural-areas-factmr-projects-c-band-to-remain-preferred-frequency-band-301404693.html
[13] https://www.infineon.com/cms/en/product/promopages/60GHz/

In each step of the test chain that is followed by the components that will form an end product, different requirements and methods are used. This chapter will discuss both the separate steps and the connection between those steps, regarding the product and data flow.

Areas of testing needed during a product life cycle are:

- during development to prove functionality and de-bug devices
- qualification testing
- pre-production validation
- in-process production testing to assure product quality, reliability and to improve yield.

This section contains an overview of PICs made on InP, SiN, SiPh, GaAs, Polymers and CMOS platforms. Elements such as fiber couplers, fiber arrays, lenses, optical and electrical interconnects and the standardization of test port positions (optical, DC, RF) will also be discussed. The kinds of testing required vary over the life cycle of a product (Figure 2). This figure lists typical optical device test activities and requirements during the life of a device from conception through the in-use and end-of-life phases. A roadmap of quantified key attribute needs is available **at this link**. Considering that data, a projection of the key industry needs is shown in Table 1.

Table 1. Key challenges with respect to test between 2020 and 2040

	2020	2025	2030	2035	2040
Adopt semiconductor EIC industry test practices					
Test procedures from custom to standardized					
Standardization of test structures					
Test data exchangeability and analysis					
Technology agnostic testing					
Test automation					
Design for test					
Application agnostic testing					

Red: Not current industry practice; **Orange:** Partial industrial coordination; **Yellow:** Significant industrial coordination and compatibility; **Green:** Established Industry standard.

Each category is broken down in more specific subcategories in the following tables, following the same roadmap guidelines. Each table addresses areas such as key challenges, test practices, transition from custom to standardized procedures, transfer of data, adopting semiconductor test practices, and design for test both at the die level and the software level. The tables show competences going out beyond 5 years and emphasize relative strengths for each area.

Table 2. Adopt semiconductor EIC industry test practices

	2020	2025	2030	2035	2040
6 Sigma methodology					
Documenting and reporting					
The same metrics but methods may vary					
Optimized test at wafer-level					
DC testing in electrical – electrical domain					
Revised accept-reject methodology					

Table 3. Transition from custom to standardized procedures.

	2020	2025	2030	2035	2040
Standards instead of custom approaches					
Prioritize tests across full PIC value chain					
Testing metrics					
Relevance of a test					
Standardized test structures					

Table 4. Transfer of test data across the PIC value chain

	2020	2025	2030	2035	2040
Implementation in PDK **Improved design tools (EPDA)**					
Correlation of the test outcomes **Improved processes** **Identification of redundancies**					
Accessible scope – potential IP issues					

Table 5. Technology-agnostic testing

	2020	2025	2030	2035	2040
Across (currently) major technologies InP, SiPH SiN, Electro-Optic (EO) polymers					
Open for emerging platforms polymer, diamond, rare earth ion doped, three-dimensional (3D) PICs, SoC (high temperature)					
Hybrid integration photonic cross platform electronic-photonic chip level (EPICs) electronic-photonic PCB-chip					
Testing PICs with CMOS circuits/testing					

Table 6. Automation of test at wafer, bar, die, module and system level testing

	2020	2025	2030	2035	2040
Wafer - level					
Bar and die – level testing					
Standard test interfaces (layout templates)					
Technology agnostic					
Scalability					
On-chip self-diagnostics (Utilizing electrical-to-electrical testing)					

Table 7. Design for test

	2020	2025	2030	2035	2040
Test oriented layout templates					
Implementation in PDKs					
Test scripts for generic die testing					
Training of PIC designers					

Situation Analyses

A situation analysis of photonic testing is available **at this link**. It covers topics such as:

- Manufacturing processes
- General Test Equipment
- Critical Infrastructure Issues
- Technology Needs
- Prioritized Research Needs
- Prioritized Development and Implementation Needs
- Workforce Development

Gaps and Showstoppers

STANDARIZATION

Standardized testing metrics and procedures are essential for developing PIC markets further. Some specific killer applications (interconnects, automotive, sensors, etc.) are needed to accelerate standardization. Necessary test items depend on a particular application, and a specific application makes them clear. A large market opportunity provides a powerful incentive for PIC companies such as PIC device companies, PIC foundries, and PIC testing equipment companies.

Necessary test items should be standardized across the full PIC value chain. Testing designs and procedures are then standardized. The design tools for testing should be implemented in EPDA and PDK. Testing should be accurate and fast. On-chip self-diagnostics like that for EICs will be needed in the future.

PIC device engineers need to clarify testing equipment specifications (electrical and optical probes, functionalities, accuracy, speed, etc.). They should collaborate closely with PIC testing equipment engineers.

Standardization seems a difficult challenge in this field because it needs many people's efforts and some sufficiently attractive markets. If this challenge is achieved, we will be able to develop various kinds of PIC products with a minimum of effort.

PLATFORM-AGNOSTIC TESTING

The basic testing setup is common in a variety of PIC technologies (SiPh, InP, GaAs, SiN, polymer, etc.). Technology-agnostic testing is very important. The standardized testing equipment should be used for a variety of PIC testing with minor modifications. Various PIC companies should cooperate with each other across technical boundaries. The PIC devices are tested at a variety of sample shapes (wafer, bar and die). Sample-shape agnostic testing is also very important.

AUTOMATION

Fully automated PIC testing equipment is essential for developing PIC markets further. Mature EIC industry test practices should be emulated, and original PIC industry test practices should be developed. Various types of fully automated transceiver testing (OOK, PAM4, QPSK, 16-64QAM, etc.) will be needed. In addition, as co-packaged PIC and EIC devices ramp, the availability of a comprehensive PIC/EIC ATE based test solution will become critical.

HIGH SPEED (RF BANDWIDTH) TESTING

PIC testing equipment must measure both low-speed and high-speed properties. Fully automated high-speed electrical test (>10-100 Gbps) at wafer level is not easy. Adding to this the need to optically connect to the DUT via either a horizontal or vertical coupling approach, and the challenges become both risky and costly.

OPTICAL TESTING FOR MANUFACTURE

Contactless and non-destructive inline optical testing equipment with no particle pollution, which is acceptable for a PIC fab, will be needed. Inline optical testing can improve product yield.

USER SUPPORT

User-friendly GUIs and a variety of testing scripts are needed. PIC tests are generally difficult because electrical and photonic knowledge are needed. Helpful training manuals and courses are necessary.

ANALYSIS OF TESTING RESULTS

We have to research relationships between testing results at each level and product performance. A PIC accept-reject methodology should be established for each product. For example, one faulty sub-system does not necessarily disqualify functionality of the full circuit. In addition, statistics and analysis of testing data should effectively be transferred across the PIC value chain.

COST

Fully automated optical and electrical testing equipment will be very expensive. We should share expensive testing tools based on standardization and platform-agnostic testing. Testing time (including setup, calibration, wafer load and unload, etc.) should be short enough because time is money. But testing should be accurate enough.

We have to make the best use of testing results to achieve a good product yield and high product performance. The testing results should also be used to revise a product design and develop new products with much higher performance.

HIGHER PIC TECHNOLOGIES

Some specific applications help to solve the above problems. Higher PIC technologies are necessary to realize such applications. For example, low-loss propagation, low power consumption and high-speed optical modulation,

photo detection and amplification, high temperature stability, high r33 materials etc., which translate into high performance, will be expected in SiPh, InP, GaAs, SiN, polymer, etc.

- The 50 GHz barrier resulting from conventional CMOS capability forcing parallel solutions rather than higher baud rates.
- Low speed of suitable assembly, test and other process equipment resulting in high costs.
- Inability to overcome the cost-driving, rate-limiting step/bottleneck of manufacturing/testing such as the number of assembly steps or length of time to perform test, especially BER testing. Lengthy test times increase expense.
- Limits resulting from adapting existing equipment, materials and methods to optical test as more specific equipment is not available. Currently the demand for such specialized equipment is not sufficient to incentivize equipment manufacturers to make it available due to high non-recurring engineering (NRE) costs and low return on investment.
- Designing for Manufacturing and test:
  - Maximizing output to reduce cost
  - Studying designs to trade off accuracy and speed
- Inability to utilize materials or processes due to environment-related constraints (RoHS, REACH, WEEE, etc.)

Recommendations for Potential Alternative Technologies

1. Silicon waveguides to 1D/2D photonic crystal waveguides or plasmonic waveguides. Some devices become much smaller (leading to higher-density photonic integrated circuits).
2. Combinations of active and passive polymers for alternative Silicon (and other) PIC designs and automated test, calibration and verification procedures.
3. Utilize laser processing to make optical waveguides in-situ for effective optical connections and optical structures.
4. Utilization of plasmons to minimize size and maximize functionality.

Contributors

Tom Brown, University of Rochester - chair
Dave Armstrong, Advantest - chair
Sylwester Latkowski, Photonic Integration Technology Center, Eindhoven University of Technology - chair

Robert Pfahl, iNEMI
Dan Evans, Palomar Tech.
Bill Bottoms, 3MTS
Makoto Okano, AIST
Tobias Gnausch, Jenoptik
Jeroen de Coster, IMEC
John MacWilliams, Consultant
Chris Coleman, Keysight
Keren Bergman, Columbia University
Sam Salloum, Tektronix
Scottie Wyatt, Tektronix
Rocio Banos, VLC Photonics
Robert Polster, Columbia University
Ignazio Piancentini, Ficontec
Eugene Atwood, IBM
Yi Zhang, Teradyne
Fen Guan, Global Foundries
Graham Reed, University of Southampton
Richard Otte, Promex Industries Inc.
Lionel Kimerling, MIT
Martin Möhrle, Fraunhofer HHI
Michael Lebby, Lightwave Logic Inc.
Chris Roeloffzen, LioniX International
Willem Vos, University Twente
Jan Mink, VTEC
Gordonn Liu, Huawei
Jan Peters Weem, Tektronix
Iñigo Artundo, VLC Photonics
Michael Garner, Stanford University
Philip Schonfield, RIT
Shangjian Zhang, UEST
Zhihua Li, IME
Zoe Conroy, Cisco
Carl Buck, Aehr Test Systems
Tsuyoshi Horikawa, Photonics Electronics Technology Research Association (PETRA), Japan

We would like to thank all who contributed to the Test TWG, both in workshops, on-line conferences and with written comments.

Section 3: Logic Testing

Introduction

The use of heterogeneous integration to combine several chiplets into a multi-die package has more than offset the established slowing of Dennard scaling; the moniker of this being the "More-than-Moore Era" is quite apt. As measured purely by area, the amount of silicon in such a package can now far exceed that possible in a traditional monolithic package. For example, Intel's Ponte Vecchio package contains 47 chiplets with a total active silicon area of 2330 mm^2 [1] compared to the enormous monolithic Nvidia A100 GPU at 862 mm^2 [2]. As measured by logic complexity and the associated test requirements, a package containing this much silicon brings with it the challenge of testing for subtle defects in transistors and wires, but at the scale of what was a motherboard's-worth of functionality only a few years ago. This is in addition to the new test requirements associated with the 2.5D and 3D integration methods themselves. In total, the move to heterogenous integration has created a substantial increase in the number and difficulty of the tasks facing the DFT and test engineering communities. This section considers these tasks by grouping them into categories: access, yield, cost, quality, and time to market.

The first group of these new tasks involve basic access to on-chip test features, both at wafer sort, where the fine pitch of chiplet interconnects makes traditional probing problematic or impossible (see probe section of this Roadmap), and in the package, where only the package pins on the base die are accessible, through which all the other die must be tested. Besides these physical constraints, the bandwidth of the interface through which test data is exchanged with the device is another key consideration: test time and thus cost are directly affected. Furthermore, the emergence of a chiplet ecosystem where third-party providers can contribute silicon for package integrators to utilize is strongly dependent on standardized test interfaces which facilitate interoperability. A standard which should enable test access is IEEE 1838 which provides a method for describing, retargeting and distributing tests as well as physical interfaces for both data and control.

The second group of tasks revolves around yield. In heterogeneous integration, the cost of a test escape (i.e., a defective chiplet which nevertheless passes its (inadequate) wafer sort test) is no longer just the cost of that piece of silicon and the package; it includes the cost of all the other good chiplets as well, since reworking a package is considered to be impossible. This situation will likely drive two different responses. First, integrators may demand known good die from their silicon providers, which in turn will drive the test community to grapple with the cost, quality, yield maximization and die harvesting topics described next. Second, silicon providers and package integrators may collaborate on fault tolerant schemes such as repair and redundancy for yield recovery, some of which may even be used throughout the life cycle of the product to gracefully deal with degradation over time.

The third group of tasks around cost extends those mentioned in the yield category by also considering the cost of the test features and the production test flow. Internal test features (scan, memory and logic BIST, I/O loopback, on-chip instruments, repair, redundancy, etc.) greatly enhance the testability of a device, but come at the price of silicon area, functional performance, and power. Similarly, adding extra steps in the manufacturing test flow (screening at multiple operating points, performing partial-assembly testing, burn-in, system-level test, etc.) and applying adaptive test techniques (part average testing, good die in bad neighborhoods, outlier detection, etc.) can reduce the number of test escapes, but increase the cost of goods sold. Die-to-Die interfaces between chiplets also present cost challenges as they are expensive to probe with today's methods and coverage is provided at later test steps which can result in higher scrap cost (mitigated with repair and spare lanes). Finding the appropriate features and flows to support the financial models will require many trade-offs.

Quality has a strong bearing on cost and yield as described above, but takes on two other important roles in a heterogeneous integration environment. First, given that a single device may contain silicon from several fabrication facilities and go through a multi-stage assembly process, managing the value delivery chain will be extremely challenging unless each participant in it measures and delivers to very high-quality standards. Second, since the products which utilize these multi-die packages will initially be in high-end markets (e.g., hyper-scale data centers, supercomputers, automotive, etc.) where data integrity is crucial, the absolute level of quality is a key consideration.

Lastly, despite the rising complexity of the devices, levels of integration and the increasing challenges of manufacturing, Time to Market (TTM) is of paramount performance. The time for tests to be developed and qualified for release has not increased with respect to product development time. To ensure that chiplets continue to support TTM efficiency, tests will need to be developed as IP which is retargetable at the various levels of integration and further standardization of test delivery interfaces to ensure interoperability between multiple vendors.

These five groups are clearly intertwined: high quality requires excellent test coverage which often involves expensive test time but can be modulated with high-bandwidth test access and internal test features, but those come at the cost of extra silicon area which can reduce yield and raise both cost and the likelihood of defects (not to mention

the negative impact on mission-mode performance and power). Finding the optimal path through these More-than-Moore challenges will require solid engineering. As chiplets are integrated from multiple providers, collaboration on test approaches and coverage methods will be of increasing importance. The following sections address these topics in more detail to help address this engineering work.

Key take-aways in the sections that follow:

- Test content continues to grow with the number of transistors at the die level
- Chiplets will provide additional challenges to traditional logic test with mixing methodologies and approaches
- Quality levels will need to improve to support product economics, and new test methods will be required
- New test methods are emerging for deploying logic test
- Silent Data Corruption (SDC) is driving logic testing methods into deployed products
- Chiplet vendors will need to provide retargetable test IP for the next level of integration into SIP or system

Addressing the architectural bottlenecks of logic scan infrastructure

In previous versions of the HIR logic roadmap section, we have assumed that the fundamental approach for Logic ATPG would remain the same. As such, the roadmap focused on metrics such as scan data volume, data rate of interface, compression factor, and test time. In this version of the roadmap, we are highlighting the impact further integration has had, where the economics of test have driven a repartition of how scan is delivered to a Device Under Test (DUT) and how it is applied. To help delineate which challenges are classical logic scan challenges and which are changes to architecture, our discussion is broken into sections: Traditional scan challenges; emerging use-cases; updated scan architecture; and evolving logic test beyond scan testing.

Traditional scan challenges

With the progression of logic density, we continue to see the proportional growth of test data volume. As was noted in the 2021 roadmap, the effectiveness of compression at the block level is slowing. Other techniques of data compression for multiple instances of identical cores have shown increases at the chip level. If classical scan delivery methods are employed, then the scan frequency is also limited. In the emerging scan challenges, we will highlight new approaches that provide further improvements. The role of continued scan pattern growth is highlighted in Figure 1, which illustrates the resultant test time growth that explodes as multiple die are integrated into a single package ("SOC" in the figure). Further modeling will be done in the next roadmap update to capture the impacts of the trends discussed in this document.

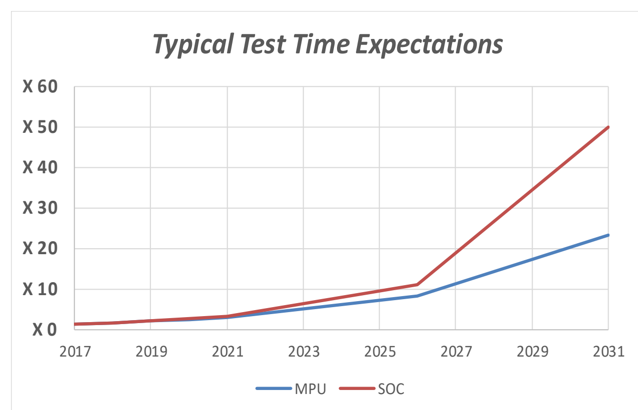


Figure 2: 2019 HIR prediction of test time growth

Though it is the easiest to model, scan-based testing is not the only driver of test time increases. Other test actions (BIST, functional test, parametric test, analog test, trimming, repair, volume diagnosis data collection, etc.) contribute as well, and have also been growing. The mix of these test types, along with the insertions (wafer sort, package test, system-level test) in which they are applied, factor into the calculation of overall test time. There is no industry consensus on what constitutes the optimal solution for maximizing quality while minimizing cost, and the specific implementations vary by market segment and by company. Given that reality, the remainder of this analysis will focus on scan-based testing.

Very few SoC sub-cores begin their development process without considering a scan compression architecture into which they will fit. If standard scan architectures are applied, it is typically for very small components or IPs which will later be embedded in larger blocks which will then include a compression architecture. Scan compression has succeeded in reducing test times for manufacturing test and reducing data storage and transport costs. Typical SoCs use a homogenous approach among all cores within a die. But this is not always the case. It is anticipated that heterogeneous integration of disparate die may also include compression schemes from various EDA vendors. Helping to support hierarchical integration and the notion of merging pattern formats from various sources is a common goal of core wrapping. Most SoCs (and therefore heterogeneous package assemblies) are composed of

wrapped cores. The patterns for these cores are developed at the core levels and retargeted (or ported) to the top level for eventual ATE application. Die stacking simply adds more hierarchical porting layers to the retargeting solution.

Practical issues users should consider when merging these many core-level pattern sets together into a manufacturing test pattern set include ATE resources, wafer or package-level access resources, test time, and power and thermal requirements and constraints. Test is a power-hungry application, and thermal issues are exacerbated by heterogeneous packaging. Solutions which integrate patterns for all these cores should consider topological proximity, and power and thermal responses when combining patterns for simultaneous application. Compression schemes have incorporated built-in power-reduction techniques for some time to help alleviate the shift switching activity profile for an individual compression codec. In addition, there are hardware resources one can add automatically to further reduce capture power or help ATPG easily reduce capture power. To help automate test scheduling of modules across a stack, more sophisticated power-related data may need to be introduced along with physical topological information to help test schedulers shorten test times while not overrunning power and thermal constraints.

Recently, test data propagation fabrics have emerged from EDA tooling to help address resource allocation issues in multi-core and multi-die packaging applications. Moving large amounts of test data long distances, or simply making use of various data types from circuits sprinkled across a vast surface area, has presented a problem not unlike functional compute and memory applications have always had. Again, heterogeneous packaging applications have only exacerbated the issue. Today, several "scan fabric" solutions are available. These might present a fixed-rate scan bus which adapts its bandwidth to the core endpoints as data moves from tester to core. Moreover, this interface might branch and maintain data speeds as fast as the intervening technology would allow, ramping down clock frequencies and adapting to core-level endpoint resource requirements as necessary.

In addition, there are solutions that seek to reduce ATE data requirements by leveraging the fact that many designs contain multiple identical cores. Of course, broadcasting a single set of stimuli to a group of cores reduces data volumes. But to help further reduce test data volumes, unique solutions exist which collapse response data to a minimum and reduce test times as well. For example, the response data can be scanned in and broadcast along with the stimulus. Each core then can determine its own correctness and store that, or scan out a composite result to help reduce data volumes. Or a MISR can be employed at the compressor outputs to further compress the resulting signature of a passing or failing test pattern or pattern set to a minimal amount of data. One can even initialize the MISR such that the resulting signature for a passing pattern set is zero (all 0 values) and this is easy to compare at the core level to compress the pass/fail result to a single-bit response at the end of the entire test.

Emerging Use-Cases

Several emerging use cases are further driving silicon sensor IP applications and DFT architectural decisions. In particular, re-use of DFT resources past the manufacturing stages and into the field have increased the value of these resources. Performance, safety, reliability, and debuggability applications have emerged as DFT IP and infrastructure has risen to address new functional challenges. Examples of these include solving adaptive voltage and frequency scaling applications, addressing the reliability crisis that silent data corruption (SDC) presents, leveraging system monitor IP to support debug operations in complex system environments, and using functional high-speed IO ports for in-system diagnosis scenarios.

Interestingly, the same IP that is used for in-system process, voltage, and temperature alarms and characterization can be used to support performance enhancements or reactions to measurements which exceed certain thresholds. For example, under a specific operational (software) load, a device could determine that there is headroom left for increasing processor speeds to address the running application. Additionally, one could use embedded monitors to determine that a device will soon fail catastrophically if not replaced due to path margin measurements on internal connections or between devices. Tester failures could be correlated with sensor data to aid the diagnosis process. And all of these resources could be accessed in-system during debugging operations. System debug availability is important. The ambient operational environment afforded by ATE is usually much cleaner and less stressful when compared to system applications. Functional high-speed IO can present a novel entry to solving these problems in-system, where and when they occur.

High-speed IO port use for supporting test and debug operations solves an interesting factory test application problem, as well. By leveraging a high-speed functional port or ports, getting data into and out of the device is no longer slowed by the limited availability of slow-speed pins on a package or die. Once the data is beyond the IO periphery, it can be expanded and slowed to frequencies more in line with the technologies and power constraints presented by each die in the package. When functional ports are leveraged for system-level debug, several considerations should be examined. First, the high-speed port type used has complexities of its own that may need

to be tested prior to use. Applying an IEEE Std 1149.10 protocol and architecture to this application may help alleviate some of the manufacturing test complexities associated with high-speed ports in the factory. However, the tester will also need to support the IEEE Std 1149.10 protocol. In addition, for use in the field, 1149.10 may also need to be leveraged by the attached debug environment. On the other hand, the functional architecture and protocol can also be used. Still, one must consider the manufacturing test environment and the field application context before locking in a solution set. Second, data and system security should always be considered. A holistic approach is required to make sure user data and device circuitry is protected from abuse by those wishing to steal that data or leverage those circuits for improper or illegal purposes.

Updated scan architectures

Test compression schemes introduced the first level of separation between external interfaces and scan chains. This helped increase the number of internal scan chains as well as reduce the scan chain length, thereby optimizing both test data volume and test application time. However, with heterogenous integration of multiple dies on a single package, ever-decreasing pin-to-gate ratios, and the dwindling number of available data pins (for example GPIOs), the ability to deliver scan data is a big challenge both for wafer and package-level manufacturing tests. To address the scan bandwidth issue, there are two things that need to be considered:

- Delivery of scan pay-load at a much higher speed via a small number of GPIOs or functional interfaces
- Distribution of scan-data within a die or across dies using a scan network/bus that can be operated at a much higher speed relative to the traditional scan rate

The delivery of scan data at a faster rate addresses the concern related to the volume of scan data that needs to be delivered using a narrow interface. The ability to deliver large amounts of scan data enables concurrent testing of hundreds of cores for large modern designs targeted for a wide-range of applications. The data organization at the interface can now be separated from the structure needed at the IP blocks. As such, the user can think of the scan data as pages of information or packets of information (note this is different than protocol packetization which includes encoding schemes). The packetization of scan data further helps in reducing the dependency on the number of IOs available for every codec within a design. This makes the tasks of test planning and test reuse much simpler, as any number of internal codec pins can be driven when delivering packetized scan data via a scan bus. A benefit of this architecture is that data payloads no longer require padding to balance scan chains, so memory can be used more efficiently by the test equipment.

When test compression was introduced, it relied on having a codec driving a large number of short chains within a core. It exploited the small number of specified bits needed to target faults in a design, and therefore, implemented lossless compression techniques by delivering the required information via a few scan channels. In the 2021 roadmap document, it was indicated that test compression ratios obtained via classical techniques will taper-off with increasing design complexity and improved ability for ATPG tools to pack more faults into a single pattern. Instead, compression will have to rely on a design trend with numerous identical cores, where ATPG tools (in addition to compressing test data) will have to re-use the same pattern set for identical cores within a design, thereby reducing scan data volume. Moving forward, with heterogenous integration of cores, packetized scan data delivery allows usage of data throttling (control the flow of data depending on cores that need the most) to manage integration of tests across multiple cores and pushing compression of test data even further. In other words, test compression improvements in the future will depend on a variety of techniques that are dependent on design characteristics and styles that go beyond just test data sparsity.

One of the characteristics of a modern design is the presence of hundreds of cores. Having a bus-based scan architecture allows delivery of scan data to hundreds (or thousands) of identical cores in parallel, and either observing the test responses or performing a local compare of the responses on-chip assuming the responses along with the masking data is also streamed to individual cores. This results in further improvement of test efficiency by reducing the test data that needs to be stored and improving the performance as data doesn't need to be read back and compared on an ATE.

Power dissipation during test has always been a major concern. With the ability to deliver the scan payload via a high-speed bus to many cores simultaneously, power dissipation becomes a bottleneck related to how many cores can be tested in parallel. This calls for localized generation of scan control signals such that one can perform independent shift and capture for each core in an asynchronous fashion. Asynchronous shift and capture between cores allow one to manage the voltage droop or IR drop that are usually associated with scan test in a much more efficient manner, thereby not only increasing the number of cores that can be tested in parallel but, in many cases, help in increasing the shift frequency.

For designs with hundreds of identical cores, broadcasting the stimuli, responses, and masking data to these cores reduces the volume of test data that need to be stored. However, there is a need for implementing efficient techniques to facilitate volume diagnosis. For example, when implementing on-chip compare for identical cores, one can determine the failing cores by inspecting a (failure flag) sticky bit at the end of test. Once the failing cores are identified, those cores can be targeted for re-testing and the failing responses can be observed to drive failure capture at ATE and diagnosis. In addition, there is ample opportunity during manufacturing test to optimize a test session based on how different tests can be scheduled and applied by considering test time requirements, as well as various environmental factors such as power, thermal gradients, power supplies, ATE throughput, etc. Additional factors that impact how tests are applied can also be related to failure data collection needs and limits. Based on the conditions in the DUT or the test needs, ATEs can play a significant role to modify and optimize the test sessions. It can drive data collection that would help modify and adapt the tests for subsequent test insertions. The diagnosis and power use-cases highlight that if tests are augmented with additional meta data, the ATE could provide further intelligence for execution which will also result in additional memory savings.

Evolving logic test beyond scan testing

As the cost of test escapes grows, it is important to try to move as much of the test content as far left in the manufacturing process as possible. Many complex devices still rely on some amount of functional testing or system-level tests to close the gap between what is testable though structural DFT techniques and mission mode. As the complexity of the chiplets of the system has grown, more of the design can be put into modes that more closely match mission-mode during test. This in enabled by having enough on-die memory so that tests can be executed internally in the chip. It also requires system hooks to support running without the external devices that would be seen in a full system. Some mission-mode capabilities such as power state and clock control can be quite challenging to shift to a production test environment.

The resurgence of functional testing has driven innovation in how tests are generated and deployed. One issue with functional tests is how effective a generated test is at detecting a fault, given a limited set of interfaces and a finite amount of time. Today, the use of functional tests is largely based on empirical experience of test escapes where symptoms of an undetected fault have impacted a software application running on the hardware. Manual effort identifies and transforms useful code snippets into functional tests; this is analogous to scan testing 30 years ago prior to the extensive automation of structural test. Extending such automation into the functional test domain will require the tools to create tests and measure their fault coverage to enable an efficient test suite for production testing. One promising technology is the Portable Stimulus Standard (PSS) which was proposed by the Accellera System Initiative. PSS takes a requirement definition, design model, and available interface descriptions for tools to generate tests that cover each requirement definition. This technology was developed for chip-level verification to prove that designs meet their operational requirements. The challenge for the test industry is to optimize the mapping of the fault space into the requirement space for coverage while also optimizing the test run time to make each test economical. These functional tests may benefit from another interesting technology called Quick Error Detect (QED, developed at Stanford University) which instruments functional tests using temporal and spatial duplication to speed detection and that can backtrack a detected error to a physical fault condition to guide how to precondition the hardware with minimal test time. These technologies will be required to make SLT testing more effective by limiting the time per test, making each test more effective, and enabling test coverage metrics.

In-System and SLT test requirements driving new logic test requirements

Testing of logic has been extending past the traditional factory test insertions of wafer and package to new areas that span from initial manufacturing throughout the device's operational lifetime. Previous heterogeneous integration roadmaps highlighted the rising use of functional testing with a System Level Test insertion as well as the use of MBIST and LBIST as part of the ISO26262 standards for periodic testing of electronic components in the field. What has gained more attention recently is the vulnerability of circuits to Silent Data Corruption (SDC) impacting complex digital devices in the data center. The risk of SDC is not new, but with the scale of modern data centers the occurrence of such errors has become measurable, and their detection, mitigation, and impact cost to the service provider has become an important topic. In 2022 at the International Test Conference, a major service provider stated that SDC events in a data center could affect as many as 1 in 1000 devices and manifest as applications producing incorrect results. However, there is not yet a consensus on how to measure SDC, nor is there a definitive breakdown of the root causes for these events. The industry sentiment is that we are only seeing the tip of the iceberg of this fault type, and new techniques will be required over the next five to ten years to drive down their rate of occurrence.

Historically, SDCs have been primarily thought of as a symptom of radiation-induced bit flips, and successfully mitigated accordingly. Today, there are multiple additional hypotheses about the possible causes of SDCs: 1) manufacturing defects that were not detected with traditional test flows; 2) latent defects that emerge due to aging effects; and 3) electrical effects (such as di/dt-induced voltage droops, IR drops, thermal gradients, etc.) caused by computational workloads which reduce design margins. The test industry is uniquely positioned to confirm or deny these hypotheses using techniques like extended characterization, root cause diagnosis, and in-situ monitoring. The best-known-method is still being explored and discussed and may well be a combination of approaches.

In the last decade we have seen an improvement in the physical realism of fault models by using the Cell Aware methodology, and this technique is expected to continue to evolve with emerging transistor technology (with related impacts reflected in the vector depth prediction of this section). Using superior fault models addresses the first SDC hypothesis by producing patterns that close the gaps from traditional methods that result in test escapes. It is important to note that, no matter how good the fault models, scan-based structural tests do not mimic the electrical conditions present during mission mode, so functional test will also play a role in catching test escapes. In addition to scan, functional techniques like PSS from Accellera, described in the last section, can be used to augment the test coverage.

Up to this point the discussion has focused on detectable faults at time zero; the degradation of circuits over time is the second hypothesis to consider as a cause of SDCs. Even if devices were all made perfectly, given the tiny size of the transistors as well as the stresses during use (temperature, voltage, current, mechanical, etc.), the way they operate over time will shift. For example, the resistivity of the power grid could increase over time due to thermal variations and current load. When the power grid changes, it will result in lower voltage delivered and the transistors will operate more slowly resulting in less margin. There are also well-known effects at the transistor level that will impact the design margin with respect to operation (NBTI, HCI, TDDB, etc.). One way to prevent SDCs from occurring may be understanding how the performance changes over time with respect to key performance parameters like timing margin, voltage, temperature, and device activity, then compensating for aging by adjusting the supply voltage or clock frequency accordingly. One challenge is how to implement such an in-situ control system to minimize the cost both in circuit area and impact to the end system.

One example from ISO26262 for automotive products involves the application of "key-on/key-off" tests which perform MBIST and LBIST in the field to re-validate the absence of faults before and after every use. The impact to the end system is defined by a required run time and the periodicity of testing, along with the higher-level architectural features to initiate the test and evaluate the results. It has also been noted that on-chip variation within a device has been increasing, so it is expected that the aging of each individual path will also become more important to measure. The solutions of the future must look at how the critical circuits or paths change over time and be monitored (ideally) while the system is running to measure the reduction of design margin over time.

These requirements are different from our traditional testing techniques that are focused on structural correctness, not operational impact. To better understand the root causes of aging, more sensors at the block level within a device will likely be deployed. These test methods will also need to comprehend how often to collect data, the data flow within the device, and driving measurements to actions within the final product. In some cases, this will be done on-device in the field for mission critical systems or in the cloud for fleet monitoring testing applications. As a result, we are presented with a new opportunity of where and when digital test is applied and how the outcome of testing will impact end-product operation. New features such as extending life with active voltage variation, predictive maintenance, or new repair methods at subsystem levels for compute elements are all within the realm of possibility when test features are made available in the field.

The third hypothesis about SDCs is that they arise when the dynamic effects of stressful workloads push the electrical environment on the chip past design margins. To maximize hardware performance in this era of post-Dennard scaling, aggressive design margining has become common – including the use of Dynamic Voltage and Frequency Scaling (DVFS) to alter operating conditions in real time based on the workload (generally to maximize performance per watt – which corresponds to minimizing excess margin). Furthermore, in modern digital design processes, detailed automation tooling accounts for various parameters like switching factors to predict reasonable device activity which leads to other design features like current estimates and power grid sizing to ensure that no excess margin is left on the table. However, to achieve the highest performance without risk of exceeding design margins, one must be able to understand the impact of the software running on the hardware which enables reducing guard bands and reaching the highest performance. In large multicore architectures, this leads to adjusting the scheduling of cores to ensure balanced activities across the chip with the best performance. To realize this, additional sensors must be deployed to characterize and monitor the impact of software running on the hardware to adjust operating point parameters over time (as DVFS uses to ensure correct operation with optimal energy use). In the

future, there will be the need for new data sources (extensions to voltage, temperature and timing margin) to enable further performance improvements.

ATE equipment challenges with the logic testing

Multiple trends are driving the test industry to develop new test methodologies which leverage high-speed IO (HSIO) to communicate data to the DUT in new ways. Most high-end devices (the ones which have the biggest test challenges) also have one or more high-speed protocol-based interfaces such as USB or PCIe available on them. Using this high-speed interface can provide two core values; 1) they provide a high-data bandwidth conduit for test, and 2) they provide a consistent test interface which can be used throughout the lifetime of the device.

Leveraging the existing HSIO interface provides an efficient way to enable many different types of tests, such as:

- Scan Test (including scan test networks)
- Functional Test
- Processor-enabled BIST
- On-chip instrument access (e.g., sensors): e.g., internal I/F to IJTAG
- MBIST and LBIST

Many traditional tests, such as scan and BIST, can be initiated over HSIO. Also, functional tests can be performed because they can be based on data payloads when the ATE interfaces are considered. The HSIO provides a fast way to load test setup information (such as arrays of coefficients) and test data sets (such as training sets) into the device for real-world confirmation of convergence and functionality. Additionally, functional tests can be executed between different cores on the die or between one chip and another in a heterogeneous integration situation, perhaps under the enablement of an on-chip processor.

The consistent HSIO test interface also allows leveraging test content from between test steps such as wafer, final, system level test, and in-situ testing after deployment. As such, it efficiently provides value through test consistency and reuse across many test insertions including end-of-life (RMA).

Enabling this type of testing, however, does require a new type of instrument in the ATE system. Key characteristics of this new ATE resource include:

- High-performance signal integrity
- The ability to enumerate the HSIO successfully, and if unsuccessful to diagnose the problem
- An industrial grade, integrated high-performance compute and software environment which mirrors the targeted real-world
- Very deep data storage array
- The ability to control the device JTAG interface
- The ability to do simple DC continuity testing

It is likely that many devices will retain, if possible, both a HSIO port for scan and functional test as well as traditional GPIO and JTAG interfaces in order to avoid the cost of additional test instrumentation at ATE-based wafer probe and package test insertions. The HSIO interface would be leveraged at system-level and in-situ test insertions where the other interfaces are not accessible.

A critical component of success, if an HSIO is used for scan and functional test, is that the interface adheres to a standardized protocol such as IEEE 1149.10 or standard PCIe. If the interface is based on some proprietary protocol, then it is difficult, if not impossible, to replicate that protocol on commercial test equipment due to implementation or IP protection difficulties.

It is assumed payload information is customized based on the DFT implementation and/or data security concerns. As noted above, this will drive the need for significant computational resources in the test equipment to construct and de-construct payload information in real time using custom software.

Lastly, it is critical that some DFT is available to validate the basic functionality of the high-speed interface prior to any other testing. Ideally, a self-test can be performed using scan and, if possible, an at-speed loopback test that utilizes an internal test path to eliminate the need for high-speed switching on the test fixture that would be needed for an external loopback.

Updated predictions of Test Metrics

Trend	Short term 0-5 years	Long term >5 years	Challenges
Scan pattern growth of >30%/year	Initial adoption of high-speed interfaces – USB and PCIe. Move towards packetized scan methods with test fabrics	High-speed serial interfaces carrying packetized scan data: more scan bandwidth Extending to D2D interfaces	Rate of adoption of new scan interfaces
Functional test resurgence	Beyond SLT, further adoption focused on portable stimulus	Functional test on ATE and SLT using software test libraries	Establishing coverage metrics
Demand for in-field testing growing due to functional safety	Re-use of DFT-based instruments at power-up e.g., MBIST and LBIST	BIST + software test libraries at power-up and on-line Safety critical requirements driving new functions	Integration of DFT-based BIST with mission mode control and reaction
Increasing IO interface challenges	Sacrificial pads and dedicated DFT interfaces	ATE infrastructure to contact advance interfaces	Electrical, optical, and mechanical interface sensitivities
Logic testing extending into field	Initial methods to describe aging and workload impacts to hardware	Provide coverage methods for Aging and SDCs	Impact and root cause of SDC and aging continues to evolve

Conclusions

It is an exciting time for logic test. In this section we have highlighted the challenges and directions of logic testing. We have shown that the classical challenges of increasing logic density are still driving the need for increased testing. Given the volume of data and emerging fault types, we also discuss new methods for test delivery as well as expansion of where logic test will occur. Heterogeneous integration will provide further product economic pressure to accelerate solutions for the challenges outlined above. Many of the emerging use case solutions are starting to be addressed with initial solutions and we will evaluate in the next roadmap the adoption rate and impact to the test economics.

References

[1] https://www.techpowerup.com/292250/intel-details-ponte-vecchio-accelerator-63-tiles-600-watt-tdp-and-lots-of-bandwidth
[2] https://developer.nvidia.com/blog/nvidia-ampere-architecture-in-depth

Section 4: Specialty Device Testing

A classification of *specialty devices* was defined in industry roadmaps beginning in 2006, driven by strong high-volume market demand, but having odd test requirements. Examples are CMOS image sensors, LCD drivers, MEMS devices (including multimode sensors), actuators, bio-MEMS, and similar non-standard devices.

Trends Impacting this Technology Area

The novel applications of mobile personal devices, IoT, healthcare/artificial organ, automotive/ADAS, smart industry, and emerging energy fields are key drivers of specialty devices where innovative testing technologies are needed to enable future processes such as 3D , chiplets, and heterogeneous integration with high yield during mass production.

The trends for technologies (Near Term < 5 years)

- The trends for multi-mode MEMS sensors are toward fusing multiple sensing functionalities together in one device with artificial intelligence processors.
- The technology trends for image sensors lead to highly integrated multiple wafers using a 3DS (three dimensional stacking) process with Cu-Cu (copper to copper) connection technology for directly connecting pixel chips and logic circuit chips. Cu-Cu connection does not require a specialized area for connecting pixel chips and logic circuit chips, as needed for conventional TSV connections. The first successful implementation of 3DS wafer processing of an image sensor was the BSI (Back Side Illumination) process which bonded a photo-sensor wafer together with a back-side mixed-signal data processing wafer. The next step in the image-sensor wafer-integration process adds a memory-cell wafer between the photo sensor wafer and mix-signal data processing wafer, which could enhance image performance and the speed of data processing in a variety of imaging applications such as 3D imaging, face recognition, and image capture, with frame rates over 1000 frames/second.
- The trends in new WLP (wafer level packaging) for image sensors are WLO (Wafer Level Optics) and WLCM (Wafer Level Camera Module), which stack optical systems on the image sensor wafer using a wafer-level packaging process to reduce the size of optical systems and increase efficiency of mass production.

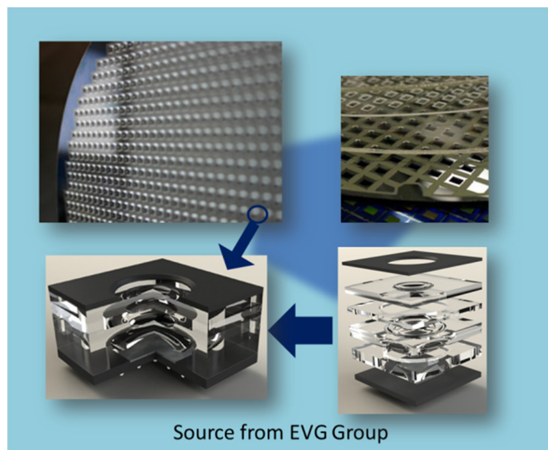


Figure 1. Image sensor WLO (Wafer Level Optics) packaging

The trends for technologies (Middle to Long Term < 15 years)

The automotive, robotic, medical and intelligent artificial organ fields are next-wave drivers of specialty devices which impact technologies in the medium and long term:

- Reliability will become critical for specialty devices. Burn-in and tri-temperature testing will become necessary test procedures during mass production.
- Built-in self-diagnostics, self-calibration and compensation, and self-repair technologies will become important design skills to apply to specialty devices for enhancing reliability performance.

Concerns: Test Challenges

LCD display drivers:

LCD display drivers are unique because of die form factor, which can have larger then 10:1 aspect ratio and thousands of very narrow gold bump pads requiring contact for test. In 2022, in-line and stager probing pad width for LCD display drivers already was down to 11µm in production and 8µm in development. Right now, only the cantilever probe card provides a major cost-effective solution for achieving probing of the LCD driver with such narrow and fine pitch pads with gold bumps in mass production.

An upcoming test challenge is that the data transfer speed for I/O will increase to 2.5 Gbps and is predicted to be up to 6.5 Gbps within 10 years. We need to overcome the challenge of probing fine-pitch bumping pads with high-speed signals with economical probing solutions.

Image sensor devices:

Testing of image sensor devices needs to consider special test requirements for optical systems and the resulting massive image data processing. Special requirements for optical test systems will be different and be relative to applications.

Table 1. Special specifications for optical test systems and applications

Illuminator Specification \ Application		Industry	Automotive	Consumer	Mobile
Wavelength Range	UV (100~400 nm)	V			
	Visible light (400nm~780nm)	V	V	V	V
	NIR (780~1400nm)	V	V		
	SWIR (1400~3000nm)	V			
High Intensity	> 10,000 Lux		V	V	
High Resolution	< 0.1 Lux		V	V	V
Polarized Light	0~360°	V			
Laser	PWM (Pulse Width Modulation)	V	V		
LED	LFM (LED Flicker Mitigation)		V		

Automotive ADAS applications and intelligent machine vision need the functionalities of image sensors with wide spectrum (from UV to FIR). high dynamic range and good S/N (Signal to Noise) ratio, fast data frame rate, and better quality and reliability, which challenges test system design. The burn-in solutions also need to include optical stress for sorting out defects in the coating process on photo sensor surfaces.

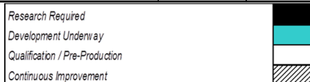
MEMS devices (Sensor, Actuator and Biological)

MEMS were successfully applied on various sensors for sensing motion, magnetic field, optic, sound, air pressure and vibration, flow, chemical composition of air, DNA sequencing, and other characteristics, and the market volume is increasing rapidly due to IoT, healthcare and automotive applications. Testing MEMS sensor devices with suitable physical stimulus and cost-effective solutions for the various types of sensors is difficult and tricky. Testing the expanding kinds of fusion sensors will bring many test challenges.

Table 2: Specialty Device Odd test potential solution for a MEMS Fusion Sensor

Table TST20-Speciality Device odd test potential solution

	Process integration		Test Method	Challenges	2022	2023	2024	2025	2026	2031	2036
MEMS Fusion Sensor	IMU sensor (Accelerometer + Gyro)	CP / wafer probe	Probing MEMS wafer (DC only)	Probe card technology							
			Full functions (Multi-insertion)	Motion Prober system							
			3DS wafer, full functions (Single-insertion)	DFT design and implement							
		WLP	Test after dicing (Wafer form)	DFT design and implement							
			Test after singular (Package form)	Handling small size package							
		FT	Full functions (Multi- insertion)	Test cost is high							
			Full functions (Single- insertion)	Reduce test coverage rate							
				DFT design and implement							
			Burn In Test	SLT							
				BISX (Build-In-Self Test, Diagnostic, Correlation, Compensation/Repair)							
	Navigation (G-sensor+ Gyro+ Magnetic sensor + Barometer)	FT	Full functions (Multi- insertion)	Test cost is high							
			Full functions (Single- insertion)	Reduce test coverage rate							
				DFT design and implement							
			Burn In Test	SLT							
				BISX (Build-In-Self Test, Diagnostic, Correlation, Compensation/Repair)							
	Environmental Sensor (Pressure + Humidity + Gas) Sensor	FT	Full functions (Multi- insertion)	Test cost is high							
			Full functions (Single- insertion)	Reduce test coverage rate							
				DFT design and implement							
			Burn In Test	SLT							
				BISX (Build-In-Self Test, Diagnostic, Correlation, Compensation/Repair)							



DFT for MEMS sensor devices is new technology and needs research and innovative development for different kinds of sensor structure. MEMS sensors DFT needs to develop the stimulus source and sensor together in the MEMS structure as a BIST (Build-In-Self-Test) cell. When testing, the cloned control signal of physical stimulus is generated from the MEMS ASIC to enable the MEMS BIST cell to imitate physical stimulus for testing the sensor cell to achieve the DFT goals. This concept could also implement the technologies of BISD (Build-In-Self-Diagnostic), BSIC (Build-In-Self-Correlation/Compensation) and BSIR (Build-In-Self-Repair) to enhance reliability of MEMS sensors for automotive and medical applications. The key during testing is to make sure this BIST cell works well.

Beyond MEMS sensors, there are also actuator and biological applications such as micro-mirrors, MEMS speakers, RF switches, energy harvesting, microfluidics, micro-dispenser and artificial organs, plus others. The testing challenges for testing MEMS actuators and biological devices are that test methods are hard to standardize and depend on the structure for each different kind of MEMS device. Especially for testing biological devices, the test environment can be severe and there is a need to pass safety certification based on the laws for different grades and countries.

Summary

Specialty devices as defined have odd test requirements and are driven by strong high-volume market demand. Under these two conditions, the trends for specialty devices will be driven toward highly integrated multi-functions in one smaller unit to overcome ASP (Average Sale Price) erosion, and testing procedures will move toward high parallelism to reduce test cost. Test challenges will follow the same trends for heterogeneous integration to address testing for specialty products though cost-effective solutions.

Team Leader: Wendy Chen

References:

1. “EVG’s wafer-level optics (WLO) manufacturing solutions”, EVG press release, September 11 ,2017

Section 5: Memory Test

Summary:

- Memory is a growing segment within the semiconductor industry (~30% in 2021 up from ~10% in 2000).
- Higher bit density drives increased interface speed, power, and thermal management requirements.
- Smaller physical geometries challenge electro-mechanical interface capability of wafer and component test.
- NAND densities are projected to grow into >8Tb/die by ~2024, driven by continued growth in vertical scaling.
- DRAM bandwidth and densities are growing to meet growing demands of CPU/GPU applications.

From 2020 to 2021 worldwide semiconductor market revenue increased 26.3%, with Memory comprising ~29% of overall production, and growth in the Memory and Storage segments increased from ~10% of overall semiconductor revenue in 2000 to ~30% in 2021[1]. End applications for the primary segments of the memory space (DRAM and NAND) have shifted slightly over the last few years, with changes in the DRAM segment seeing a flattening in demand for Mobile and PC applications, and an increase in demand for Datacenter applications, and with the NAND segment realizing the largest growth in the SSD segment for both Enterprise and Client applications[2,3]. NOR FLASH remains stable but becoming less relevant as NAND and DRAM growth continues.

As the demand for Memory applications has continued to grow and evolve over the last several years, the associated bit output has also grown due to innovations in architecture and technology that scale the density at a faster pace than package unit output (Chart1)[2,3].

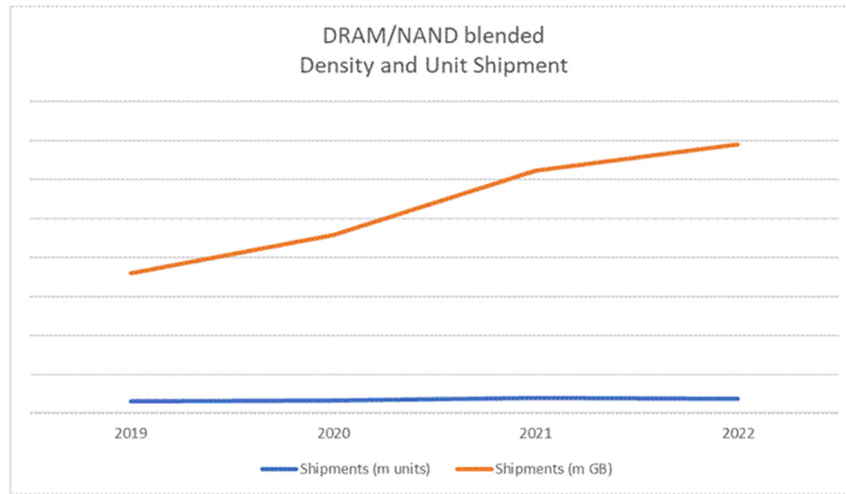


Chart 1

Effective use of this increased density relies on higher interface speeds (UFS, PCIe, PAM) to access the data. The scale of these increased speeds for both DRAM[4] and NAND[5] (Chart 2)[6] begins driving additional power and thermal management requirements.

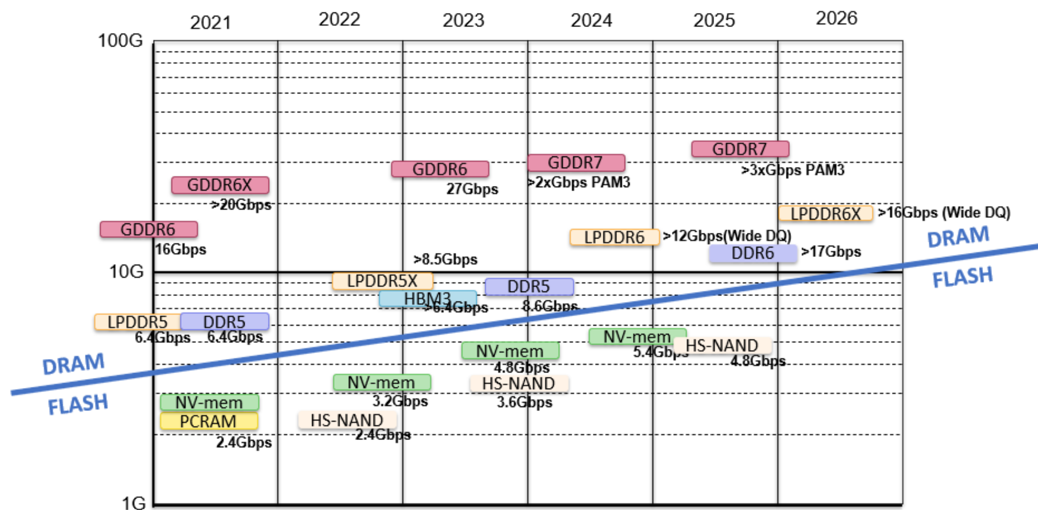


Chart 2

From a Memory Test perspective, as the increases in bit output, interface speed, power, and thermal management requirements scale in both NAND and DRAM, challenges arise to meet the intersection of capability. Die sizes continue to shrink either through geometry or integrated scaling, resulting in higher Die Per Wafer at increased device density and speed. These shrinking die sizes create challenges at wafer test in terms of interface constraints – in many cases, the number of die that can be tested must be reduced in order to route signals, and to enable contact to the wafer. The interface pad size and pitch are also projected to shrink below 50um in size, and the pitch of the pads creates challenges in signal routing, power delivery, and in some cases touching the pads has impact to the bondable pad area used for device assembly. As these key contact interface features scale smaller, and expansions in thermal demands grow to include coverage from -40C to 125C for automotive needs, wafer test interface thermal scaling must be proactively managed to ensure effective test coverage. From a power/thermal management perspective, with more power being delivered to smaller devices through the required range of test temperatures, proactive power dissipation at the device level also becomes a critical concern. For example, the overall growth from 2017 to 2022 shows a 5-year trend of ~9% reduction in voltage, but an increase of ~550% Die Per Wafer (DPW), and ~450% increase of power dissipation requirements at wafer test (Table 1)[6].

Table 1

	~2017	2022	Scaling
device voltage	1.2V	1.1V	-9%
dies per wafer (DPW)	~500	2500-3000	500-600%
test equipment device power dissipation (wafer prober)	*100-150W/wafer*	*500-600W/ wafer*	*400-500%*
speculative trend			DPW scaling faster than test equipment power dissipation despite lower device voltages.

As the die size shrinks with higher interface speeds, Signal Integrity (SI) and Power Integrity (PI) become more challenging because the signals become more tightly arranged with smaller interfaces. Similar to challenges faced at wafer test with smaller pads, packaged die are also facing scaling issues, including BGA interfaces that are shrinking below 125um balls and less than 250um pitch. Added challenges include decreasing solderball heights (<100um), thinner packages (<500um), and increasing contact points on the interface which drive issues related to contact, thermal management, power dissipation, and handler drive force to optimally scale interfaces to the desired parallelism.

As bandwidth requirements increase, higher speeds and new interface technologies are emerging (e.g. PAM3, PAM4, wide I/O), and there is often a lack of agreement at standards consortia until very late. This challenges the development of tester technology to meet the evolving device interfaces in terms of technical risk, schedule, and cost.

For all test insertions, as the device density grows, more and more bits are required to stream from each device back to the tester for processing and analysis, potentially driving changes in tester architecture and IT infrastructure to manage growing bandwidth considerations.

NAND

Key NAND applications today include enterprise data and edge compute centers, the ADAS automotive cloud, plus local storage, gaming, and 5G applications. Data creation in these key spaces in 2022 hits a remarkable 100ZB, and is projected to grow to 200ZB by 2026, with a resulting 32% CAGR[7]. NAND device bit density growth from 2015 to 2021 grew from 64Gb/die to 512Gb/die[8], roughly doubling every 2 years, resulting in a 10-20% growth of associated test time every year[9]. Future bit growth is achieved in the transition from today's 2xx layers at ~2Tb/die, into projected >300 layers by 2024 resulting in ~8Tb/die[10]. This density increase is achieved through vertical scaling with thinner layers, lateral scaling with higher density layer interconnect, architecture scaling moving from CNA to Multi-bond, and in logical scaling moving from SLC to PLC. This bit growth in NAND is particularly challenging, as the industry has been testing all die at a wafer level in a single touchdown for the last 10-15 years[9], and with no simple way to scale interface parallelism, there is significant growth in demand for testers to meet Si output. Device speed performance is also increasing to move data from the device to the outside world. Asynchronous random reads enable faster bit access, and SSD interface speed growth is moving from 2.4Gbps to 3.2Gbps to 4.8Gbps[10] to improve bandwidth and reduce latency. Interface standards compliment the trend as they move (for example) from PCIe G4 to G5 to G6, and the addition of high-speed SERDES interface memory controllers such as UFS 4.0 23 Gbps and PCIe G5 32Gbps provide the necessary support to double interface speeds about every 4 years[11,12].

In next-generation interconnect and speed, CXL3.0 is driving towards the next hyperscale applications. The CXL fabric architecture is intended to solve cost and bandwidth issues that DRAM-only solutions cannot address, all at a projected 64GT/s with no added latency above CXL2.0[13]. This adds further complexity in signaling and throughput from a Memory Test perspective, as many traditional ATE interfaces are architected for adaptable re-use, and not architected for high bandwidth applications.

DRAM

PC DRAM transitions are beginning to occur from mainstream DDR4 to DDR5, largely in an effort to increase effective bandwidth to CPU cores (Chart 3)[14,15]. As this transition occurs, the increase in data volume and speed will result in some key Test challenges both at the wafer and package level. Challenges include: Higher power to service the increase in bandwidth; delivery of power and signal to the device with sufficient fidelity to achieve the higher

speeds (while device size shrinks as noted above, which will challenge interface routing and development); and thermal management of the device at wafer and package level to appropriately dissipate and control device heating.

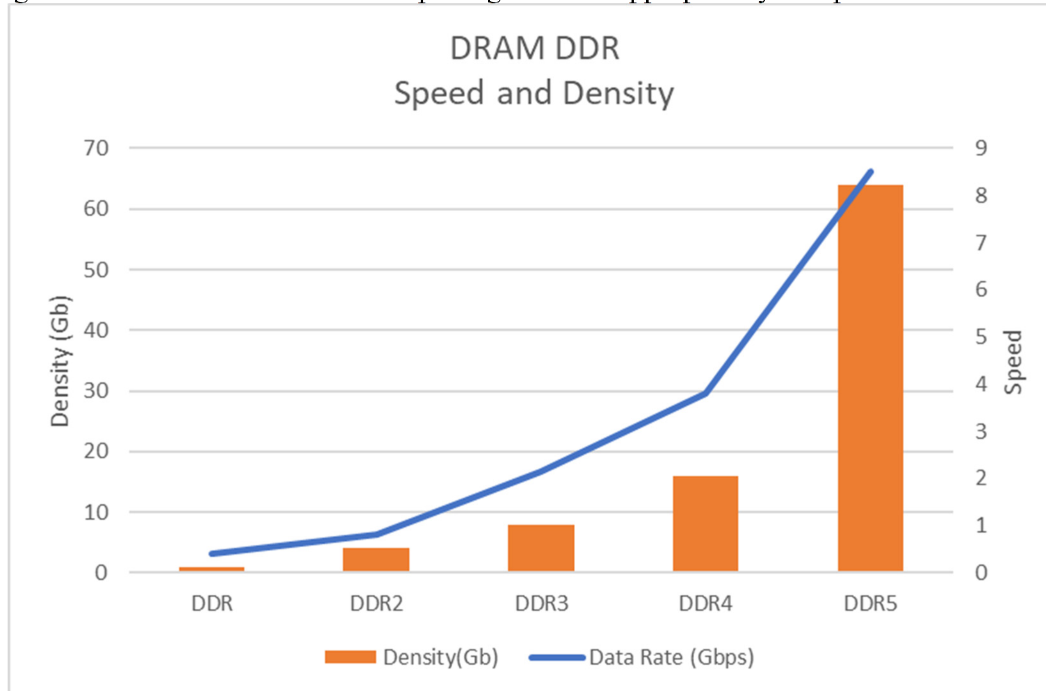


Chart 3

High Bandwidth Memory (HBM) is also growing in application for near-processor applications to improve graphics and AI applications; and GDDRx speeds are continuing to increase speed to accelerate graphics performance – GDDR6 at 24Gbps[16] is available today, with continued speed increases expected. All these advances improve the speed and ability of users/systems to effectively access data with decreased latency. These advances will further challenge speed, power, and thermal management in similar ways as observed in the DDR transitions noted above.

References:

1) Micron, "Memory Coalition of Excellence"
2) Yole Group "Status of The Memory Industry Report 2022"
3) Forward Insights "NAND Quarterly Insights Q3/22"
4) JEDEC JESD79-5B "DDR5 SDRAM"
5) JEDEC JES220F "UNIVERSAL FLASH STORAGE, Version 4.0"
6) Advantest Summary of Industry trends
7) Jim Elliott, Samsung Keynote, Flash Memory Summit 2022
8) Yole Group "NAND Market Monitor Q4 2022"
9) Teradyne analysis of broad industry trends
10) Scott Nelson, Kioxia Keynote, Flash Memory Summit 2022
11) Wikipedia, "Universal Flash Storage, Version Comparison"
12) Wikipedia, "PCI Express, History and revisions"
13) Compute Express Link "CXL 3.0 Press Release"
14) Synopsis, "DDR5/4/3/2: How Memory Density and Speed Increased with each Generation of DDR"
15) JEDEC, "JEDEC Publishes New DDR5 Standard for Advancing Next-Generation High Performance Computing Systems"
16) Samsung, "Samsung Electronics Launches Industry's First 24Gbps GDDR6 DRAM to Power Next-Generation High-End Graphics Cards"

Contributors:

Contributor	Company Affiliation	Contributor	Company Affiliation
Anthony Lum	Advantest America Inc	Phil Byrd	Micron Technology
Vineet Pancholi	Amkor	Jerry McBride	Micron Technology
Paul Okino	Teradyne		

Section 6: Analog and Mixed Signal Test

Executive Summary

The economic benefit of monolithic integration (SoC) and system in package (SiP) is well established and continues. This integration has combined digital logic with processing, analog, power management, and mixed signal 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.

An 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), also known as Multi-Site Efficiency (MSE), high; this is of increasing importance to avoid severely impacting Units Per Hour (UPH).

A similar concept but in a dimension relating to the single device itself is testing multiple IP cores within the device in parallel (concurrent test). This has many of the requirements and challenges of parallel test, but also includes some unique ones. A key one is having the ability in the design of the IC to test IP cores independently, in parallel. Test Access Mechanisms (TAMs) are the ability of IP cores to be accessed and controlled independently from other IP cores. The most powerful economic advantage results when being able to test multiple IP cores in parallel, while at the same time testing multiple devices in parallel.

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 typically non-deterministic and must be post 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. The requirements for the TAM are covered in the DFT SOC Device Testing section. The requirements for DC trim accuracy are included in the Mixed Signal tables (see Table 1).

DC Accuracy updates for 2020

The 2020 update for DC accuracy includes ever-increasing low-end accuracy requirements driven by lower VDD values and more fuse blowing and servo techniques being used to cost effectively make the DUT more accurate and improve the specifications and yields.

COT is always important and more parallelism in terms of IP blocks within a device (IP block) and multi-site parallelism is key to this.

Quality also needs to be improved with these accuracy improvements. Pre and post inline checking and the comparison of lot runs looking for common tests that always pass or fail will be aided by using Artificial Intelligence (AI) and Machine Learning (ML) to handle and simplify large volumes of data. Other quality improvements include inventorying the tests that have been run and having more quantitative (actual value) versus qualitative (pass/fail) testing. There is always a cost trade-off balance.

Power updates for 2020

The other end of the spectrum for 2020 is high power (current and voltages) being driven primarily by server farm power needs and automotive and battery management systems as shown in table 1 in the Note 8 section.

Because of the higher power, some tests that run a device at full power must be run very quickly and then turned off so as not to damage the parts that require special cooling. In these cases, precision pulses are required on tests like RDSon which pulses a high current at a very short pulse width to test the on-resistance of a switch.

Quality improvements here would include thermal testing and management throughout the test flow. For example, high power tests which would generate a lot of heat could be interleaved with low power test to allow the device to cool down.

Handlers with built-in cooling for the device is another option to be looked at for devices requiring the cooling.

Some process technologies once considered niche are gaining mainstream acceptance, including GaN (gallium nitride) and SiC (silicon carbide) devices.

SiC is projected to hit $1.5B by 2023 for these types of applications[14]:

- Electric Vehicle
- Train
- Charging Infrastructure
- Motor Drivers
- Photo Voltaic (PV)
- Wind Power

GaN is projected to hit $500M by 2023 for these types of applications[15]:

- Data Centers
- Fast Charger
- LiDar
- Wireless Charging
- Electric Vehicle

Power devices using GaN and SiC have higher band gaps compared to their silicon counterparts. The benefits are[16].

- Higher power density
- Smaller size (smaller wafer & die)
- Better high temperature performance because their band gap is higher than silicon
- Higher frequency response
- Lower ON-resistance
- Lower leakage, so there is a need for sourcing higher test voltages, as well as appropriate low current measurement sensitivity.

The test requirements to test GaN and SiC devices are

- Breakdown voltages up to 3000 V or even higher
- More than 100 A
- Junction capacitances for dc biases up to 3000 V
- High SiC and GaN voltages and fast switching speeds
- Testing these devices at their specified voltage, current and power rating
- Test fixturing:
    - A proper test fixture solution is extremely important to ensure safety (due to the high voltages and currents used)
- Supporting the wide variety of power device package types.

The breakdown voltage test has special techniques being investigated involving Paschen's Law. To summarize: above a certain pressure, increasing the pressure raises the breakdown voltage or allows a narrower gap without breakdown at a set voltage.

Analog Mixed Signal Updates for 2020

Pulse Amplitude Modulation – 4 levels (PAM4) (Note: Optical PAM4 is not addressed in this update)

The attributes of PAM4 include:

- 4 amplitude levels
- 2 bits of information in every symbol: ~ 2x throughput for the same Baud rate, ie, 28 GBaud PAM4 = 56 Gb/s
- Lower SNR, more susceptible to noise
- More complex Tx/Rx design, higher cost

It is used extensively in the JESD 204B/C standard.

The transmitter (Tx) can be measured with high-speed digitizers, samplers, digital oscilloscopes or even a digital comparator. The receiver (Rx) signal is generated by RF DACs. RF design rules come into play at these high frequencies.

[14] https://www.systemplus.fr/wp-content/uploads/2018/07/YD18027_Power_SiC_2018_Materials_Devices_Applications_July2018_Yole_Sample-1.pdf
[15] https://compoundsemiconductor.net/article/106038/Would_Apple_Change_The_Power_GaN_World%7BfeatureExtra%7D
[16] https://www.powerelectronics.com/technologies/power-electronics-systems/article/21860727/testing-gan-and-sic-devices-faqs

DSP is required to get an optimal eye opening which entails equalization for both PRE and POST processing. PRE processing is used to clean up the stimulus to the Rx, and POST processing is used to clean up the measured data from Tx. Amplitude accuracy is important because of the 4-level algorithm of PAM4. High-accuracy timing and low jitter are important to get a good eye opening.

Challenges in Analyzing PAM4 signals include:

- Sampling Point: Finite rise times and different transition amplitudes create inherent ISI and make clock recovery more difficult (TransImpedance Amplifiers have CDR integrated into them).
- Quantization error plays a role when you take PAM4 measurements versus NRZ. Transition times of the PAM4 data signal can create significant horizontal eye closure due to the higher transition density.
- Noise Tolerance: Instead of having the full amplitude range, there is only 33% of the amplitude because the voltage range is divided into four levels (refer to the Figure 1). Lower PAM4 insertion loss compensates for the 9.5-dB loss in SNR because the eye height for PAM4 is 1/3 of the eye height for NRZ, SNR loss = 20* log10 (1/3) = ~9.5 dB. When other non-linearity is included, it is approximately 11 dB.
- Non-Linear Eyes: The system-margin bottleneck lies with the worst eye. Nonlinearity starts right at the Tx output, and is composed of RLM loss + SNDR loss + other losses like SNDR (ISI).
- Clock Recovery is used on the Rx side to minimize low frequency jitter.
- Fixturing – getting the signal to the DUT
  - Integrated resources are difficult to design at these speeds but are sometimes easier to fixture. External Boxes are available but then are more effort and expense to route to the device. Line loss and jitter are a challenge.

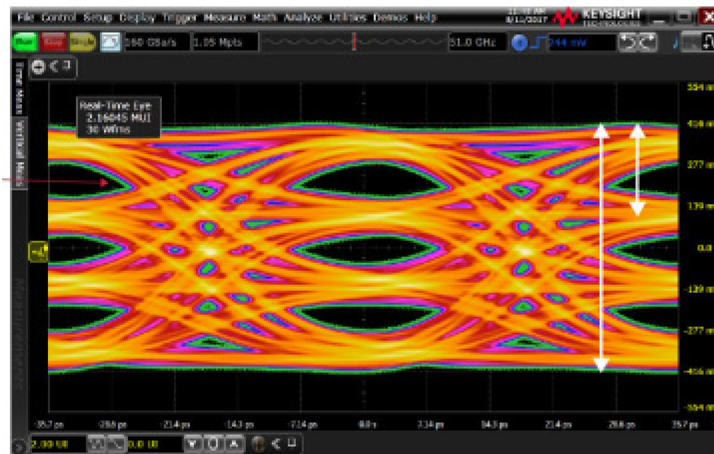


Figure 1. Scope Capture of PAM4 Signal

A typical test list for PAM4 looks like this:

Tx using a PRBS13 waveform:

- Output waveform
- Level Separation Mismatch Ratio
- Eye Symmetry
- Eye Height (amplitude) and Width (timing)
- Transition Time
- Signal-to-noise-and-distortion ratio (SNDR)
- Output Jitter
  - Jrms
  - Even-Odd Jitter (EOJ)
- Spacing of the PAM4 levels
- Eye Linearity: ratio of min to max PAM4 eye amplitudes as shown in Figure 2
  - Eye linearity = min(AVupp, AVmid, AVlow) / max(AVupp, AVmid, AVlow)

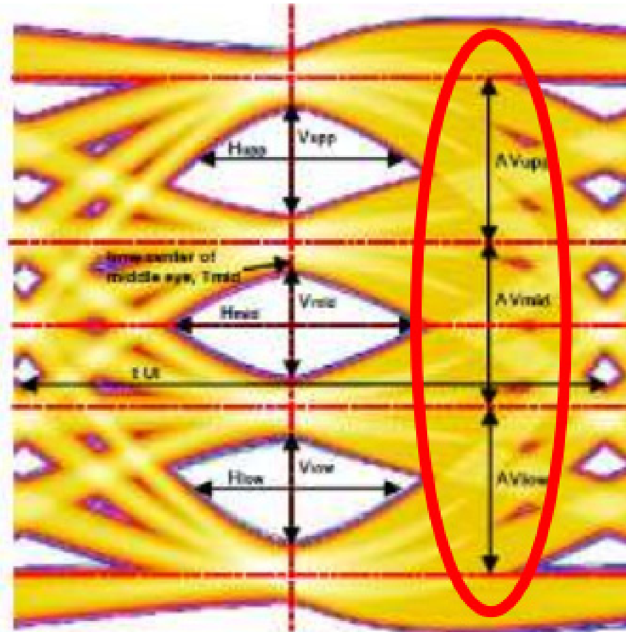


Figure 2: Eye Linearity

Rx tests

These tests involve how much distortion and jitter can be placed on the incoming signal to the receiver and still "read" the correct data stream.

- Jitter tolerance defined as how much jitter the receiver can tolerate
- Other potential receiver "stress tests"
  - Eye Skew (Timing)
  - Eye non-linearity (Amplitude between levels)

Key Test Trends

Short-Term Trends (< 5 Years)

There are three important trends. The first is to deliver adequate quality of test. Most analog/mixed-signal testing is done through performance-based testing. This includes functional testing of the device and then analyzing the quality of the output(s). 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. See the Mixed Signal Test tables (Table 1) for updates and future needs.

The second key trend is the need for higher DC accuracy. Many of the converters and precision references are made more accurate by doing a measure and trim step. The trim can be accomplished through several means; one of the more recent and cost-effective ways is through register programming of the device. The trim takes a relatively lower performance device and adds high accuracy to it through a DC test and register programming. In the past, this was done for medium performance devices, but now the test methodology has matured, and it is being applied to high accuracy/resolution devices. The change is that in this class of devices, much higher DC accuracy is required to make a valid test.

The third key trend is to enable the economics of test through instrumentation density and Parallel Test Efficiency (PTE). The level of parallelism requires an increase in instrumentation density.

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

Analog/mixed-signal DFT and BIST techniques continue to lag. No proven alternative to performance-based analog testing has been widely adopted 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.

The Ethernet trends are continuing into higher speeds – 28, 40 Gbps per channel and even beyond.[1] There continues to be the need for backwards compatibility to the many existing digital communication standards.

Table 1: Mixed-signal and DC Test Requirements

	2020	2021	2026	2031
Low Frequency Waveform [Note 1]				
SFDR	145	145	145	145
SNR	120	120	120	120
THD	140	140	140	140
BW-Minimum (kHz)	50	50	50	50
BW-Maximum (kHz) [Note 2]	500	500	500	500
High Frequency Waveform Source / Measure [Note 3]				
Level V (pk–pk)	<4	<4	<2.5	<2.5
BW (MHz)	250	250	500	500
Sample rate (MS/s) [Note 5]	500	500	1000	1000
Resolution (bits) AWG/Sine	16	16	18	18
Noise floor (dB/RT Hz)	-140	-140	-150	-150
Very High Frequency Waveform Source / Measure [Note 4]				
Level V (pk–pk)	<4	<4	<4	<4
Accuracy (±)	0.50%	0.50%	0.50%	0.50%
Measure BW (GHz) (under sampled)	9.6	9.6	15	15
Capture Depth Mwords	4	4	4	4
Min resolution (bits)	8-10	8-10	8-10	8-10
DC Accuracy (Note 6)				
DC force (uV)	50	50	50	50
DC measure (uV)	50	50	50	50
DC force (nA) (Note 7)	5	5	1	1
DC measure (nA) (Note 7)	5	5	1	1
DC Power (Note 8)				
DC force V Constant	120	120	140	140
DC measure V Constant	120	120	140	140
DC force A Constant	80	80	100	100
DC measure A Constant	80	80	100	100
DC force V Pulse	80	80	100	100
DC measure V Pulse	80	80	100	100
DC force A Pulse	30	30	50	50
DC measure A Pulse	30	30	50	50
Ethernet				
Speeds (Gbps)	40	40	100	400

Manufacturable solutions exist, and are being optimized

Manufacturable solutions are known

Interim solutions are known

Manufacturable solutions are NOT known

NOTES:

1) Audio / Precision; Source & Measure specifications (22 KHz BW)

2) Major testing condition

3) Target Devices are Wireless Baseband, xDSL, ODD, Digital TV (Track Mobile Baseband)

4) Target Devices are HDD, Radar, WiGig

5) For Measure Sample Rate: Dependent on method, tracking or Front End filter.

6) The purpose of DC accuracy for this table is for high resolution force/measure and trim

7) Devices may also need high current with the less accuracy

8) Markets include Automotive, Battery Management and Power. This does not include high voltage breakdown test.

Difficult Challenges in the Short Term

- As reflected in the tables, manufacturing solutions exist for the immediate future testing needs. However, high DC accuracy for sourcing, measuring and for trim/fuse blowing/register-setting in a manufacturing environment could be at issue depending on how high a resolution/accuracy the DUT is. Also 40 Gbps Ethernet has known manufacturing solutions, but none are optimized.
- Time-to-market and time-to-revenue issues are driving test to be fully ready at first silicon. The analog/mixed-signal test environment can seriously complicate 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.

Medium-term Trends (6 to 10 years out)

- For Wireless Baseband, xDSL, ODD, and Digital TV (Track Mobile Baseband) devices, the source and measure bandwidths, sampling rates and resolutions increase, while the noise floors are decreasing.
- Additionally, DC force and measure accuracies get more challenging.
- Ethernet speeds trending to 100 Gbps [2] have only interim solutions identified.
- Higher speeds and modulation will necessitate PAM to handle the increased data bandwidth – for example, PAM4, 8 or 16 at speeds of 32 GBPS. [3], [4]

Difficult Challenges in the Medium Term

- As the capability requirements increase, there are solutions available, but they do not lend themselves easily to high volume manufacturing.
- Basic physical and electrical properties come more into play. For example, a -150 dB noise floor is possible, but special fixturing is required that is difficult to deploy into a manufacturing environment.
- Ethernet speeds of 100 Gbps [2] have only interim solutions identified.

Long-term Trends (10 years+ out)

- Ethernet speeds trending to 400 Gbps [5], [6]

Difficult Challenges in the Long Term

- Ethernet speeds of 400 Gbps do not have known manufacturing solutions identified.

SUMMARY

Cost continues to be the most critical pressure and concern for analog mixed signal because much of the volume for this is consumer oriented. However, in the medium and long term, performance starts becoming an issue for high-volume manufacturing in terms of bandwidth, sample rate, resolution and noise floor to keep up with the newer devices on the horizon. Ethernet in the medium and long term has manufacturing challenges both in optimization and known solutions.

References

1. https://www.cisco.com/c/dam/en/us/products/collateral/switches/catalyst-6500-series-switches/white-paper-c11-737238.pdf
2. https://www.networkcomputing.com/data-centers/25-50-and-100-gigabit-ethernet-data-center/1422885308
3. http://www.ieee802.org/3/100GNGOPTX/public/mar12/plenary/tremblay_01_0312_NG100GOPTX.pdf
4. http://www.ieee802.org/3/bm/public/sep12/ghiasi_01a_0912_optx.pdf
5. http://searchdatacenter.techtarget.com/feature/Understanding-tomorrows-Ethernet-speeds
6. https://www.keysight.com/main/application.jspx?ckey=2716254&id=2716254&nid=-32176.0.00&lc=eng&cc=US
7. https://en.wikipedia.org/wiki/Effective_number_of_bits

Section 7: Wafer Probe and Device Handling

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.

Device Handling Trends

Increased parallelism at wafer probe drives a greater span of probes across the wafer surface and significantly increased probe card complexity. Prober and probe card architecture should evolve to simplify the interface, however just the opposite is happening: ATE tester complexity is decreasing and more technology and complexity is built into the probe card 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 test temperature. 3D integration technology requires very precise probing technology in X, Y and Z, as micro-bumps may be easily damaged during the probing process. MEMS applications require a variety of testing environments such as pressure, magnetic, and vacuum environments; also, wafer shape and package style are 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 reduced test times), faster handler speed, or process improvements such as asynchronous test or continuous-lot processing. 3D integration technology requires new contact technology for the intermediate test insertion which will be added between conventional front-end process and back-end process. New contact technology to probe on the singulated and possibly thinned die's micro-bumps or C4 bumps after the die is mounted on an interposer is needed. For the die-level 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/receives 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, die-level handlers as well as package 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 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 still being cost effective in a competitive environment is a significant challenge.

Gravity feed, turret, and strip handlers have been added to the table while retaining the pick and place type handler. The gravity feed handler is used on SOP, QFN, and DIP packages. Turret handlers are widely used on discrete-type QFN devices. Strip handlers are used on the frame before singulation. 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 is quite different but important for various type of LSI.

Table 1: 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°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 kit-less 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°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(ΔT) during test
	Both Logic and Memory wafer generates more wattage/heat, demand of Heat dissipation performance improvement is expected. Especially Heat Dissipation at Hot temperature is challenging technology for wafer prober.
	There are wafer handling requirements of non-SEMI standard such as 3DI, MEMS, WLCSP and PsP applications. Those are thin, thick, unique shape so customized wafer handling technique/technology is needed. Wafer cassette is needed to be customized to meet the request as well.
	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 application requires very thin wafer which drive need for 'Taiko Wafer' and 'Ring attached wafer' handling and more high voltage chuck technologies.
	Enhanced Probe Z control is needed to prevent damage to pads, there are solution in the market but those must be optimized to integrate onto wafer prober to meet needs of test cost requirement.
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

Table 2 (part 1): Wafer Probe Technology Requirements

Year of Production	***2019***		***2020***		***2021***	
MPU, ASIC, SOC and Mixed Signal Products						
Wirebond - inline pad pitch	**40**		**35**		**35**	
Wirebond - stagger pad pitch	**45**		**30**		**30**	
Bump - array bump pitch	**30**		**30**		**30**	
Sacrifical pad pitch in a field of bumps	**100**		**100**		**100**	
I/O Pad Size (µm)	**X**	**Y**	**X**	**Y**	**X**	**Y**
Pad Materials						
Wirebond	**30**	**30**	**30**	**30**	**30**	**30**
Bump	**30**		**30**		**30**	
Sacrifical pad in a field of bumps	**45**	**45**	**42**	**42**	**42**	**42**
Wafer Test Frequency (Hz)	**2.4G**		**2.4G**		**2-10 GHz**	
Wafer Test Frequency (Hz) for HSIO	**25Gbps/12.5GHz**		**56Gbps PAM4 28Gbps NRZ @ 14GHz**		**100Gbps PAM4 @ 28GHz**	
Probe Tip Diameter Wirebond	**7.5**		**6.5**		**6.5**	
Probe Tip Diameter Bump	**25**		**25**		**25**	
Probe Force Bump(gf) - at recommended overdrive	**1.5**		**1.5**		**1.2**	
Size of Probed Area (mm^2)	**20000**		**20000**		**20000**	
Number of Probe Points / Touchdown	**180000**		**200000**		**200000**	
Maximum current per probe >130um pitch	**2A**		**2A**		**2A**	
Maximum current per probe <130um pitch	**1A**		**1A**		**1A**	
Maximum contact resistance	**<0.5**		**<0.5**		**<0.5**	
Probe test temperature range	**-55**	**200**	**-55**	**200**	**-55**	**200**
Automotive Radar						
Wafer Test Frequency (GHz)	**80GHz**		**80GHz**		**80GHz**	
RF I/O Geometry	**Solder Ball**		**Solder Ball**		**Solder Ball**	
I/O Size (um)	**100um Cu Pillar SB**		**100um Cu Pillar SB**		**100um Cu Pillar SB**	
I/O Pitch (um)	**300um**		**300um**		**300um**	
RF Ports per Site	**14**		**14**		**13**	
Sites being probed together	**2**		**4**		**4**	
Total Number of RF Ports	**28**		**56**		**52**	
High Speed Digitial (TIAm CDR, VCSEL, etc.)						
Wafer Test Frequency (GHz)	**67GHz**		**67GHz**		**67GHz**	
RF I/O Geometry	**X**	**Y**	**X**	**Y**	**X**	**Y**
I/O Size (um)	**50**	**50**	**50**	**50**	**50**	**50**
I/O Pitch (um)	**80um**		**80um**		**80um**	
RF Ports per Site	**24**		**24**			
Sites being probed together	**2**		**8**		**8**	
Total Number of RF Ports	**48**		**96**		**96**	
802.11ad						
Wafer Test Frequency (GHz)	**64GHz**		**64GHz**		**64GHz**	
RF I/O Geometry	**Solder Ball**		**Solder Ball**		**Solder Ball**	
I/O Size (um)	**80um**		**70um**		**70um**	
I/O Pitch (um)	**150um**		**125um**		**125um**	
RF Ports per Site	**32**		**32**		**32**	
Sites being probed together	**8**		**8**		**8**	
Total Number of RF Ports	**256**		**256**		**256**	
5G						
Wafer Test Frequency (GHz)	**45GHz**		**73GHz**		**50-60GHz**	
RF I/O Geometry	**Solder Ball**		**Solder Ball**		**Cu Pillar w/o cap**	
I/O Size (um)	**100um**		**70um**			
I/O Pitch (um)	**150um**		**130um**		**130um**	
RF Ports per Site	**34+**		**38+**			
Sites being probed together	**8**		**8**			
Total Number of RF Ports	**64**		**>100**			

Table 2 (part 2): Wafer Probe Technology Requirements. NOTE VCSEL and PIC have different requirements

Year of Production	***2019***		***2020***		***2021***	
Optical Probe - ***NOTE VCSEL and PIC have different requirements***						
Minimum pitch between fibers (um)	**127**		**120**		**120**	
Fiber optical alignment accuracy (Multi-Mode)	**< 5um**		**< 10um**		**<10um**	
Fiber optical alignment accuracy (Single-Mode)	**< 0.1um**		**< 0.1um**			
DRAM						
Wirebond - inline pad pitch	**50**		**50**		**50**	
I/O Pad Size (µm)	**X**	**Y**	**X**	**Y**	**X**	**Y**
Wirebond	**40**	**50**	**35**	**40**	**35**	**40**
Sacrificial Pads	**45**	**50**	**40**	**40**	**40**	**40**
Wafer Test Frequency for Sort(Hz)						
Test Frequency(Hz)	**250M**		**400M**		**400M**	
Shared Signal Line Test Frequency (Hz)	**125M**		**200M**		**250M**	
Minimum pulse width	**2.0nS**		**2.0nS**		**2.0nS**	
At Speed Wafer Test						
Test Frequency(Hz)	**3.2G**		**3.2G**		**3.2G**	
Probe Tip Diameter	**8.5**		**8.5**		**8.5**	
Probe Force(gf) - at recommended overdrive	**2.5**		**2.5**		**2.5**	
Size of Probed Area (mm^2)	**100% of wafer**		**100% of wafer**		**100% of wafer**	
Number of Probe Points / Touchdown - Memory	**130000**		**150000**		**150000**	
Maximum Current (mA)/pin	**Probe Tip**	**DC Leakage**	**Probe Tip**	**DC Leakage**	**Probe Tip**	**DC Leakage**
	250	**<.001**	**250**	**<.001**	**250**	**<.001**
Maximum Resistance (Ohm)	**Contact**	**Series**	**Contact**	**Series**	**Contact**	**Series**
	<0.5	**<3**	**<0.5**	**<3**	**<0.5**	**<3**
Probe test temperature range	**-45**	**150**	**-45**	**175**	**-45**	**175**
NAND						
Wirebond - inline pad pitch	**80**		**80**		**65**	
I/O Pad Size (µm)	**X**	**Y**	**X**	**Y**		
Wirebond	**50**	**60**	**50**	**60**	**50**	**60**
Wafer Test Frequency for Sort (Hz)						
Wafer Test Frequency(Hz)	**100M**		**133M**		**133M**	
At Speed Wafer Test						
Test Frequency(Hz)	**600M**		**600M**		**2.4G**	
Probe Tip Diameter	**10**		**10**		**10**	
Probe Force(gf) - at recommended overdrive	**3**		**3**		**3**	
Size of Probed Area (mm^2)	**100% of wafer**		**100% of wafer**		**100% of wafer**	
Number of Probe Points / Touchdown - Memory	**80000**		**80000**		**80000**	
Maximum Current (mA)/pin	**Probe Tip**	**DC Leakage**	**Probe Tip**	**DC Leakage**	**Probe Tip**	**DC Leakage**
	250	**<.001**	**250**	**<.001**	**250**	**<.001**
Maximum Resistance (Ohm)	**Contact**	**Series**	**Contact**	**Series**	**Contact**	**Series**
	<0.5	**<3**	**<0.5**	**<3**	**<0.5**	**<3**
LCD driver Products						
Bump - inline pad pitch	**18**		**16**		**16**	
Bump - stagger pad pitch	**10**		**8**		**8**	
I/O Pad Size (µm)	**X**	**Y**	**X**	**Y**	**X**	**Y**
Inline	**11**	**50**	**11**	**50**	**11**	**50**
Stagger	**15**	**30**	**12**	**40**	**12**	**40**
High speed I/O pin freq (Mobile/TV)	**4.5Gbps / 6.5Gbps**		**4.5Gbps / 6.5Gbps**		**4.5Gbps / 6.5Gbps**	
Probe needle structure	**Cantilever / Vertical**		**Cantilever / Vertical**		**Cantilever / Vertical**	
Probe Tip Diameter (um)	**8**		**8**		**8**	
Probe Force(gf)	**2**		**2**		**2**	
Size of Probed Area (mm^2)	**5600**		**6800**		**6800**	
Number of Probe Points / Touchdown	**12000**		**12000**		**12000**	
Maximum Current (mA)/pin	**Probe Tip**	**DC Leakage**	**Probe Tip**	**DC Leakage**	**Probe Tip**	**DC Leakage**
	300	**<.001**	**300**	**<.001**	**300**	**<.001**
Maximum Resistance (Ohm)	**Contact**	**Series**	**Contact**	**Series**	**Contact**	**Series**
	<0.5	**<3**	**<0.5**	**<3**	**<0.5**	**<3**

Table 2 (part 3): Wafer Probe Technology Requirements

Year of Production	***2019***		***2020***		***2021***	
CMOS Image Sensor						
Wirebond - inline pad pitch	**90**		**80**		**70**	
I/O Pad Size (µm)	**X**	**Y**	**X**	**Y**	**X**	**Y**
Wirebond	**60**	**70**	**60**	**65**	**60**	**60**
WLCSP	**46**	**100**				
WLCSP (TSV construction)	**55**	**55**	**40**	**40**	**40**	**40**
	200M		**200M**			
High speed I/O pin freq (Hz)	**2.5G**		**3G**			
Probe needle structure	**Vertical / MEMS**		**Vertical / MEMS**		**Vertical / MEMS**	
Probe Tip Diameter Wirebond (um)	**12**		**10**		**7**	
Probe Force Wirebond(gf)	**2**		**2**		**2**	
Size of Probed Area (mm^2) [3]- Visible light	**300x300**		**300x300**		**300x300**	
Number of Probe Points / Touchdown - IR [4]	**5000**		**10000**		**10000**	
Maximum Current (mA)/pin	**Probe Tip**	**DC Leakage**	**Probe Tip**	**DC Leakage**	**Probe Tip**	**DC Leakage**
Visible light sensor	**250**	**<.001**	**250**	**<.001**	**250**	**<.001**
IR sensor [5]	**1000**	**<.001**	**1200**	**<.001**	**1200**	**<.001**
Visible Light Sensor / Optical Fiberoptic Transmission						
Maximum Current (mA)/pin	**Probe Tip**	**DC Leakage**	**Probe Tip**	**DC Leakage**	**Probe Tip**	**DC Leakage**
Visible light sensor	**250**	**<.001**	**250**	**<.001**	**250**	**<.001**
Parametric (Process monitor)						
Inline pad pitch	**40**		**40**		**40**	
Inter-row pad pitch	**35**		**35**		**35**	
Pad Size (µm)	**X**	**Y**	**X**	**Y**	**X**	**Y**
In line pads	**20**	**20**	**20**	**20**	**20**	**20**
Probe Tip Diameter	**6**		**6**		**6**	
Number of pad rows	**2**		**2**		**2**	
Probe Force(gf) - at recommended overdrive	**2**		**2**		**2**	
Number of Structures /Touchdown	**8**		**8**		**8**	
Maximum Capaciance (pF pin to pin)			**1**		**1**	
Maximum Leakage (pA)/pin (10V / 1 Sec test)	**0.2**		**0.2**		**0.2**	
Maximum Contact resistance (Ohms)/pin	**0.3**		**0.3**		**0.3**	
Maximum Path resistance (Ohms)/pin	**3**		**3**		**3**	
Maximum Probe temperature Range (degrees C)	**-50**	**200**	**-50**	**200**	**-55**	**200**
Maximum test Frequency (GHz)	**3**		**6**		**6**	

Table 3: Wafer Prober Requirements

Year of Production	*2019*	*2020*	*2021*
Wafer Handling			
Wafer Size [inch]			
200mm Prober	6, 8	6, 8	6, 8
300mm Prober	8, 12	8, 12	8, 12
Min Bump Size[um]	15	15	15
Min Wafer Thickness[um]	200	100	100
Max Wafer Thickness[um]	3000	3000	3000
Max Wafer Weight[g]	350	350	350
Min Wafer Exchange Time (sec)	30	30	30
Tester Docking			
Test Head Weight[Kg]	1500	1500	1500
Probe Card			
Probe Card diameter[mm]	580	725	725
Probe Card PCB Thickness[mm]	10	18	18
Probecard Total Height [mm]			
Prober			
XY Accuracy (Probe to Pad) [±um]			
200mm Prober	2.0	2.0	2.0
300mm Prober	2.0	1.0	0.8
Z Accuracy (Probe to Pad) [±um]			
200mm Prober	5.0	3.0	2.0
300mm Prober	5.0	2.0	2.0
Chuck Planarity [±um]			
200mm Prober	7.5	7.5	7.5
300mm Prober	7.5	5.0	5.0
Chuck Maximum Force [Kg]			
200mm Prober	60	60	60
300mm Prober	450	450	500
Set temperature range [ºC]			
200mm Prober	-55 to +300	-55 to +300	-55 to +300
300mm Prober	-55 to +250	-55 to +250	-55 to +250
Chuck Temp. Accuracy [±℃]			
200mm Prober	1.0	1.0	1.0
300mm Prober	1.0	1.0	1.0
Chuck Leakage [pA]			
200mm Prober	0.1	0.1	0.1
300mm Prober	0.1	0.1	0.1
Total Power Logic (W/Die)			
300mm Prober	200	200	200
Total Power Memory (Watts Per Die)			
300mm Prober	0.75	0.80	0.80
Max Voltage [V]			
200mm Prober	10000	10000	15000
300mm Prober	10000	15000	15000
Max Electrical current [A]			
200mm Prober	300	300	300
300mm Prober	300	300	300

Table 4 (part 1): Test Handler Requirements

Year of Production	***2019***	***2020***	***2021***
Pick and Place Handlers (High Performance)			
Temperature set point range (°C)	-20 to 125	-20 to 125	-20 to 125
Temperature accuracy at DUT (°C)	±1.0	±0.5	±0.5
Number of pins/device	2500	4000	5000
Throughput (devices per hour)	2-10K	2-10K	2-10K
Sorting Categories	3-6	3-6	3-8
Maximum Power Dissipation (W/DUT)	400	500	700
Maximum socket load per unit (kg)	80	120	200
Maximum Package Size(mm)	50x50	75x75	90x90
Minimum Package Thickness (mm)			
Pick and Place Handlers (Consumer SoC/Automotive)			
Temperature set point range (°C)	-55 to 190	-60 to 200	-75 to 200
Temperature accuracy at DUT (°C)	±1.0	±1.0	±1.0
Number of pins/device	1000	1200	1200
Throughput (devices per hour)	2-30k	5-30k	5-30k
Sorting Categories	3-6	3-8	3-8
Maximum Power Dissipation (W/DUT)	40	40	40
Maximum socket load per unit (kg)	80	80	80
Minimum Package Size(mm)	2x2	2x2	2x2
Minimum Package Thickness (mm)	0.2-1.8	0.2-1.8	0.2-1.8
Pin/land pitch (mm)	0.3	0.3	0.3
Pick and Place Handlers (Memory)			
Temperature set point range (°C)	-55 to 155	-55 to 155	-55 to 155
Temperature accuracy at DUT (°C)	±1.0	±1.0	±1.0
Number of pins/device	50-1000	50-1000	50-1000
Throughput (devices per hour)	20-75K	20-75K	20-75K
Index time (sec)	2-3	2-3	2-3
Sorting Categories	5-9	5-9	5-9
Minimum Package Size(mm)	4x6	3x5	3x5
Minimum Package Thickness (mm)	0.2-1.8	0.2-1.8	0.2-1.8
Pin/land pitch (mm)	0.2	0.2	0.2
Ball edge to package edge clearance (mm)	>0.1	>0.1	>0.1
Gravity Feed Handlers			
Temperature set point range (°C)	-55 to 175	-55 to 200	-55 to 200
Temperature accuracy at DUT (°C)	±2.0	±1.0	±1.0
Parallel testing:	8 (2x4)	16 (2x8)	16 (2x8)
Throughput (devices per hour)	50k	50k	50k
Index time (sec)	0.6-0.8	0.6-0.8	0.6-0.8
Sorting Categories	3-10	3-10	3-10
Minimum Package Size(mm)			
Minimum Package Thickness (mm)			
Conformity tube type (mm)	280-580	280-580	280-580
Turret Handlers			
Serial testing	2-4	2-4	2-4
Index time (sec)	0.072	0.072	0.072
Throughput (devices per hour)	50k	50k	50k
Minimum Package Size(mm)			
Minimum Package Thickness (mm)			
Sorting Categories	5-9	5-9	5-9
Impact load to PKG (N)	3	3	3

Table 4 (part 2): Test Handler Requirements

Year of Production	*2019*	*2020*	*2021*
Strip L/F Handlers			
Temperature set point range (°C)	-55 to 155	-55 to 155	-55 to 155
Temperature accuracy at DUT (°C)	±1.0	±1.0	±1.0
Number of pins/device	6-250	6-250	6-250
Parallel testing:	1-256	1-256	1-256
Throughput (devices per hour) 1-16 parallel	20-120K	20-120K	20-120K
Index time (sec)	0.15	0.15	0.15
Sorting Categories	32	32	32
Min. Pkg. Size(mm)	0.8x0.8	0.8x0.8	0.8x0.8
Max. Strip Size(mm)	300x100	300x100	300x100

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 the 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 for these higher requirements. Current models show difficulty in making sockets for high ball count devices and achieving I/O bandwidths of > 20GHz.

Socket Trends

Table 3 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 their 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 and stroke, and to 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, mainly used for testing BGA-DRAM devices, 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. 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.

Table 4: Test Socket Technology Requirements

Year of Production	***2019***	***2020***	***2021***
TSOP – Flash (NAND) – Contact blade			
Commodity NAND Memory			
Lead Pitch (mm)	0.3	0.3	0.3
Data rate (MT/s)	133	133	266
Contact blade			
Inductance (nH)	5-10	5-10	5-10
Contact Stroke (mm)	0.2-0.3	0.2-0.3	0.2-0.3
Contact force (N)	0.2-0.3	0.2-0.3	0.2-0.3
Contact resistance (m ohm)	30	30	30
Slit width (mm)	0.17	0.17	0.17
BGA – DRAM – Spring Probe			
Commodity DRAM (Mass production)			
Lead Pitch (mm)	0.25	0.25	0.2
DRAM RM GT/S	5.3	5.4	6.4
Spring Probe			
Inductance (nH)	0.2	0.2	0.15
Contact Stroke (mm)	0.2	0.2	0.2
Contact force (N)	<0.2	<0.2	<0.2
Contact resistance (m ohm)	100	100	100
BGA – SoC – Spring Probe (50 ohm)			
Lead Pitch (mm)	0,3 mm	0,25 mm	0,25 mm
I/O data (GT/s)	56 G/s	56 G/s	112 G/s
Spring Probe (50 ohm)			
Contact force (N)	0,3 (N)	0,2 (N)	0,2 (N)
Contact resistance (m ohm)	28 mOhm	28 mOhm	15 mOhm
BGA – SoC – Conductive Rubber			
Lead Pitch (mm)	0,3 mm	0,25 mm	0,25 mm
I/O data (GT/s)	56 G/s	56 G/s	112 G/s
Conductive Rubber			
Inductance (nH)	0,1 nH	0,1 nH	0,05 nH
Contact Stroke (mm)	0,1 mm	0,1 mm	0,05 mm
Contact force (N)	0.1	0.1	
Contact resistance (m ohm)	20 mOhm	20 mOhm	10 mOhm
Thickness (mm)	0.5	0.5	
QFP/QFN –SoC – Contact blade + Rubber			
QFP/QFN –SoC			
Lead Pitch (mm)	0.3	0.3	0.3
Data rate (GT/s)	20	40	40
Contact blade + Rubber			
Inductance (nH)	0.15	＜0.1	＜0.1
Contact Stroke (mm)	0.2	0.2	0.2
Contact force (N)	0.2-0.3	0.2-0.3	0.2-0.3
Contact resistance (m ohm)	30	30	30

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 roadmap, but it is difficult to reduce contact force without decreasing the compression travel.

Contact blade + Rubber, generally used for testing QFP/QFN high frequency SoCs, 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 for a Contact Blade type contact.

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.

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 over the roadmap timeframe are expected to exceed 20 GHz, which will greatly challenge impedance matching and 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 ensure 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 stringent mechanical requirements. To handle these high-frequency signals, the user has to carefully consider the signal integrity of the overall test system including board design/components/socket.

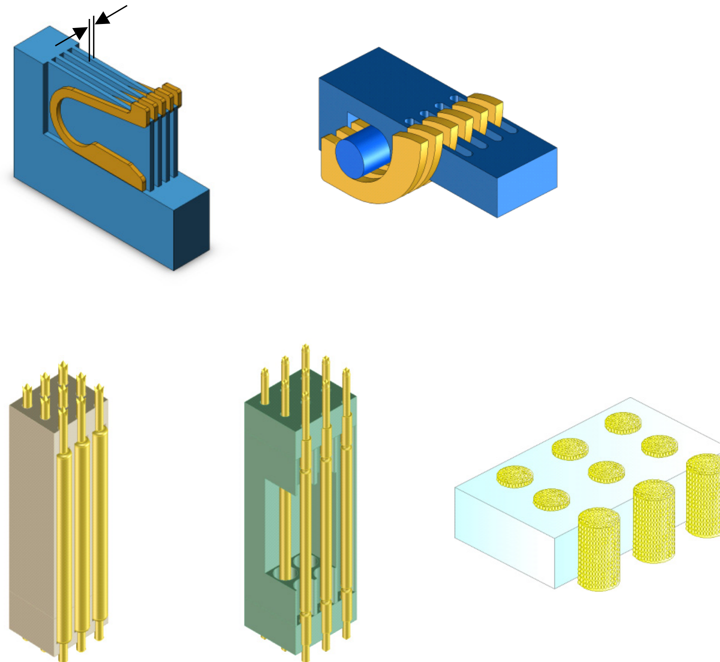


Figure 1: Contactor Types

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 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-direction), 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.

Section 8: System Level Test

The section dedicated to system level test (SLT)[17] was introduced for the first time in the 2019 edition of the HIR Test Chapter.[18] As such, it was written much like a whitepaper covering historical background, then-current practices, gaps, challenges, and future needs discernable at the time. In the few years since then, broadened penetration of semiconductor electronics into the multitude of systems that govern our daily lives has significantly affected the role of SLT to meet user expectations in aspects such as quality, reliability, and safety. This update will focus on what's next for SLT from the refreshed perspective of today. For readers less familiar with SLT, a review of the 2019 edition as well as some recent topical papers are recommended.[19] [20] [21]

Executive Summary

While increasing integration and complexity continue to drive the need for SLT, two recent trends are impacting SLT from additional directions:

1. The rise of "bespoke" silicon optimized for specific application domains dictated by system architects.
2. Integration of chiplets in advanced packaging to realize optimized end-system products.

Behind these trends is the accelerating demand in computing and communications far outpacing slowing performance improvements offered by continued semiconductor technology scaling. Both trends impact upstream testing of the components and sub-systems that eventually form the final system. In essence, even if it's not feasible to perform full-fledged SLT, some aspects of the end-system need to be considered in the way individual components are tested. Thus, instead of viewing of SLT as a traditional last-stage test insertion, various forms of system-oriented testing need to occur at every stage from wafer sort, through die stack, packaging, to assembled sub-systems.

Rapid proliferation of AI applications in the cloud and at the edge has made the importance of energy-efficient computing paramount. With the death of Dennard scaling and untenable increase in multi-core complexity, system providers are resorting to novel architectures to meet power and thermal constraints. Architects resort to bespoke silicon and chiplets to maximize performance for specific use cases and data types via SW-HW co-optimization.[22] [23] [24]

However, during stand-alone testing of individual components prior to integration, more nuanced settings of test conditions and pass/fail criteria are needed when the full system SW-HW context is lacking. System scenarios that may unduly create stress conditions on individual components causing system failure are hard to anticipate. Overall system performance and reliability are degraded by the weakest member in the set of assembled components. The key challenge is thus the mapping of system context to the upstream testing of individual components.

[17] https://eps.ieee.org/images/files/HIR_2019/HIR1_ch17_test08.pdf
[18] https://eps.ieee.org/technology/heterogeneous-integration-roadmap/2019-edition/hir-test-chapter.html
[19] Beyond Structural Test, the Rising Need for System-Level Test, https://ieeexplore.ieee.org/document/8373238
[20] Exploring the Mysteries of System-Level Test, https://ieeexplore.ieee.org/document/9301557
[21] System-Level Test: State of the Art and Challenges, https://ieeexplore.ieee.org/document/9486708
[22] https://semiengineering.com/ic-architectures-shift-as-oems-narrow-their-focus/
[23] https://semiengineering.com/bespoke-silicon-rattles-chip-design-ecosystem/
[24] https://semiengineering.com/rise-of-the-fabless-idms/

Enablers and Challenges of System-oriented Test

- Flexible DFT architecture allowing delivery and execution of both structural and functional test content on multiple tester platforms spanning ATE and in-system.
- Low-cost multi-site high-throughput system functional testers.
- Tight link to system verification for rapid functional test development on both ATE and SLT platforms.
- Transient fault modeling and analysis to better reflect failures at system level.
- Effective SW-HW system failure diagnosis methods for efficient root-causing and yield learning.
- Deep extraction of component internal parametrics that can be correlated with system behavior via advanced data analytics.
- Creation of deep data models can predict how a component will likely behave in the system as well as finding a set of compatible components to meet integrated system performance targets.
- Closer collaboration among supply chain parties to share data and create effective predictive models.
- Standards and practices to meet security requirements despite potentially enlarged threat surface caused by increased data access and sharing.

Section 9: Data Analytics

Background

An IEEE Xplore® database search yields publications on Data Analytics for Adaptive Test and Yield Learning dating back 30+ years. While advances have been achieved over the last several decades, it's challenging to apply the techniques holistically across the full semiconductor value chain. Limitations on our ability to efficiently collect, store, and analyze the massive amounts of available data have limited adaptation to well-defined and self-contained applications. During the past 5-10 years, multiple technological advances have combined to change this landscape significantly:

- The Internet of Things has facilitated the efficient collection of massive amounts of data
- Cloud Computing and Big Data technologies have turned data silos into Data Lakes and Data Meshes
- Tremendous advances in computational power and parallel processing have facilitated the adoption of advanced Data Analytics and machine learning models
- The combination of all the above has enabled rapid advancements in algorithm design and implementation

The foundation is now in place to strategically improve Adaptive Test and Yield Learning, by implementing Data Analytics, Big Data, and Machine Learning techniques.

Why is Data Analytics Important for Semiconductor Manufacturing and Test?

Today's challenges of increased design complexity including Heterogeneous Integration (HI) packages, functionality, shrinking process nodes, increased quality and reliability requirements, and shortened time to market have combined to drive an exponential level of pressure to improve the semiconductor value chain. A massive amount of data – we conservatively estimate multiple terabytes (TB) of data (device and operational) per day for a fully-loaded high volume back-end operation – is collected across the semiconductor manufacturing supply chain and test flow [1]. That data contains a wealth of information that can help optimize the overall test flow and discover hidden issues and relationships across process steps. For example, if the correlation between process drifts and yield is fully understood, immediate actions can be taken to maximize profit and ensure supply (e.g., predictive analytics). A multitude of key insights can be unlocked by using advanced Data Analytics. Data collection during production test should strategically be designed to take full advantage of new and different analytic techniques.

Data collected at test is critical for driving learning and optimization during the product lifecycle using automated data analytics. This includes:

- Cost of test and back-end operations (including test content optimization across test steps)
- Yield (optimized across all test steps)
    - to drive repair/redundancy, die matching – including chiplets, on-die trim, dynamic voltage scaling and fail data collection for diagnosis
- Product Quality – including shipped DPM and product reliability

- Product Performance – such as speed, power and functionality/repair
    - this includes data gathered from on-die monitors and sensors
- Supply chain traceability – such as all components that are used in a HI package
- Time-to-Market, efficient product introduction, feedback to design

A clear requirement is that all test results (e.g., wafer probe test, final test, SLT) and other data from across the semiconductor value chain will need to be merged and available for these analytics, while maintaining high levels of data security and trust for both data at rest and data in motion across pipelines between entities.

Database and IT infrastructure is critical to enable data analytics. Cloud Technology is a key enabler for end-to-end test data analytics across multiple test steps in the value chain from silicon to system test (full product lifecycle). Analytics will be applied at multiple levels including off-line in the Cloud, local to the tester, and at the Edge (for reduced latency and real-time decision-making).

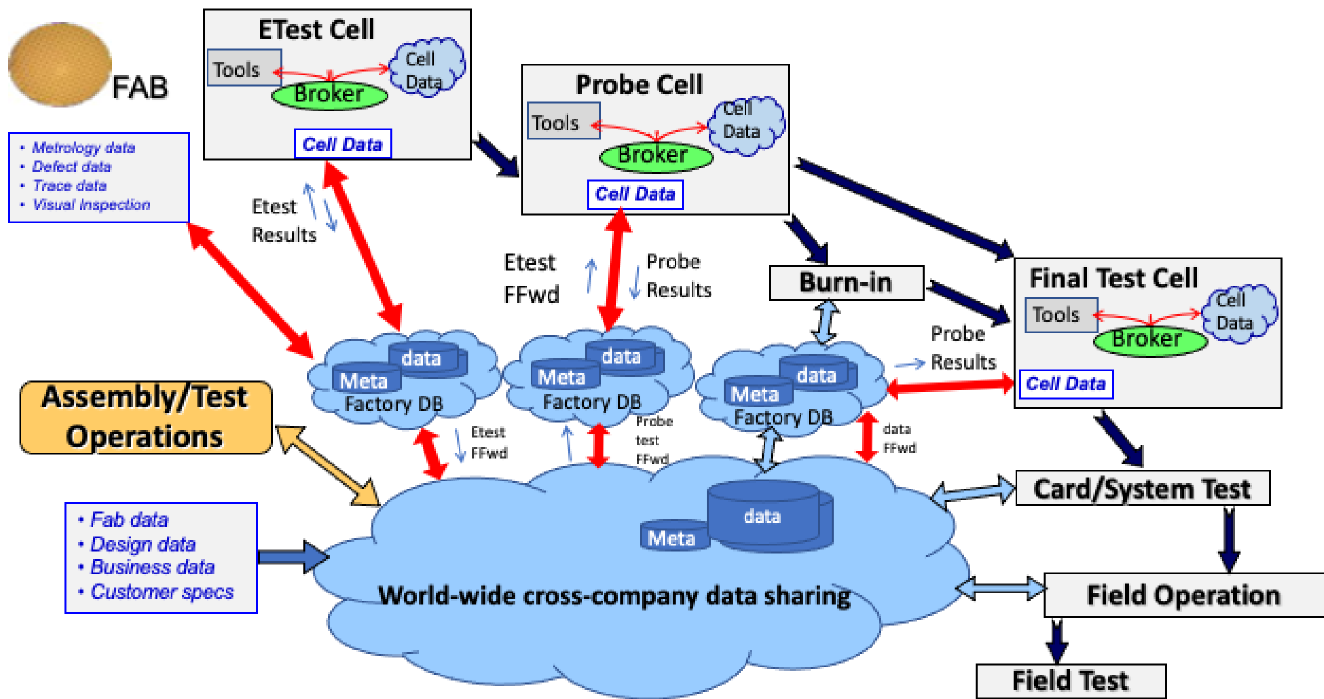


Figure 1: The architecture of Adaptive Test organizes each insertion's test data into one or more databases. A waterfall of manufactured parts may insert, join or query databases for test flow decision-making.

Data analytics requires real-time analysis to enable capabilities such as Adaptive Testing, as shown in Figure 1. This analysis is done either local to the tester or at the Edge, within the required latency to drive production test and dispositioning.

Heterogeneous integration is increasing the importance of data analytic capabilities due to the complexity of combining many different dies – sometimes from multiple suppliers – onto the same package. Tasks such as yield analysis require the merging and analysis of a wider set of data from these dies and packages.

Failure or delay in applying modern Data Analytics holistically across the semiconductor value chain will lead to increased costs and risks as design, fab, assembly, and test complexities increase, and stop-gap measures are implemented to reach quality targets. Attempts to optimize manufacturing process steps individually, without full consideration of the interactions and dependencies across the entire process flow, will lead to diminishing returns. A test escape anywhere in the process flow reduces the quality level of the overall flow.

Tactical, localized solutions to manufacturing challenges are usually costly. One example is adding a System-Level Test (SLT) insertion as a back-end quality screen. By the time issues are detected, the manufacturing process is typically so far downstream that the effort required to truly root-cause and resolve them is only justified for the most major and systemic issues. With Data Analytics that relate SLT fails to other process and test data, issues can be caught sooner and SLT becomes one of a series of test insertions rather than a backstop.

Multiple benefits of applying advanced Data Analytics are described in the sections below.

Transforming the Backend to an Industry 4.0 Smart Factory

In essence, Smart Manufacturing is the trend towards automation, enhanced data connectivity and advanced analytics to improve efficiency. It involves automating repeatable tasks, using data from process, production, assets maintenance, and production planning to gain actionable insights through analytics. The path to achieving this includes the use of cyber-physical systems (CPS), the Internet of Things (IoT), Industrial Internet of Things (IIoT), cloud computing, cognitive computing, and machine learning.

Today, the big push towards a smart factory is to:

- Reduce costs
- Improve overall equipment effectiveness (increased uptime, accelerated output, decreased faults)
- Improve quality
- Enable secure data exchange across the value chain to improve visibility and productivity

The key requirement is the ability to collect, share and act on the data. For Smart Manufacturing there are three dominant data perspectives:

- Historic state - Review, analyze and model historical performance
- Current state - Monitor current state to enable real-time control
- Future state - Use history and current data to identify and plan for future improvements

It is crucial in Smart Manufacturing to measure and characterize every aspect of the manufacturing process, including logistics, products, machines and processes, without scrambling to consolidate data. A consistent and detailed strategy for collecting, analyzing, and categorizing data is essential.

Optimizing Cost of Test

Savings from cost of test reductions are easily quantifiable, as they drop directly to the bottom line as increased profit. This benefit must be balanced with other factors that can potentially have significantly greater impact on profitability and competitiveness, such as improved yield, quality, and reliability. (See *Section 11: Key Drivers and Test Costs* for a detailed comparison of these impacts.) Advanced Data Analytics are key to achieving this balance, through their ability to identify complex effects and interdependencies, both within a specific test insertion and across the entire test flow. Examples:

- Data Analytics-driven decisions, including those made on-the-fly based on results from the current and/or previous insertions, support a smart, adaptive approach for optimizing test coverage at reasonable cost.
- HI-related technologies such as chiplets drive a "shift-left" of testing to earlier insertions to guarantee known good die (KGD) as well as providing the data necessary for speed binning and die matching. Data Analytics facilitates an efficient and cost-effective shift-left strategy through correlation of results across all test insertions from wafer probe through SLT.
- Correlation of test results across multiple insertions facilitates moving test seconds to lower-cost (or even fully depreciated) equipment while maintaining required test coverage.
- Incorporating additional design-for-test (DFT) circuitry can reduce test system requirements but must be applied carefully as it uses valuable chip real estate. With the complexities that come with advanced packaging, embedded sensor IP provides an effective means for monitoring chip performance and functionality at a deeper level. Data Analytics play a key role in maximizing the information that can be inferred from this sensor data across test insertions as well as in-field.
- Machine Learning models are being successfully used to reduce test cost by replacing time-consuming searches (for example, determining trim values or setting test parameters such as voltage levels) with fast predictions based on previously collected data [2, 3].

This ability to optimize across the entire flow becomes especially important to keep test costs under control for complex devices requiring an added System-Level Test insertion, or devices for automotive applications that have rigorous multi-temperature testing and burn-in requirements. Optimizing test across the entire flow requires tools and standards that support the efficient combination of data from different processes, devices, and equipment.

Improving Quality Assurance

The primary objective of production test is quality assurance. Data analytics provides a powerful means for ensuring that devices are meeting functionality, quality, and reliability requirements by inferring additional information on existing failure modes and potential quality and reliability issues while maintaining an economically

viable test strategy. Applying data analytics cohesively on test data from multiple test steps further increases overall effective test coverage. This capability is especially valuable in market segments that require high reliability such as automotive, military, aerospace, and medical devices, which strive for "zero defects" outcomes.

Some defects do not manifest during testing or initial operation. For example, recent advances in the awareness of Silent Data Errors have led to calls for additional screening and outlier detection, particularly on devices that exhibit some degree of abnormal behavior even when passing all tests. For this reason, adding more test coverage and/or insertions, which adds cost, may not meet the stated goal of zero defects. Instead, Advanced Data Analytics can be used to minimize test escapes by optimizing the test content at each insertion, and inferring additional valuable information from the combined results data.

Traditional outlier detection techniques utilizing statistical post-processing are well understood, but may not be adequate for catching potential reliability issues at Final Test or System Level Test. Near-real-time statistical techniques will be particularly valuable for devices that do not have individual device traceability, since re-binning in near real-time allows the prober or handler to re-bin devices before they get lost in the population.

Real-time outlier detection offers a potentially useful addition to the set of tools for achieving high reliability. Near real-time analytics is relatively inexpensive compared to additional test time, and is capable of identifying test process issues such as site-to-site bias, enabling corrective action sooner than would be possible with post processing.

Data feed-forward methods are used to analyze upstream test data, to adaptively determine the appropriate downstream test content and minimize test escape rates. Data feed-backward methods are used to adjust the manufacturing process and shorten the time to achieve entitlement yield and quality. Correlations across test insertions can identify drift or other issues. When available, historical data should be mined to set baselines, screening limits, and guardbands. As new test methods are deployed, data analysis can measure the impact to ensure no new test escapes are created when displacing other forms of testing.

When applied as described, data analytics can contribute greatly to reduced Time To Quality (TTQ) and therefore to reduced Time To Market (TTM). Achieving this vision will require more standardization of data formats and traceability wherever feasible. Analytics software will need to be demonstrably secure, and capable of running on multiple data systems.

Improving Yield

Heterogeneous integration presents difficult challenges in terms of both yield prediction for the chiplets (the "known good die" or KGD problem) as well as diagnosing yield losses for the packaged product. Full electrical testing of the individual chiplets prior to package assembly is technically challenging and cost prohibitive for the supply chain and, moreover, defects may occur not only at the chiplet level but throughout the entire package manufacturing and assembly process. The chiplets may be manufactured in different process nodes and at multiple foundries, leading to a vast Pareto of possible defect types, and assembly processes such as wafer-to-wafer stacking or die-to-wafer stacking introduce even more defect sources. This creates test coverage challenges in the fully packaged product, leading to extensive and costly electrical testing. Furthermore, late detection of bad chiplets at package test leads to costly loss of the other good chiplets in the package.

Collecting data at all stages of the manufacturing process can provide complete material traceability and overcome gaps in the conventionally recorded genealogy of the packaged part (e.g., ECID). This richer data set enables new data analytics to trade-off cost versus resolution of test and diagnosis throughout the HI process, and enables die matching to improve overall HI product yield. The time lags inherent in chiplet silicon manufacturing and package assembly processes couple with test and diagnosis challenges to create time-to-yield issues resulting in time-to-market issues, making yield improvement throughout the heterogeneous integration process a critical component to product success.

DFT techniques developed originally for SoC products must also be incorporated in the heterogeneously integrated products without driving increased resources or test time. Resilience must also be designed into the chiplet and package architecture to realistically achieve full functionality. This can be accomplished by additional resources such as redundant TSV's or bonds as well as redundant memory and logic circuits. Data analytics across the entire supply chain will play a crucial role in collecting and model building to allow for the optimization of the system resiliency architecture.

Performance Grading/Binning

Performance grading and binning of devices has been a common technique for many years for tiered products, for example memory or processors. More recently, with mobile and energy-conscious applications, there is a need for an improved performance understanding which could lead to either traditional product binning/grading or product

applications for improved energy/performance trade-offs. With Heterogeneous Integration, understanding of device performance at the wafer level and concepts like die matching or calibration will be of paramount importance. To help achieve these goals, there have been improvements in on-chip sensing technology for both process variation and operational parameter monitoring under various operating conditions. The combination of sensor networks and advanced data analytics provides new signatures at the die level for enhanced binning.

Performance binning for the frequency vs voltage trade-off is also changing. Traditionally this has been accomplished by shmooing voltages and frequency to determine the maximum operating point. These test techniques can be expensive from both a test time and test intensity perspective. Trained models are starting to be deployed based on early characterization data to pre-determine operating points. This can be done as a data feed forward to later test/assembly steps or as an in-situ decision for binning. Doing so results in improved product economics - yield, test time and optimal operating conditions.

Previously, data sources have been focused on voltage, temperature and process. Other measurement parameters have also emerged as critical on-die measurements. An example is a margin measurement that is placed on critical timing paths or interfaces. It provides visibility on the amount of timing margin, which will further indicate performance optimizations or more quickly determine the quality of the device grading being performed. This leads to a better understanding of design margin for optimal operation.

It is also expected that innovative approaches will emerge combining financial, sales and device data to tailor deliverables that exactly match customer requirements. This tuning optimizes manufacturing and test processes, which increases margins, improves lead time and increases supply elasticity.

Traceability Across the Semiconductor Value Chain

With increasingly stringent reliability requirements and use of HI, we must have more visibility into the assembly processes where the root cause for hard to pinpoint reliability failures often occur. With the complex supply chain for the HI assembled product, security considerations have become of paramount importance.

Analyzing failures and security events in electronic devices requires traceability at the individual device level to access the manufacturing, test and root of trust data. Virtual identifiers based on SEMI E142 [4] can provide a basis for single device traceability from any point in the supply chain (wafer, package, PCB, field) both downstream and upstream [5]. The data model is applicable from the wafer through traditional and more advanced packaging technologies such as wafer level packaging and the heterogeneous integration of chiplets. This enables precise analysis for pinpointing the root cause of a failure, for example a rare early life failure of a wire-bond in the field, or for pinpointing the source of a security attack, for example rapid detection and mitigation of counterfeits, Trojans, and malware attacks.

An on-chip electronic identifier (ECID) can be used to trace back to wafer test and further back into wafer fab for root cause analysis. However, this only applies to the primary active components with ECID and does not provide any visibility into the assembly processes . We must add traceability to assembly to capture every active and passive component, bump and wire contact, consumables, equipment, Failure Detection Classification (FDC) trace and inspection images.

Data Analytics for Test - Key Enablers Roadmap

In the table below, the key enablers for realizing the full potential of advanced data analytics to optimize the test process across the semiconductor value chain are listed. For each enabler, the current status is described, as well as the 3-5 year projection of how the enabler needs to evolve to support the bold visions described in the sections above. Importantly, progress on the enablers needs to be comprehensive, as a delay in the development of any of them can hold back overall progress on the successful implementation of advanced data analytics solutions for semiconductor test.

Enabler	Current Status	3-5 year Projection
Machine-to-machine IoT communications infrastructure	Mostly Point-to-point ethernet. Early use of MQTT M2M.	Wide use of MQTT M2M-like net.
Standardization of data formats	Several application specific formats; STDF, SECS combined with generics; CSV, text, binary.	Move to more data-centric formats which support real-time analysis, including streaming data formats.
Real-time analytics - more local processing	Dependent on tester capability. Off-line analysis common.	Cells become data centers. Local real-time processing on test. cell or Edge compute server. Distributed analysis and storage
Strategy for collecting, analyzing, and categorizing data	Most data is indexed via file paths and location. Databases are used for access. Data is mutable and hard to find.	All data indexed via metadata. Emphasis on provenance and trust. Data mesh architectures common.
Characterize every aspect of the manufacturing process	Test data is generally available locally. Non-test data is not common.	All collected data available. Continual addition of new data.
Use of device-sourced data, sensors and test structures	Some use of on die sensors for analysis. Mostly post-processing.	Pervasive use of on-die, in-package and in-system test data sourced from the entire life cycle. Efficient real-time access to on-die sensor data.
Big Data technologies - cloud - local	Storage is limited especially globally due to cost.	Distributed analysis to reduce data size impacts.
Advanced Data Analytics and machine learning models	Some well known techniques. Part Average Testing. Outliers, neighborhoods. Limited in scope due to knowledge models. Some use of machine learning models, mainly for COT reduction.	Extension to non test data. Rule based models common. Pervasive use of machine learning for test optimization, yield enhancement, and quality/reliability improvement. Greater use of unsupervised learning algorithms for anomaly detection, correlations, ...
Pipelines between entities	Ad hoc contract-based solutions. Requires experts to share.	Shareable cross domain models. Knowledge shared effectively along with the data.
Data security	Encryption is used. Some data hiding techniques.	Encrypted analytics and models reduce the need to share raw data.

Impact of COVID-19 on Data Analytics Roadmap (Special Section for 2023)

In 2020 we had expected COVID-19 to be an accelerating force in the adoption of advanced Data Analytics, with key drivers being the move to remote (work, data access, support, etc) requirements for efficient meshing of cloud and edge compute resources, increased supply chain stress, and greater reliance on predictive analytics/diagnostics. Together these factors have driven an urgent need to bridge the worlds of test engineering and data science. A key

question at the time was how strongly these drivers would persist in the post-pandemic world. This hoped-for post-pandemic world has yet to arrive, and instead the world has adapted to living with COVID-19. Ongoing severe issues such as those with the global supply chain have been exposed as systemic problems requiring new approaches rather than quick and temporary fixes. The bridging of test engineering and data science is in progress but increased focus is required to bring the level of expertise in line with the magnitude of the challenges. Government subsidies such as the US and European CHIPS Acts provide important and timely fuel for furthering the development and application of Data Analytics in the semiconductor industry. The combination of a strong requirement for better capabilities in this space and massive government funding provides a great opportunity to make fast progress, but prudent spending will be key to getting the best return on these investments.

Additional Reading

For further reading on *Adaptive Test* and *Yield Learning* topics, please see ***Data Analytics - Appendix A.***

References

1. A. Meixner, *Too Much Fab and Test Data, Low Utilization*, https://semiengineering.com/too-much-fab-and-test-data-low-utilization/, Semiconductor Engineering, Jan 2021.
2. S. Mier, *ML in Semiconductor Test - A Balanced Approach*, https://events.meptec.org/wp-content/uploads/ChipletsDataTest/DataTest2021Mier.pdf, MEPTEC Road to Chiplets: Data & Test, Nov 2021.
3. M. Eiki, K. Schaub, I. Leventhal, B. Buras, *In Test Flow Neural Network Inference on the V93000 SmarTest Test Cell Controller*, 2019 International Test Conference, Nov 2019.
4. *SEMI E142 Substrate Mapping Standard*, https://www.semi.org/en/standards-watch-2020Sept/revision-to-semi-e142.
5. *The GSA Trusted IoT Ecosystem Security (TIES)*, https://www.gsaglobal.org/iot/ties/.

Section 10: 2.5D & 3D Device Testing

Introduction

2.5D and 3D technologies (see figure 1) are characteristics of a system and should be tested as such: testing the complete package at an application level and diagnosing failures at the die and interconnect level. This section will address key test challenges, based on the evolution of 2.5D/3D. These test challenges are with respect to known good dies (KGDs), interposers, high speed interconnects and signal integrity, impact of emerging technologies, 3D TSV/interconnect, as well as 3D probing, die stacks, and stack repair.

Memory die stacks (Wide I/O, High Bandwidth Memory, and Hybrid Memory Cube) were precursors to 2.5D and 3D. Both technologies have provided insights to requirements and challenges associated with 3D and 2.5D test. 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, tend to produce relatively high yields at the stack level. As these adjacent technologies become more mature and as additional 2.5D/3D-TSV applications emerge, more and better data will improve predictions and decision making, with respect to 2.5D/3D-TSV test processes.

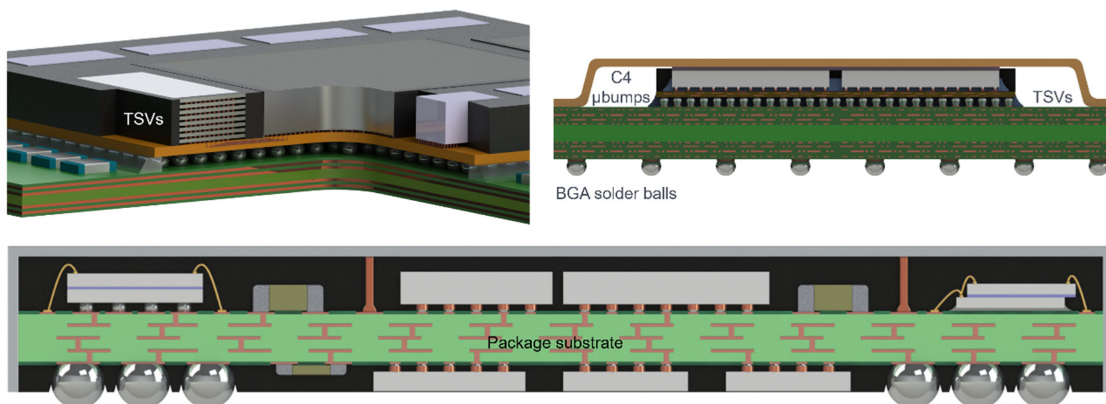


Figure 1 - 2.5D/3D Technology (Amkor Technology, Inc.) [7]

Challenges for Test

While the current state of 2.5D/3D is maturing, new enabling and supporting technologies will require advances in test access, capabilities, and costs. These emerging technologies will provide significant challenges for testing 2.5D and 3D technologies. The challenges below represent potential impacts to test, including increased costs, longer test times, and reduced yields and reliability.

- Known Good Die (KGD) Test

Due to yield concerns at the final package level, incoming bare die should have as good a quality as possible. KGD is the common industry term – but KGD does not mean that 100% of the bare dies will pass all testing at the next level of assembly. Chiplet suppliers should provide an estimate to their customers of the expected fallout at package test or later testing steps.

To achieve high quality, chiplets should see as much testing at wafer probe as possible. But there are challenges since for advanced technology chiplets (such as fine-pitch µ-bump or Copper Hybrid Bonding) not all signal IO will be probed at wafer test. Instead, Design-for-Test (DFT) solutions should enable high test coverage at wafer probe without the requirement to probe all signal IOs. (Using test-only pads is the most common solution). This DFT, combined with the IEEE 1838 standard for chiplet access, should be used to apply tests pre- and post-packaging.

- Interposer Testing

Interposer testing can be accomplished primarily by point-to-point continuity probing. Multipoint interposer testing requires significantly more probing (more time and higher costs) and requires embedded logic to coordinate point-to-multipoint connections. Over time, it is expected that probing becomes more challenging, from Point-to-Point to larger-scale Multipoint.

Known good interposers (KGI) are vital to ensure adequate yields for advanced packages. Post-package assembly, IEEE P1838 primary and secondary TAP ports allow for testing the die-to-die test access and interconnect performance integrity.

- High Speed Interconnects and Signal Integrity

Testing high speed interfaces (HSIO) requires access and ability to run test patterns specifically on each interface, whether in a loopback mode from transmit ports to receive ports or from one chip transmitting to an adjacent receiving chip in an integration. Design for test is needed to run these tests standalone via an easy-to-use interface such as JTAG 1149.1 or 1149.6, SPI, J2C (JTAG to CPU), or PCIe. These tests should have the ability to be run at wafer sort, package test and system test in characterization and production. Designers need to understand defect mechanisms of the HSIO, and what testing will cover all defect types and guarantee outgoing quality. This can be time-consuming and expensive for silicon area. The IEEE1838 standard is available, but a user must go through details of the IO DFT for each die-to-die connection today to ensure an implementation will work. Some applications have a high count of HSIO (512-1025) lanes. Current ATE generally can only test this number of IO up to low GHz range (<20GHz). High speed add-ons such as bit-error-rate testers (BERTs) and digital sampling oscilloscopes (DSOs) are available but are limited to 32-64 direct connections. Loopback testing needs DFT like BERs and PHY control built in to be most effective. Often a second test step is required to perform the loopback test.

Handling noise and thermal cross talk across multiple chiplets can erode HSIO margin, and considerations as to whether special packaging or shielding may be needed. ATE testing does not lend itself well to testing high-speed optical interfaces of photonic devices. There are no production test solutions for multiple optical port photonic devices.

3D interconnects on a product can exceed 100,000. They are becoming increasingly denser (< 3µm), the interconnect technology is evolving, and each new generation brings complicated failure mechanisms (see ref [1], [2], and [3]). A standardized test and repair methodology that considers these trends in 3D interconnects would be helpful.

- Impact of emerging technologies with respect to test

The challenge of wafer probe testing is emerging with many more chip-to-chip connections such as Copper Hybrid Bonding (CHB). Reduced pin count testing will be required for wafer probe. Design-for-Test (DFT) solutions are required to enable complete wafer probe testing.

Die-to-die interconnect testing at final package test will require solutions that enable complete testing and failure diagnostics. Emerging solutions such as UCIe and Bunch-of-Wires (BOW – see ref [5] and [6]) should be explored and ideally an industry standard will emerge to ensure chiplets and SOCs/processor chips can be tested.

Another challenge is the test methods for other types of circuits (not just digital) for 2.5D/3D packaging: for example, test methods for Silicon Photonics, RF and high-speed mixed signal. Some of these circuit types have not typically had the same level of DFT as digital circuits.

- 3D TSV/interconnect testing

Silicon interposers include interconnect and through-silicon-via (TSV) structures. Mechanical integrity of these structures during the manufacturing process ensures electrical performance. DC and AC transient pre-bond testing of interposers helps in screening micro-void and pinhole defects. Testing of interposers may require custom test fixture development and test insertion, which impacts the overall product cost. Custom implementations may require complex test techniques (see ref [4]).

- 3D probing, 3D die stacks, 3D stack repair

3D die stacks offer many potential test moments: pre-bond, mid-bond, post-bond, and final test. The more dies that are contained in the stack, the more pressing is the need for a tool that models the cost and yields of wafer processing, stack assembly, testing, packaging, and logistics, to optimize the stack assembly and test flow.

One of the major pre-bond test challenges is getting test access to the non-bottom dies, where the natural functional interfaces consist of large arrays of fine-pitch micro-bumps. State-of-the-art micro-bump pitches are 40 µm; some advanced products already push this down to 30 µm; and the scaling does not stop there. Feasibility of 40 µm probing has been demonstrated but only single-site – future research should push this to multi-site testing and to even smaller pitches (10 µm).

Once the stacking has commenced, we require specific 3D-DFT (i.e., DFT in addition to the conventional 2D-DFT) to transport test stimuli up into the stack and test responses back down. The 3D-DFT in the various dies should collaborate to form a stack-wide test access architecture. For this purpose, in 2020 the IEEE Std 1838-2019 was released; in the meantime, the three major EDA suppliers have started to provide support for IEEE Std 1838 insertion and usage. The standard supports both INTEST (testing or re-testing the internal circuitry of the die) as well as EXTEST (testing inter-die interconnects).

Stack repair makes sense cost-wise only if the spares are already included as redundancy in the stack. Spares could be individual inter-die interconnects or even full dies; for relatively small investments, spares can significantly increase the overall stack yield. Current-generation products that have seen silicon include redundant interconnects and are already implementing repair.

Long term prediction

It is important to note that 2.5D/3D is an evolving technology, and, because of that, it is difficult currently to make any predictions regarding 2.5D/3D test flows. With this said, 2.5D/3D technologies are expected to create increasingly complicated and time-consuming assembly process flows that can add cost as well as yield challenges to the mix. As packaging technologies and the associated Test challenges will continue to evolve, it is expected that DFT features that can enable yield troubleshooting across all process steps will become a focus for the industry.

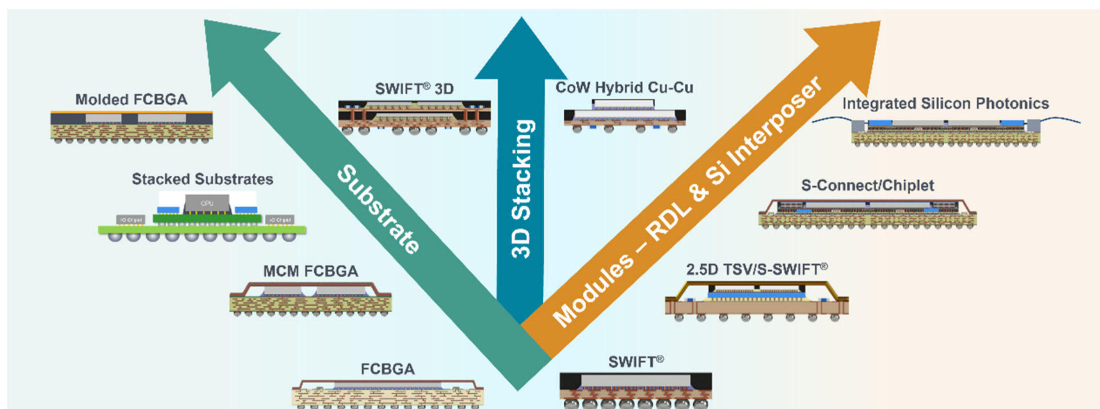


Figure 2: Packaging Technology (Amkor Technology, Inc.) [7]

Call-for-action

To help 2.5D/3D Device testing capability to mature and improve, there are several areas that need attention:

- Known Good Die DFT methods that enable high quality wafer probe test – thus reducing fallout at final test.
- Die-to-die communication standards that enable thorough testing at final test.
- Repair methods at final test to ensure yield is high.
- A standardized test and repair methodology that considers new trends in 3D interconnects.
- Yield prediction and analysis methods that ensure fallout at all levels of testing is understood.
- End-to-End data analytics capability that applies to all dies on the package (see section 9, Data Analytics)

References

[1] A. Elsherbini, et. al., "Hybrid Bonding Interconnect for Advanced Heterogeneously Integrated Processor, IEEE Electronic Components and Technology Conference, 2021
[2] Sreejit Chakravarty, "A Call to Standardize Chiplet Interconnect Testing," IEEE VLSI Test Symposium, 2022
[3] Sreejit Chakravarty, "3D Interconnect Testing Challenges," IEEE European Test Symposium, 2022
[4] Sourav Das, Su Fei and Sreejit Chakravarty, "Testing of Pre-bond Through Silicon Vias", IEEE Design and Test, July/August 2020, pp. 27-34.
[5] R. Farjadrad, M. Kuemerle and B. Vinnakota, "A Bunch-of-Wires (BoW) Interface for Interchiplet Communication," in IEEE Micro, vol. 40, no. 1, pp. 15-24, 1 Jan.-Feb. 2020, doi: 10.1109/MM.2019.2950352.
[6] S. Ardalan, R. Farjadrad, M. Kuemerle, K. Poulton, S. Subramaniam and B. Vinnakota, "An Open Inter-Chiplet Communication Link: Bunch of Wires (BoW)," in IEEE Micro, vol. 41, no. 1, pp. 54-60, 1 Jan.-Feb. 2021, doi: 10.1109/MM.2020.3040410.
[7] © 2022, All content is copyright its respective owners. SWIFT is a registered trademark of Amkor Technology, Inc.

Section 11: Key Drivers and Test Costs

Minimizing costs is a key goal of any 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 lower test times, coupled with increases in parallelism and Built-In Self-Test (BIST) capability. Most SOC devices are tested 2 to 16 at a time, and memory devices can have more than 1,000 devices tested at once.

The typical perception of test costs is that it is dominated by the test equipment itself, which is extremely costly. In reality, that is not the case. The depreciation cost of the test equipment itself (which has a useful life of 15-20 years) typically constitutes less than half the cost to operate a complete test cell, and that cost is zero after the depreciation period (typically 5 or 6 years) has expired.

For large SOC devices, it is notable that, since 2015, the cost of consumable material – material that is expected to be used and then discarded - has become the leading capital expenditure relative to test. This situation stems from:

- The increased cost of interface material (primarily influenced by probe cards and relative items). This cost is driven by finer-pitch probe pads and sockets, and the need for increased maintenance and repair, especially for high current applications.
- The decreasing depreciation period for materials utilized to produce devices used in the mobile device space, where devices have a shorter life span. In most cases, material is typically discarded not because it has ceased to function, but rather because the devices it is used to test are replaced by newer versions which drive different consumable hardware.
- The increasing use of System-Level Test (SLT), where costs are dominated by device-specific hardware. These costs recur with every new device version and, as noted above, often has a very short useful life.

For lower complexity devices, especially those that are not produced in very high volumes, test costs are dominated by capital equipment and are highly affected by Overall Equipment Efficiency (OEE). OEE measures the amount of time the equipment is doing useful work. For devices that are produced in lower volumes, the test equipment is usually taken out of production to change over to test different devices. If these configuration changes happen frequently, OEE is significantly degraded. For this reason, site counts are intentionally limited to lower idle time and increase OEE, even if cost of test per device is slightly higher.

Key Cost of Test Trends

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

- Increases in transistor count that outstrip on-chip test 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 situation will be addressed primarily with increased parallelism and new scan technology to increase external data rates and reduce the number of clock cycles required for a given scan test.
- Device configuration and one-time programming during test is causing more time to be spent 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, although these are really "repair" costs.
- The 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 to improve the yield of multi-die packages.
- Site count at probe test is limited owing to the attendant increase in the cost of consumable material (discussed above) and the limitations of Touch-Down Efficiency (TDE). TDE is discussed in more detail below.
- The continuing increase of silicon content in automotive and other end-uses such as military and satellite applications that require a high level of reliability, which drives additional test insertions for fault coverage and temperature-related test.

Continuous improvement in equipment efficiency will be offset by new device test requirements, so the overall cost of test will remain relatively flat for the foreseeable future.

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 by VLSI Research of the capital and service costs of equipment used in device fabrication, packaging and test.

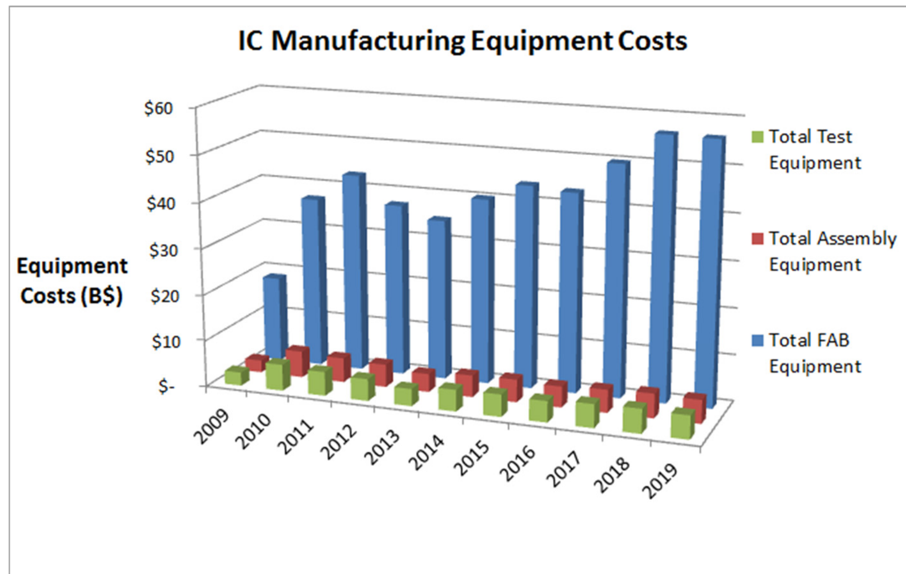


Figure 1: Relative cost of Fab, Packaging and Test Equipment

For any device, the worst-case cost scenario is to ship defective devices that cause failures later in the manufacturing process. Presuming this is not the case and defect rates are acceptable, then the next concern is manufacturing costs.

While it is helpful to focus on the cost of test itself, the impact of test costs on overall device costs varies widely by device type. For devices that use smaller die or mature process nodes, tests cost can be a significant part of overall manufacturing costs. For devices that use leading edge processes and have larger die sizes, the 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.

For these more complex devices, the highest avoidable costs in test are devices that are good but are rejected at test for some reason.

Consider the following, simplified example.

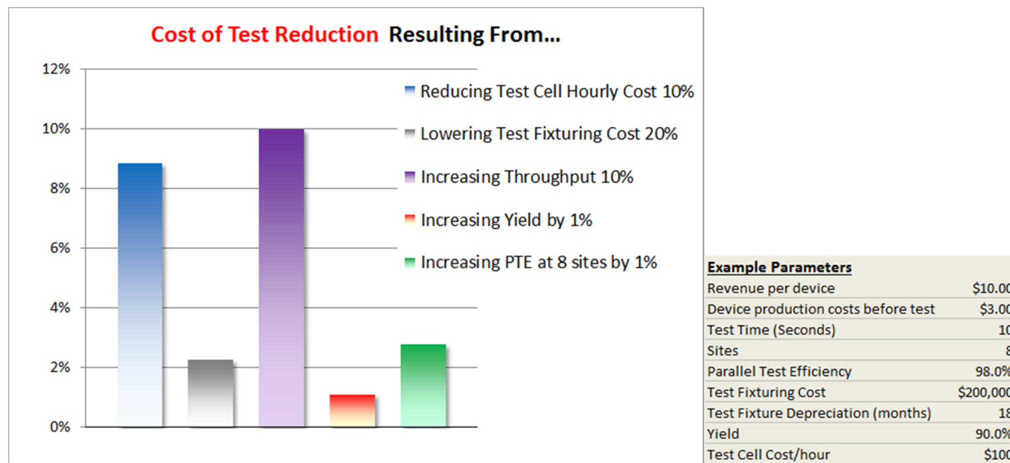
- A device costs $1.00 to manufacture, including fabrication, assembly and packaging, etc.
- Test constitutes 5% of that cost, or $0.05

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.

Figure 2 shows the effect of traditional cost reduction techniques on cost of test.



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 2: 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 savings due to improved yield becomes far more significant.

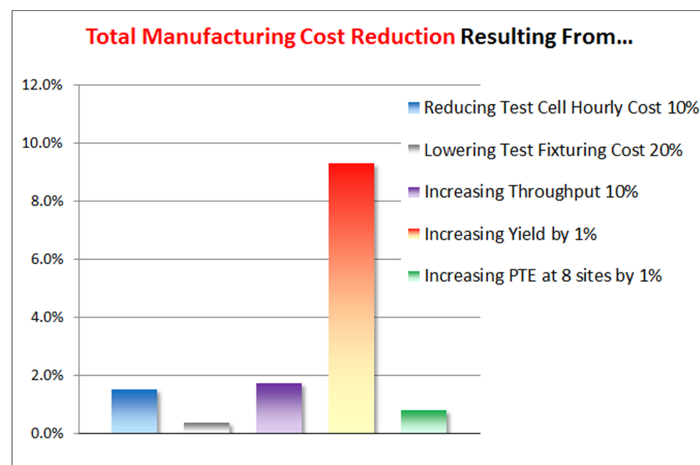


Figure 3: Total Cost of Manufacturing Reduction realized by traditional cost reduction techniques

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 need dramatically increases both the number of measurements made and the accuracy required of the test equipment. These 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 most devices produced.

Of course, the danger resulting from recovering marginal devices to improve yield is that there may be a greater chance of the device failing in the end application. While test costs for complex devices are lower than silicon and packaging costs, the cost of a failing device in an end product easily swamps out both. Striking the balance between yield and device quality has been the challenge of semiconductor test since the beginning. Optimizing for both can only be achieved through better test accuracy or greater test time, both of which drive up test costs.

The remainder of this section 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 to balance reduction in test costs with potential reductions in product yield.

Test Cost Models and Cost Improvement Techniques

The cost of semiconductor test has many drivers, which is further complicated for multi-die packages as shown in Figure 5.

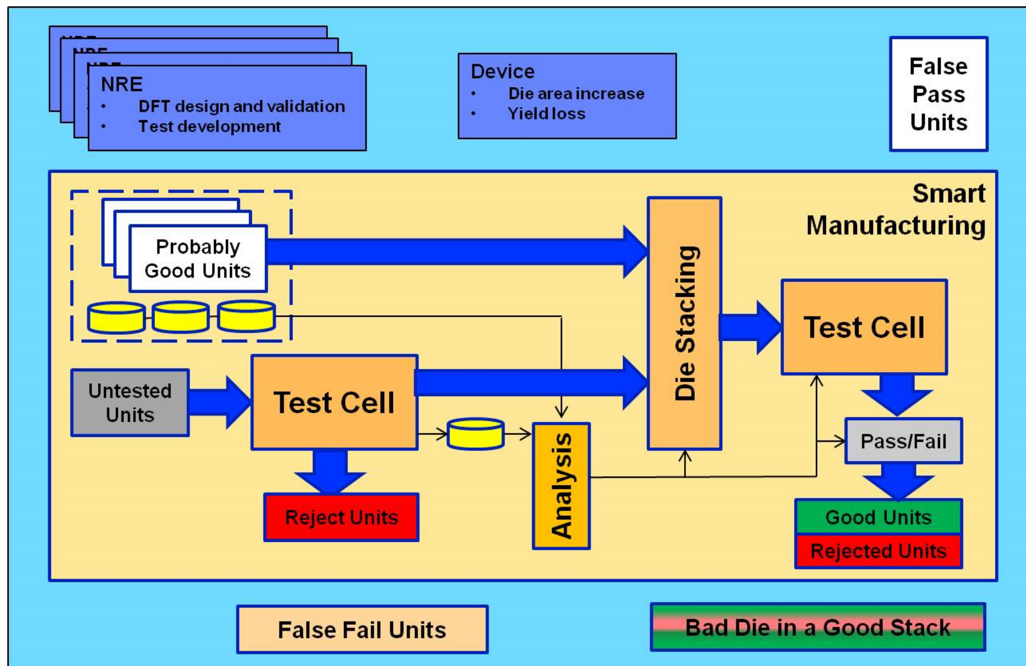


Figure 5: Multi-die Flow

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 and interface expenditures
- Facility/labor costs
- Cost of test program development
- Cost of die space used for test-only functions

Future Cost Drivers

- Increased test time due to additional scan and functional testing
- Increased testing at wafer to produce Known Good Die (KGD)
- Addition of system-level testing to augment traditional ATE test
- 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

Cost Reduction Techniques

- Multi-site and reduced pin-count
- Structural test and scan
- Compression/BIST/DFT and BOST
- Yield learning and adaptive test
- Concurrent test
- Improvements to test processes based on analysis of collected test data

Multi-site Trend

The simplest 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:

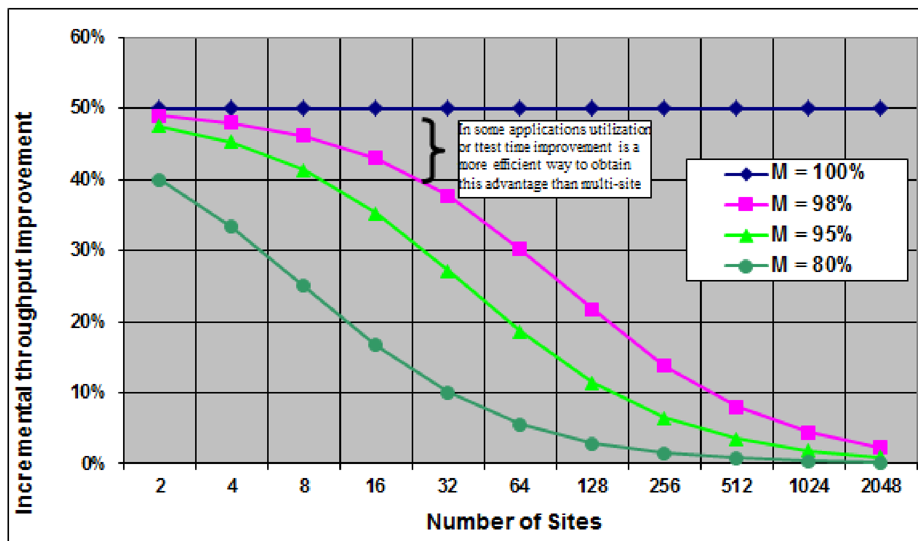
$$M = 1 - \frac{(T_N - T_1)}{(N-1)T_1}$$

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%.

There are cases where increased site count is either not possible or not effective:

- Site count is limited by equipment capability. Additional site count required additional test resources and new prober and handler capability that may either not exist or be prohibitively expensive to use as compared to existing equipment that is already depreciated.
- The test time overhead of adding sites will, at some point, begin to reverse the gains achieved by going to higher site count. This situation is discussed below.
- At wafer probe, Touch-down efficiency (TDE) is limited by the size of the die relative to the size of the wafer. Those details are discussed below.
- Additional site count is most effective for high volume devices, which will efficiently occupy test equipment over long periods of time. For lower volume devices, the down time to reconfigure test equipment between different device types will eliminate any gains made as a result of higher site count.

Figure 6: Importance of Multi-Site Efficiency in Massive Parallel Test



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 10 second

test time to 10.8 seconds. However, going from testing a single site to testing 32 sites will increase a 10 second test time to 16.4 seconds, significantly reducing the potential advantage of multi-site as shown in Figure 6.

Touch-Down Efficiency (TDE) is defined as the number of wafer touch-downs required to test all devices on a wafer, relative to the theoretical minimum. TDE is influenced for the most part by the die size (and therefore the number of die per wafer) and the pattern used to probe. For example, if a device is tested 10 sites at a time, and there are 1,000 die per wafer, then ideally a probe card would have to touch down 100 times to test the wafer and be 100% efficient. If, due to the mismatch between the round shape of the wafer and the linear or rectangular pattern of the probe card, the probe card must touch down 110 times to test the 1,000 devices, then the TDE is closer to 90%. This result is illustrated in the figures below.

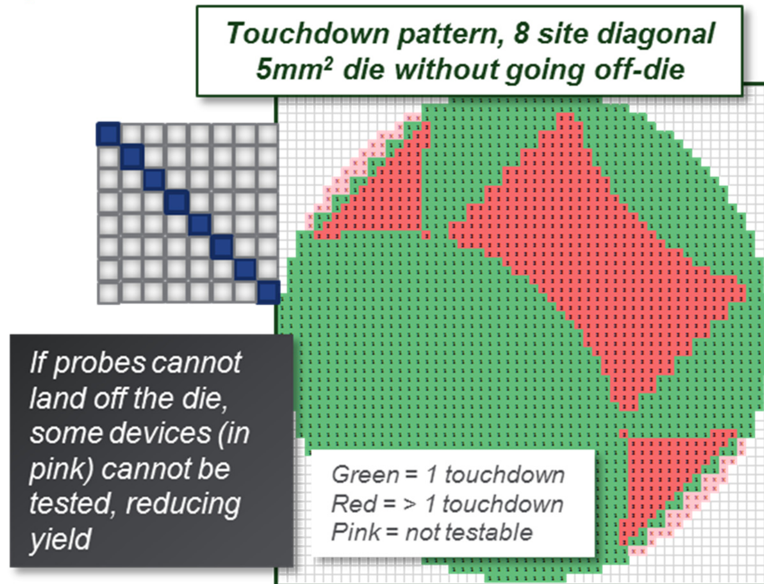


Figure 7: Probe Pattern of 5mm² die using 8-site probe pattern

As die size of a complex device increases, the TDE will continue to degrade as shown in Figure 8. This degradation of efficiency will negate any advantages of increased site count and will eventually increase the cost of test as shown in the example below. In this case, there are gaps in the probe pattern to allow for the inclusion of electrical components on the probe card required for the proper operation of the device under test.

TDE inefficiencies will primarily be addressed by the development of singulated die testing technology. There is significant work underway to allow die to be reassembled in silicon panels that have a rectangular shape as opposed to the round shape of the original silicon wafer. The deployment of this technology will re-start the increase in site count at probe that is currently stalled due to interface costs and TDE limitations.

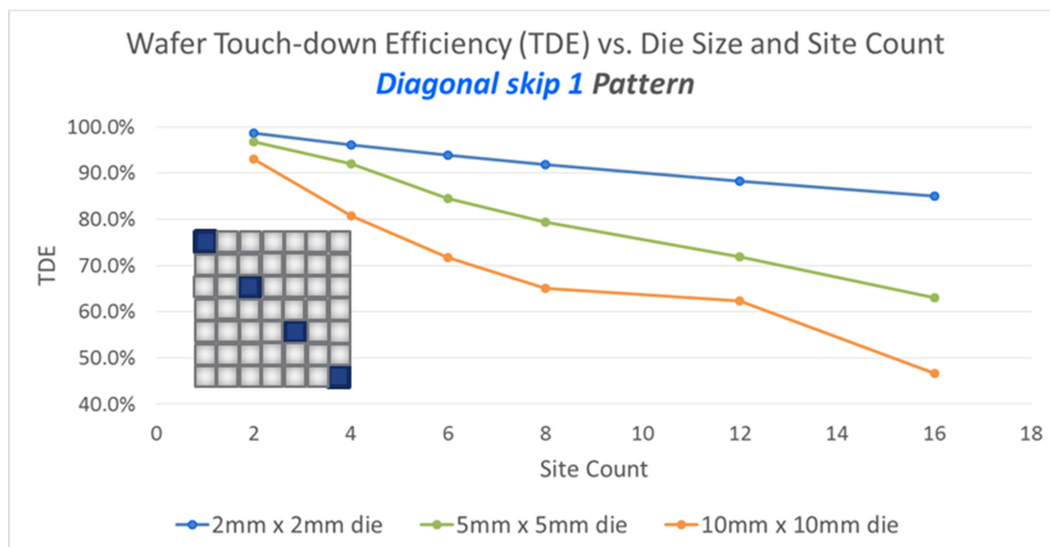


Figure 8: Touch-Down Efficiency as function of die size using a 4-site probe pattern

Because of the TDE inefficiencies of probe, more test could be deferred until devices are packaged (where there is no touch-down penalty). The overall trend, however, is to do most testing at probe because:

- Package costs, especially for more complex devices, are significant enough that the cost of discarding a package because of the bad die is greater than the cost of doing more test at probe.
- Multi-die packaging requires known-good die in order to be cost efficient since the cost of discarding good die because of one failing die is never acceptable.
- Devices are used in some form of chip-scale packaging, where traditional package handling equipment cannot be used.

Summary

Major conclusions are:

- Cost of test has been declining for some time, but the rate of reduction has slowed and will remain flat in terms of test costs per device.
- Major reasons for the slower rate of cost reduction are:
    - Packaging trends that drive more test at the wafer probe insertion where site counts are lower.
    - Increased cost of consumable material, which now dominates tester capital cost in terms of test cell costs.
    - Desire for higher yield, which has a much larger impact on overall device production costs than test costs alone.
    - Desire for higher device quality, especially for automotive applications, which necessitates more test.
- Potential solutions to decrease test costs are:
    - New probing technology which allows test of singulated die.
    - New PCB and Interposer technology to lower the cost and complexity of consumable material.
    - Improvements to the test process through increased use of data analysis and machine learning based on measured data.
    - Factory automation.
    - Cost reduction of system-level testing.

Edited by Paul Wesling