HETEROGENEOUS INTEGRATION ROADMAP

2020 Edition

Chapter 8: Single Chip and Multi-Chip Integration

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.

IEEE ELECTRONICS PACKAGING SOCIETY · ELECTRON DEVICES SOCIETY · IEEE Photonics Society · ASME SETTING THE STANDARD · semi

Table of Contents

To download additional chapters, please visit
http://eps.ieee.org/hir

Chapter 8: Single Chip and Multi Chip Integration

Section 1: Executive Summary and Scope

Introduction

Sixty plus years after the invention of integrated circuits, there have been periodic predictions of the end of Moore's Law. While significant innovations in design and process technologies are ongoing to continue the drive to the next nodes, Moore's Law economics are coming to an end, and some key performance metrics at advanced nodes are plateauing, as described in an article in the business magazine *The Economist* March 12, 2016, as "Moore's Law Saturation" (Figure 1.1). The semiconductor industry is implementing EUV, FinFET and FinFET successors. The 5 nm node is in early production, and 3 nm is in sight. Dr Moore's own prediction of a technical end of Moore's Law is closing in on the target years. The message in the March 12, 2016, article is as relevant as ever today.

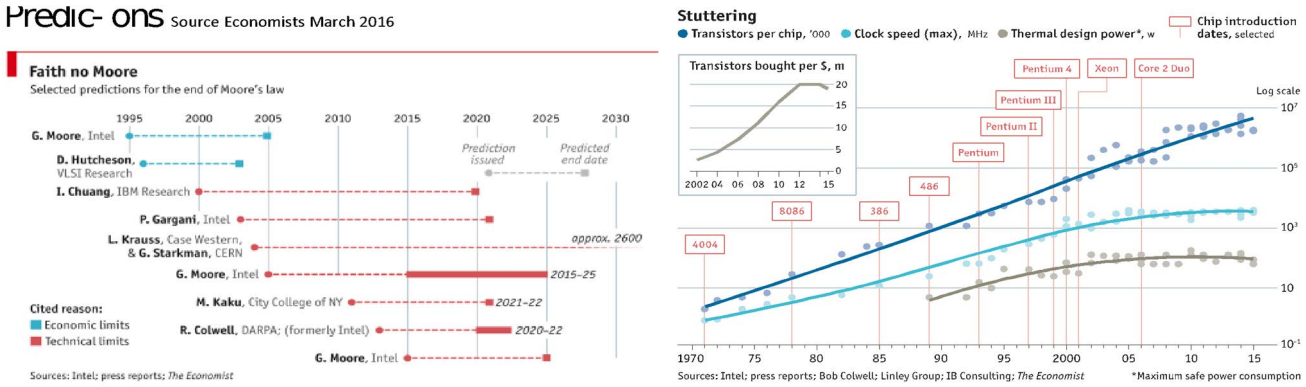


Figure 1.1 Moore's Law Saturation – Performance & Economics. Source: The Economist March 12, 2016

We are entering the era of the digital economy and myriad connectivity, with data migration to the cloud, smart devices everywhere, Internet of Things to Internet of Everything, introduction of 5G, and the emergence of autonomous vehicles. The business landscape is seeing great changes with the rise of tech companies – social media, cloud, search, online commerce, big data, artificial intelligence, leading to integrated hardware-software driven applications and unprecedented growth of application spaces. Figure 2 lists the top 10 publicly traded companies by market capitalization in 2006 and 2019. While there was only one tech company in 2006 in that list, in 2021 8 of the top 10 are tech companies, signifying the transition to the digital age.

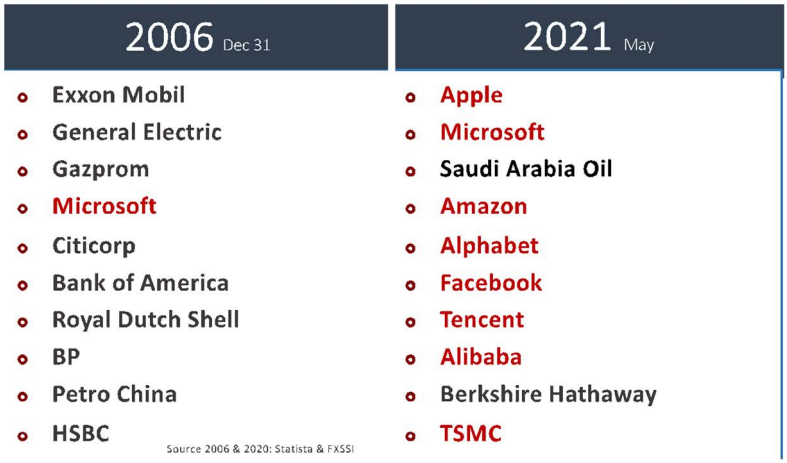


Figure 1.2: Rise of the Tech Companies

At this triple inflection point of plateauing of CMOS scaling advantage, with the transition to a digital economy, big data and artificial intelligence, and the explosive expansion of electronic products into our global society, continued progress requires a different phase of electronics innovations.

In Gordon Moore's celebrated 1965 paper *"Cramming More Components onto Integrated Circuits"*[1] his first focus was integration of transistors into integrated circuits – *"The future of integrated electronics is the future of electronics itself. The advantages of integration will bring about a proliferation of electronics, pushing this science into many new areas.*" In this task the industry has succeeded tremendously well indeed, in investment, technology and the science of scaling – "cramming" billions of transistors into integrated circuits from wafers to chips and chips to packages and packages to products, fostering a global electronics industry for the benefit of society.

In the same 1965 paper, Dr. Moore turns to a system focus: *"It may prove to be more economical to build large systems out of smaller functions, which are separately packaged and interconnected. The availability of large functions, combined with functional design and construction, should allow the manufacturer of large systems to design and construct a considerable variety of equipment both rapidly and economically."*

Following Moore's words, our purpose in Heterogeneous Integration is to build large systems out of smaller functions, which are separately designed, packaged, interconnected, qualified, manufactured, sold and integrated into large systems . This is the purpose and theme of the Heterogeneous Integration Roadmap. Heterogeneous Integration is our focus and direction in the drive to maintain the pace of progress needed to continue the past 56 years of Moore's Law, for Electronics Systems today and years beyond.

How is Heterogeneous Integration defined? Heterogeneous Integration refers to the integration of separately manufactured components into a higher level of assembly (SiP) that, in the aggregate, provides enhanced functionality and improved operating characteristics.

In this definition, components should be taken to mean any of the units whether individual die, MEMS device, passive component and assembled package or sub-system, that are integrated into a single package. The operating characteristics should also be taken in their broadest meaning including characteristics such as system level performance and cost of ownership. Source: ITRS Packaging & Assembly chapter.

Heterogeneous Integration is and will be the key technology direction going forward. It is the pivotal direction for initiating a new era of technological and scientific advances to continue and complement the progression of Moore's Law scaling into the distant future. Packaging – from system packaging to devices packaging – will form the vanguard to this enormous advance.

The Heterogeneous Integration Roadmap addresses six specific market segments: High Performance Computing and Data Center, IoT and 5G, Smart Mobile, Automotive, Wearables and Health, and Aerospace and Defense, as illustrated in Figure 1.2.

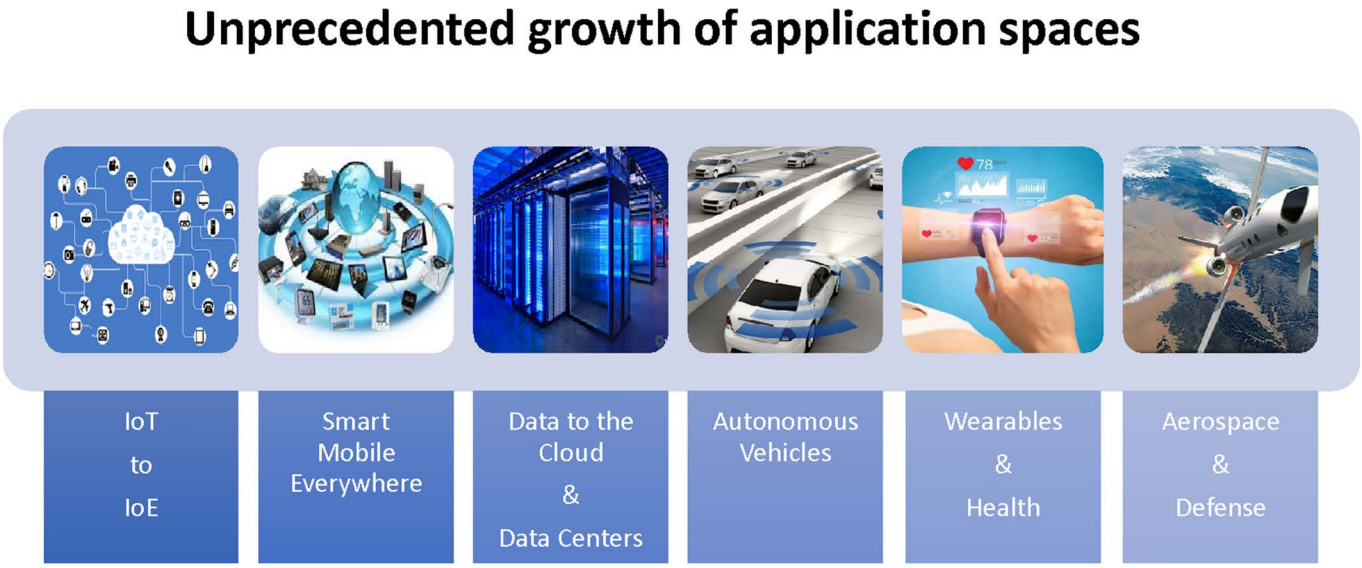


Figure 1.3 Six Major Applications Markets

This Single and Multichip Integration Chapter covers the basic knowledge-base tools and physical manufacturing infrastructure tools across all market segments. IC devices start in wafers from foundries, thinned and singulated into "chips". While wirebond remains the workhorse of the industry, there is very strong growth in Flip Chip (BGA-CSP) and WLCSP, wafer packaging (Fan-In and Fanout), and specifically a high growth rate for advanced packages shown in Figure 1.4, below.

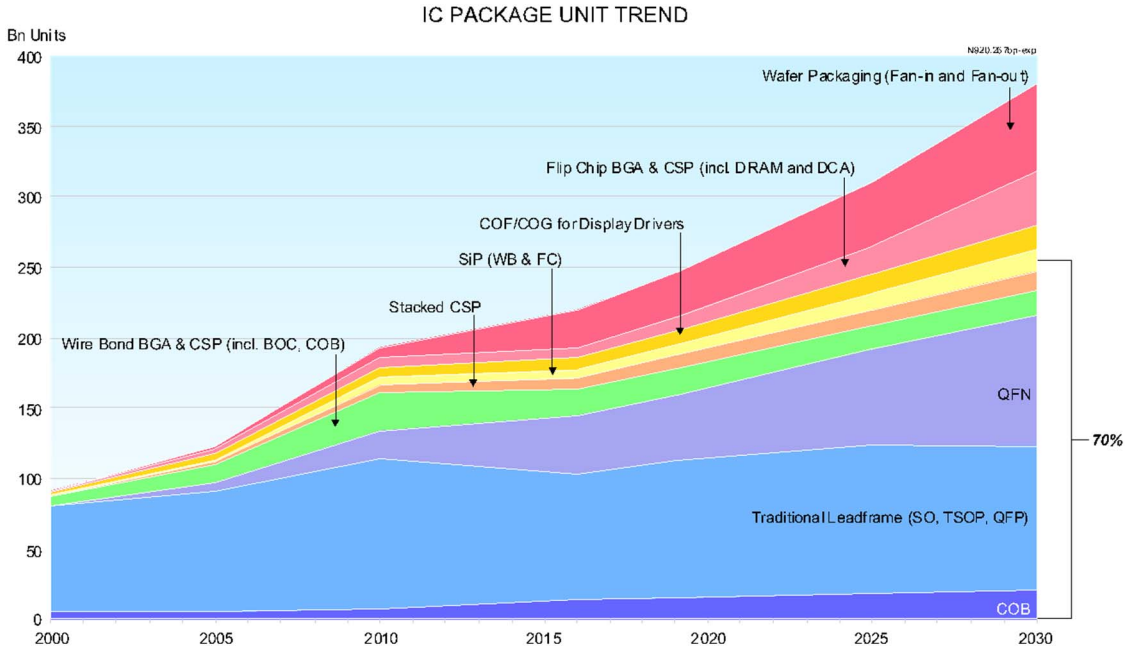


Figure 1.4 Chip Package interconnect technology trends (Bn units): Source: Prismark Partners 06-2021

The ICs are assembled and packaged on substrates into components and the components are mounted on boards. Additive manufacturing is emerging that may see implementation across all the interconnect and assembly processes. In the design of components and systems, engineers consider electrical, thermal, and mechanical performance requirements, and address quality and reliability issues such as electromigration. Together, the eleven sections of this chapter form the knowledge and manufacturing infrastructure in the implementation of Heterogeneous Integration for electronic products. Following Moore's words, our purpose in Heterogeneous Integration is to build large systems out of smaller functions including System in Package (SiP) and Chiplets – which are separately designed, packaged, interconnected, qualified, manufactured and sold.

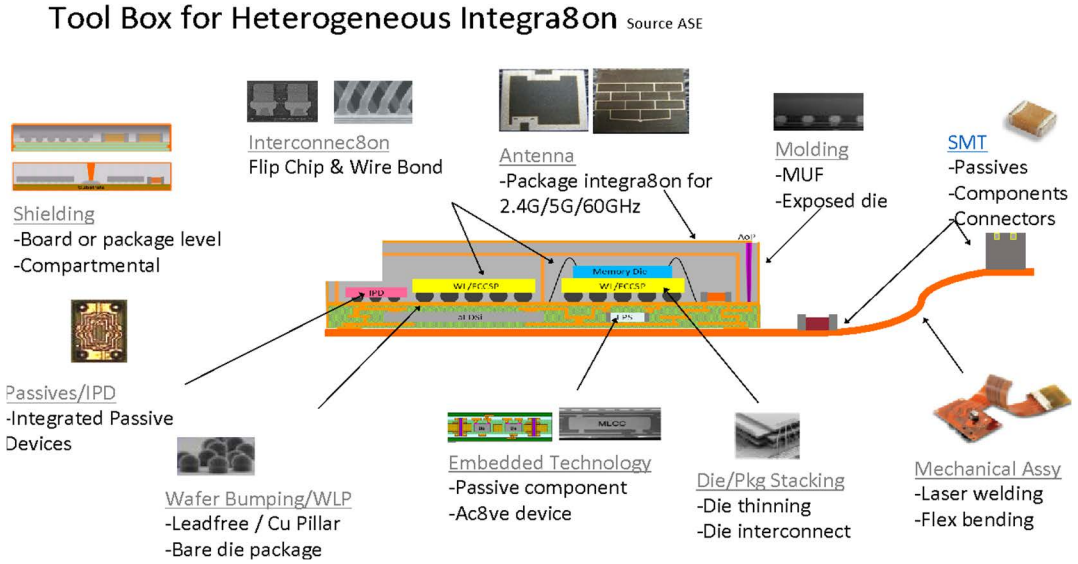


Figure 1.5 Tool Box for Heterogeneous Integration Technologies. Source: ASE

A basic tool box for SiP and Heterogeneous Integration is illustrated in Figure 1.5. This well-established design and manufacturing ecosystem has been highly productive, flexible, and responsive in producing electronic products across the whole spectrum of products serving consumers and industries large and small, with well-established companies and new startups building SiPs and through heterogeneous integration, from Home Assistants, Smart Phones, Data Centers, Automotive, Avionics, and many other products in the trillion-dollar global electronics market.

Our purpose in Heterogeneous Integration is to build large systems out of smaller functions – System in Package – which are separately designed, packaged, interconnected and manufactured. The twelve sections in this chapter articulate the basic tool sets – infrastructure and knowledge – for heterogeneous integration for all the market segments.

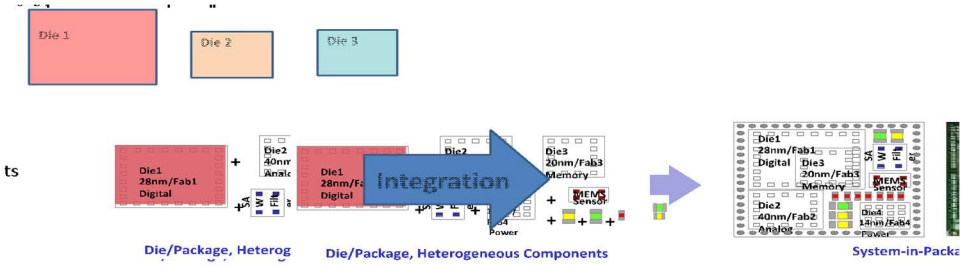


Figure 1.6 Heterogeneous Integration and System in Package (SiP). Source ASE

A very good example of Dr. Moore's suggestion is the mobile smartphone industry, which has been an effective adopter of heterogeneous integration technology through the use of System in Package (SiP) for their advantages in miniaturization and modularity for co-design, from product debut to subsequent generations of product roll out to the marketplace. This industry has adopted heterogeneous integration through SiP, very effectively developing new functions and expanding capabilities in their products while maintaining the phone form factor in this highly competitive global consumer market. The application processor, housed in a PoP package with its memory component stacked in close proximity on the top, is almost always of the most advanced node. This is the premier example of System-in Package incorporating the most advanced node processor integrated with memory die in close proximity for the demanding consumer market.

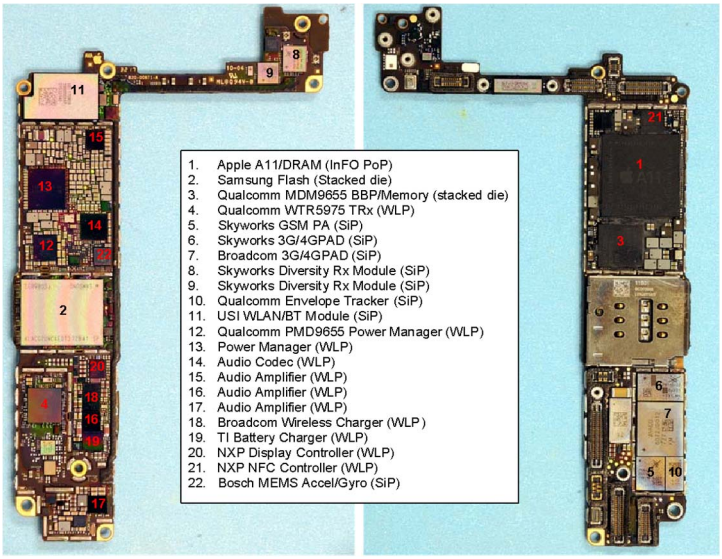


Figure 1.7: Apple iPhone X main Board. Source Prismark Partners

Shown above is the main board for the iPhone X. The main A11 processor die is housed in a PoP package flip-chip assembled on an advanced substrate on the bottom with a wirebonded memory component on the top. There are 8 other SiPs and 11 WLCSP-packaged devices tightly assembled with many passives on the rigid-flex board. To paraphrase Dr. Moore, availability of large functions in the form of SiPs, combined with functional design and construction, should allow the manufacturer of large systems (smartphone) to design and construct models from one product generation to another both rapidly and economically for the hugely competitive consumer market.

Let us now consider the High Performance Computing market application. Shown below is an example of heterogeneous integration through SiP, with the processor and HBM memory stacked on a silicon interposer platform. The package provides massively parallel, high bandwidth connectivity to the HBM, a significant power savings, and greater than half shrinkage in X-Y form factor.

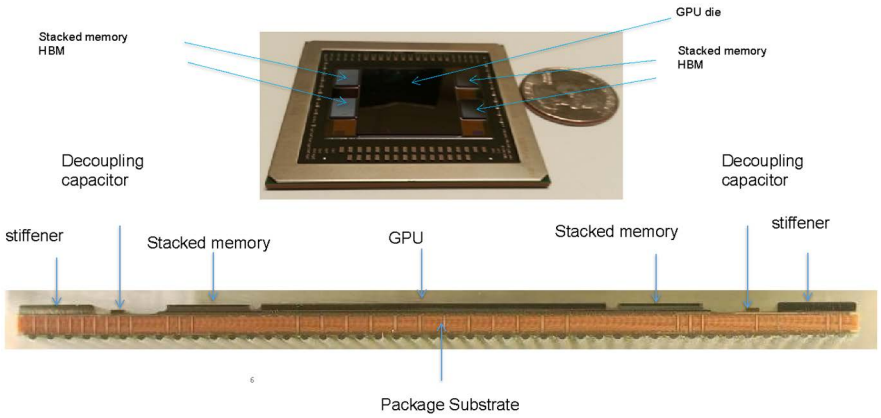


Figure 1.8: AMD Fiji GPU – HDM on Si Interposer 2.5 D Package. Source ASE

The silicon interposer is a physical substrate platform for the CHIPS program, a part of the DARPA Electronics Resurgence Initiative. CHIPS stands for Common Heterogeneous Integration and Intellectual Property IP Reuse Strategies Program. The vision is for an ecosystem of discrete modular IP blocks, to be assembled into a system using existing and emerging integration technologies.

At advanced nodes the die yield falls exponentially with die size. Splitting a large monolithic SoC into smaller tightly-coupled die, first demonstrated by Xilinx on a silicon interposer, are now being seriously considered and executed. At the same time, die cost per unit area is escalating [2]. Shown below are two generations of the AMD EPYC server processors . To the left, the large monolithic SoC has been split into four tightly coupled die (for better yield) called chiplets, in homogeneous integration on an organic substrate. To the right there are two groups of four 7 nm chiplets on each side of the larger 14 nm I/O die in heterogeneous integration to optimize unit area die cost.

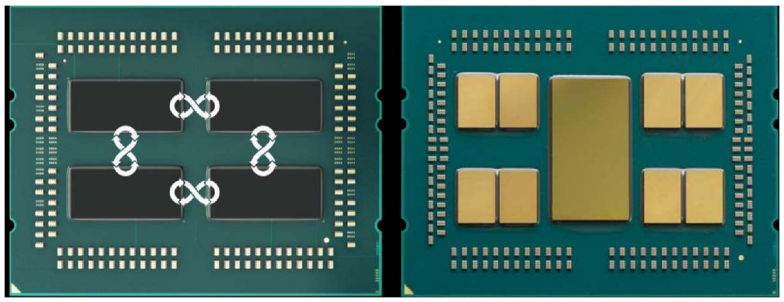


Figure 1.9A: Examples of System Integration – 1st and 2nd Generations EYPC Server Processors. Source AMD

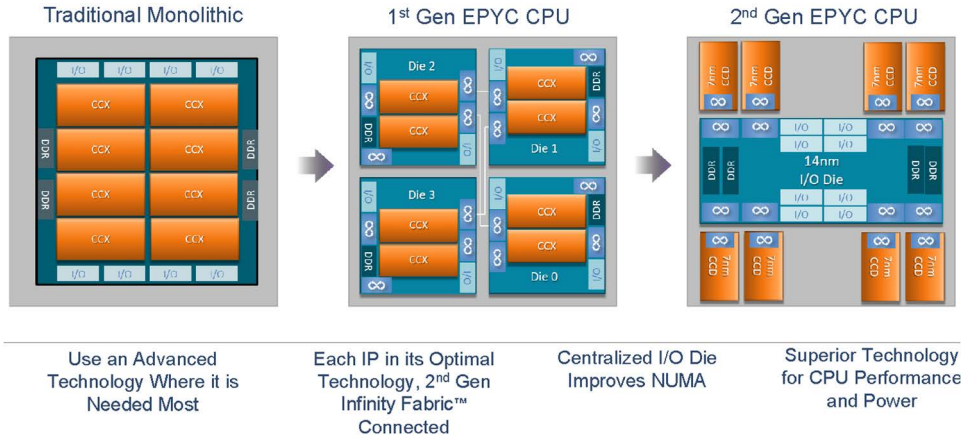


Figure 1.9B: Chiplets in Heterogeneous Integration - EYPC Server Processors. Source AMD

Components of different nodes or from different companies may be heterogeneously integrated together in one SiP, such as implemented in Intel Kaby Lake G card incorporating an Intel CPU with AMD GPU linked to 4GB of HBM2. The silicon bridge on organic substrate (EMIB) developed by Intel was used to link multiple die together in close proximity in the package [3].

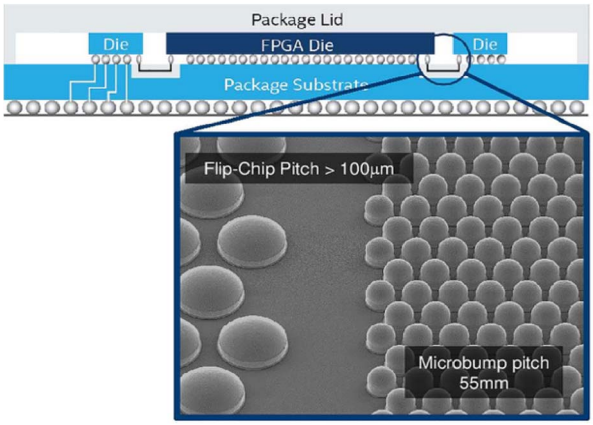


Figure 1.10: Silicon bridge on substrate technology (EMIB) from Intel for Heterogeneous Integration. Source Intel

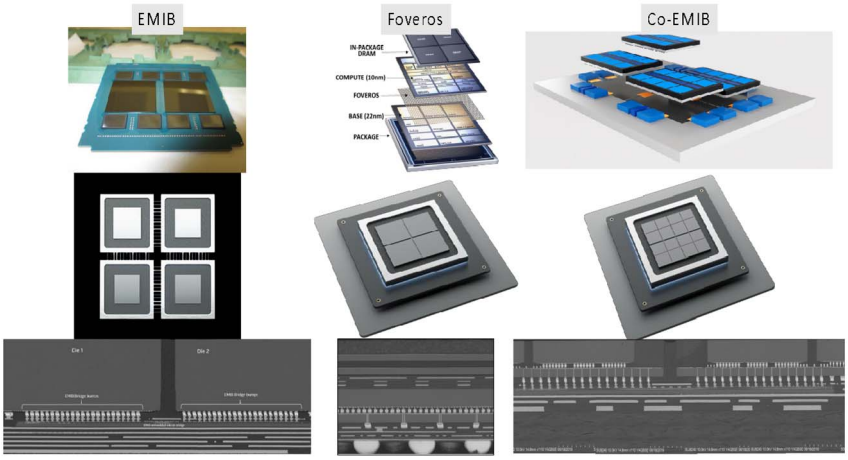


Figure 1.11: Advance Package Chiplet Integration Technologies at Intel: EMIB, Foveros and Co-EMIB. Source Intel

Extending from silicon bridge on organic substrate (EMIB), Fevoros and Co-EMIB form the advanced technology family in Intel to link multiple die (chiplets) on substrates in the package.

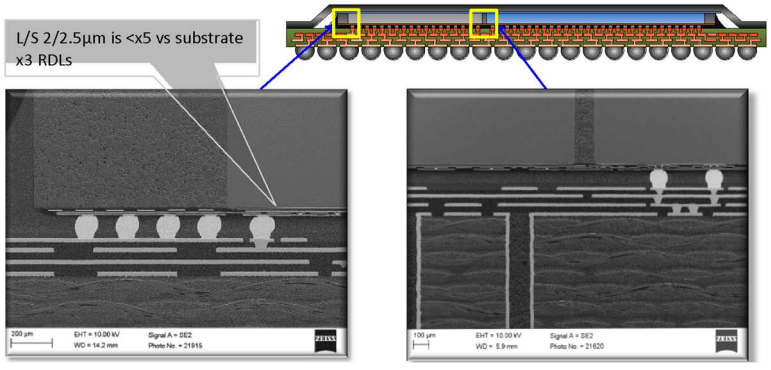


Figure 1.12: Wafer Level Fan-Out integrating 14 nm & 22 nm die in multichip package. Source ASE

Wafer Level Fan-Out technology was initially developed for addressing the WLCSP form factor for BGA balls. The same manufacturing infrastructure has been utilized to integrate two die from dissimilar nodes into one multi-die package as shown above.

The previous examples demonstrate the growing momentum for SiP and chiplets in high performance computing: taking a new look at system architecture and expanding innovations in our packaging toolbox. We are now seeing the "chiplet initiative" utilizing different packaging technologies for disparate system applications. As the industry goes further into single-digit nodes, considerations of cost and time to market trading off with high bandwidth and performance per watt considerations, multi-die (chiplets) heterogeneous integration will be an important trend going forward [4-10].

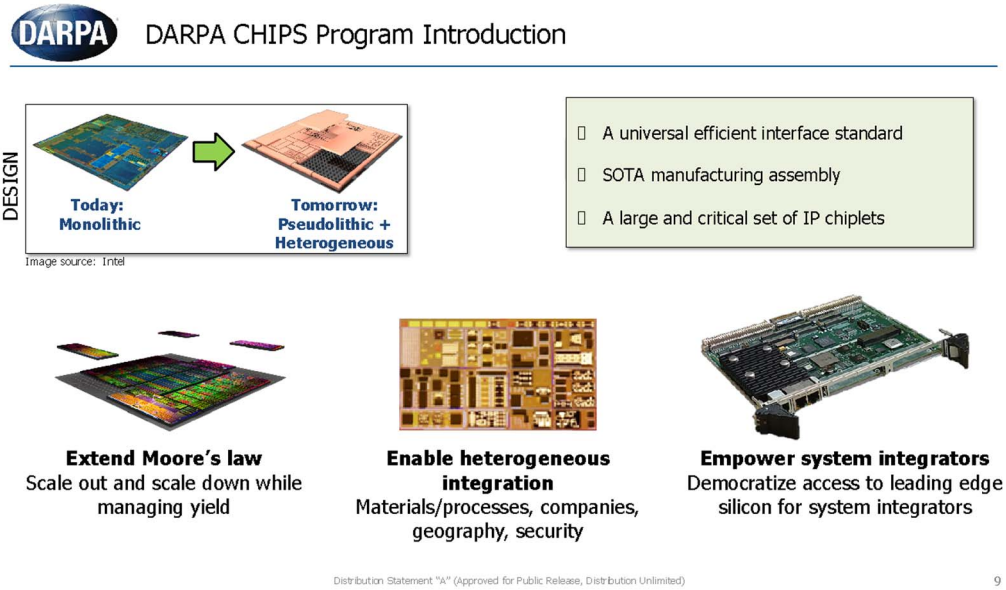


Figure 1.13 Chiplets in DARPA CHIPS Program. Source DARPA

At SEMICON West July 10th, 2019, Andreas Olofsson, DARPA Microelectronics Technology, presented a talk illustrating Chiplet in the DARPA CHIPS program. The Chiplets program is aiming for three goals (a) extend Moore's Law (b) enable heterogeneous integration and (c) empower system integrators.

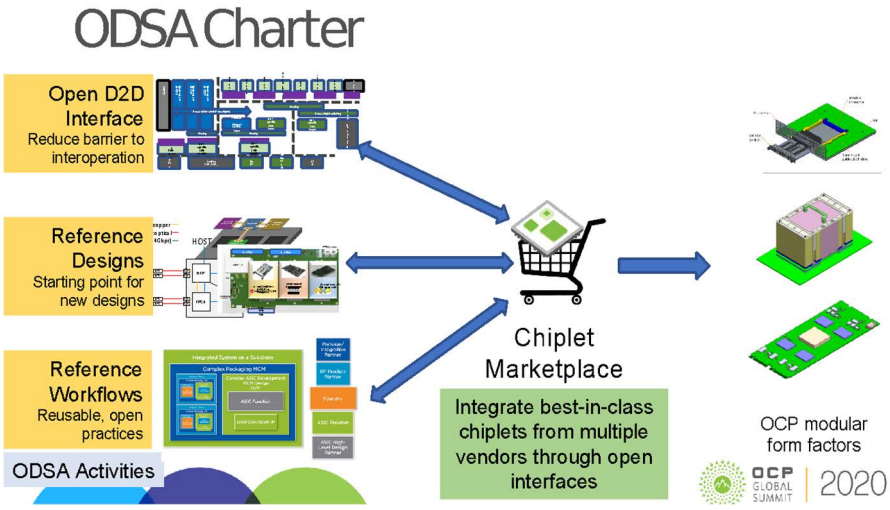


Figure 1.14. Chiplets in Commercial Perspective. Source: ODSA

At the OCP Summit May 12, 2020 and the HIR Workshop February 21, 2020, Bapi Vinnakota presented the ODSA charter which projects a vision of commercial perspective of chiplets with a commercial chiplets marketplace enabling integration of best-in-class chiplets from multiple sources in heterogeneous integration for commercial electronics products.

The scope of the Single Chip and Multichip chapter is designed to understand the current state of the art and to ask the questions: What is the status of leading edge technologies in our tool box for SiP and chiplets and

heterogeneous integration? What are the challenges looking into the future? What are the potential solutions? We address these questions, and include the electrical, thermal and mechanical technology issues from device packaging to subsystem and system packaging, from system-package-device co-design to manufacturing, inclusive of the total ecosystem. Following Moore's words, our purpose in heterogeneous integration is to build large systems out of smaller functions – chiplets and system-in-package (SiP) – which are separately designed, packaged, interconnected and manufactured. This chapter starts with a section on Scope followed by 11 sections on key technology building blocks from the Knowledge Base and Data to Manufacturing and Physical Infrastructure.

- Knowledge Base & Data
  - Electrical Analysis & System Requirements
  - Thermal Management
  - Mechanical Analysis
  - Electromigration
  - Reliability
- Manufacturing & Physical Infrastructure
  - Wafer Singulation and Thinning
  - Wire bond
  - Flip Chip
  - Substrate
  - Board Assembly
  - Additive Manufacturing

The Knowledge Base, and Data and Manufacturing and Physical Infrastructure, form the base foundation for Advanced Packaging and Integration technologies from SiP to Wafer Level Packaging, 2D, 2.5D and 3D. Working together with other TWGs in the Heterogeneous Integration Roadmap Initiative, we describe the current state of the art, and the roadblocks in the path going forward, to stimulate pre-competitive research and innovations 15 years ahead.

Acknowledgments

We appreciate very much the leadership team in making this chapter a reality for our profession and our industry.

TWG Chair: William Chen; Co-Chair: Annette Teng

Electrical Requirements	Lei Shan (Tekollect)	Wafer Singulation & Thinning	Annette Teng (Promex)
Wirebond	Ivy Qin (K&S)	Substrate	Ky-Oh Lee (Intel)
Board Assembly	Jim Wilcox (UIC)	Flip Chip	Mark Gerber (ASE)
Thermal Management	Baghat Sammakia (BU), Srikanth Rangarajan (BU)		
Mechanical Requirement	Benson Chan (BU), SB Park (BU)		
Additive Manufacturing	Kris Erickson (Facebook)		
Electromigration	Paul Ho (UT), Valeriy Sukharev (Mentor)		
Reliability	Abhjit Dasgupta (UMD), Richard Rao (Marvel), Shubhada Shasrabudhe (Intel)		

A very important announcement is that our Reliability Section is being spun off to form a separate Technical Working Group with a separate chapter for the 2022 Roadmap. We wish them the best in their new venture.

7. References

1. "Cramming More Components onto Integrated Circuits" Gordon Moore, pp 114 – 117 Electronics April 19, 1965
2. Keynote Presentation by Lisa Su, SEMICON West, July 9, 2019.
3. Presentation by R. Mahajan at SEMI 3D Summit January 2019.
4. "Making Chip Packaging Simpler" Ed Sperling & Mark Lapedus, Semiconductor Engineering, March 21, 2019.
5. "Intel's View of Chiplet Revolution" Samuel Moore, IEEE Spectrum, April 12, 2019
6. "Turn to Chiplets" Tom Simonite , Wire, November 6, 2018
7. "Modular Routing Design for Chiplet-based System" Jieming Yin et al, ACM/IEEE 45th International Symposium on Computerware Architecture, 2018 pp 726-738.
8. Open Domain-Specific Architecture Workgroup Working Paper Draft v0.9 2018 titled "Open Domain-Specific Architecture – A chiplet based open architecture"
9. "Waiting for Chiplet Interfaces", Brian Bailey, Semiconductor Engineering, June 13, 2019
10. Momentum Gains for an Open-Chiplet Ecosystem" Dylan Mcgrath, EE Times June 18, 2019

Section 2: Electrical Analysis and System Requirements

1. Introduction

As system integration migrates from "on-chip" to "in-package", I/O signal integrity (SI) and package-level power distribution effectiveness (PI) become essential to sustaining system advancement. The scale of integration varies from application to application, and so do the SI and PI requirements. In this section, three representative applications are selected for the context of electrical analysis, i.e. Memory, Mobile, and High Performance. In general, "Memory" drives the demands of integration density and bandwidth, "Mobile" drives miniaturization and power reduction, and "High Performance" drives the limits of I/O bandwidth as well as integration technologies.

Table 1: Single- and Multi-Chip Integration Technology Requirements

Year of Production	*2019*	*2020*	*2021*	*2022*	*2023*	*2026*	*2029*	*2032*	*2035*
Chip size (mm^2)									
Memory (DDR/HBM)	60/90	60/90	60/90	60/130	60/130	60/130	60/130	60/130	60/130
Smart Phone / Laptop	80/130	80/130	80/130	80/130	80/130	80/130	80/130	80/130	80/130
High-performance (note1)∏ chiplet/monolithic	500	500	200/500	200/500	200/500	300/600	300/600	400/600	400/600
Maximum Average Power Density (W/mm^2)									
Memory (DDR/HBM) (note2)	0.06/0.12	0.06/0.12	0.06/0.12	0.05/0.11	0.05/0.11	0.04/0.1	0.04/0.09	0.04/0.08	0.04/0.07
Smart Phone / Laptop	0.45	0.45	0.45	0.5	0.5	0.55	0.6	0.65	0.7
High-performance	0.9	0.9	0.9	0.95	0.95	1	1.05	1.1	1.15
Core Voltage (Minimum Volts)									
Memory (DDR/HBM)	1.1/1.2	1.1/1.2	1.1/1.2	1/1.1	1/1.1	0.9/1	0.8/0.9	0.75/0.8	0.75/0.75
Smart Phone / Laptop	0.85	0.85	0.85	0.8	0.8	0.8	0.75	0.75	0.75
High-performance	0.85	0.85	0.85	0.8	0.8	0.8	0.75	0.75	0.75
Package Pin count Maximum									
Memory (DDR/HBM)	288/3200	288/3200	288/3200	288/3200	288/3200	326/4100	326/4100	350/4700	350/4700
Smart Phone / Laptop	1155/5260	1155/5260	1155/5260	1212/7000	1212/7000	1275/7000	1396/9600	1444/9600	1588/11000
High performance (note3)	6400	6400	6400	7800	7800	9600	9600	11200	11200
Minimum Overall Package Profile (mm)									
Memory (DDR/HBM)	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Smart Phone / Laptop	0.19/0.75	0.19/0.75	0.19/0.75	0.19/0.85	0.19/0.85	0.19/0.85	0.19/0.95	0.19/0.95	0.19/0.95
High-performance	1.5	1.5	1.5	1.5	1.5	2	2	2	2
Performance: On-Chip									
Memory (DDR/HBM), MHz	800	800	800	800	800	1200	1600	2000	2400
Smart Phone / Laptop, GHz	2.8/6.4	2.8/6.4	2.8/6.4	3.2/8.0	3.2/8.0	4/9.6	5.2/11.2	6.4/12.8	8.0/16
High-performance, GHz	6.4	6.4	6.4	8	8	9.6	11.2	12.8	16
Interconnect: Chip-to-Chip (note4)									
Memory (DDR/HBM), *Gb/s*	6.4/2.0	6.4/2.0	6.4/2.0	8.0/4.0	8.0/4.0	9.6/8.0	12.8/12.8	16/16	25/25
Smart Phone / Laptop, *Gb/s*	50	50	50	100	100	100	200	200	200
High-performance, *Gb/s*	56	56	112	112	112	224	224	224	224
Interconnect: Pkg-to-Board									
Memory (DDR/HBM), *Gb/s*	6.4/2.4	6.4/3.2	6.4/3.2	8/6.4	8/6.4	9.6	12.8	16	25
Smart Phone / Laptop, *Gb/s*	50	50	50	100	100	100	200	200	200
High-performance, *Gb/s*	56	56	112	112	112	224	224	224	224
Maximum Junction Temperature									
Memory (DDR/HBM)	85	85	85	85	85	85	85	85	85
Smart Phone / Laptop	90	90	90	90	90	90	90	90	90
High performance (Note 5)	70	70	70	70	70	70	70	70	70
Operating Temperature Extreme: Ambient (°C)									
Memory (DDR/HBM)	45	45	45	45	45	45	45	45	45
Smart Phone / Laptop	45	45	45	45	45	45	45	45	45
High-performance (Note 6)	45	45	45	45	45	45	45	45	45

Manufacturable solutions exist, and are being optimized	
Manufacturable solutions are known	
Manufacturable solutions are NOT known	

Table 1 is a brief summary of the metrics relevant to system requirements – particularly signal and power integrity for various application scenarios – as well as the technology trends in both short term (5 years) and long term (15 years) in 3-year intervals. With the continuous decrease in on-chip feature dimensions, high-end semiconductor manufacturing cost increases exponentially with chip size due to yield impact, and therefore, instead of continuously growing, chips tend to settle at certain optimal sizes, as packaging-level integration delivers cost-effective system performance. Meanwhile, lower voltage and leakage current lead to power reductions in most applications, except for logic cores with ever increasing clock frequency. Particularly, both artificial intelligence and accelerator architectures involve heterogeneous devices/components, which further drives the increase in package dimensions and hence maximum pin count as well as package profile.

In the heterogeneous multi-chip packaging environment, input/output interconnections play an important role in system performance. In addition to data bandwidth and signal integrity, common I/O standards are desirable for the ease of complex system designs with devices/components from various sources. In late 2019, The Open Compute Project (OCP) announced significant progress by the Open Domain-Specific Architecture (ODSA) subproject in the development of a chiplet-based architecture. The ODSA subproject's mission is to define an open interface and architecture that enables the mixing and matching of silicon chiplets from different vendors. To achieve this goal, multiple working groups within the ODSA have been established: The ODSA PHY interface group is tasked with defining a simple, open, flexible data-rate interface between chiplets, by defining a new low-power Bunch of Wires (BoW) interface for low-cost packaging technologies; The ODSA Proof of Concept (PoC) group is tasked with validating the technology proposals from the program; The ODSA Business Working Group is tasked with defining a workflow and business processes to enable companies to assemble products from the marketplace.

More interconnect options and specifications are to be detailed in the following sections.

1. On-package interconnections

As the semiconductor industry continuously scales down feature size, costs for yielding large dies increases significantly. Compared to 250 mm² die on the 45 nm process, the 16 nm process more than doubles the cost/mm² and the 7 nm process nearly doubles that to 4x the cost per yielded mm². Moving to the 5 nm and even 3 nm nodes, the cost is expected to continue to increase. Fabricating large monolithic dies will become increasingly less economical. As one solution for easing the economics of manufacturing chips with a large number of transistors, the industry has started shifting to chiplet-based design whereby a single chip is broken down into multiple smaller chiplets that are "re-assembled" at the package level, which demands significant interconnect bandwidth. In addition, other heterogeneous components, such as HBM, GPU, and FPGA, are to be integrated in the package simultaneously. The scale and complexity of SiPs requires greater carrier dimensions as well as higher interconnect density, which in turn drives the development of innovative packaging solutions.

In addition, to take full advantage of multi-chip packaging, it is critical to provide high bandwidth, low latency connections among functional components. Specifically, for connecting multicore processor die with stacked memory dies, point- to-point interconnections are needed and the number of memory dies will be proportional to the number of cores on the processor die. Conservatively assuming that core counts scale by a factor of 1.4X per generation, 1.4 times as many memory dies need to be accommodated per generation in the SiP. Simultaneously, if we assume that advances in the stacked memory technologies enable twice as many data bits to be delivered per generation and assuming that the clock rate on the processor-memory link remain unchanged, the number of bit links between the multicore die and the stacked memory dies will have to grow by a factor of 2.8X with each process generation.

As an example, at 14 nm, Intel implements 1024-bit-wide bit links as EMIBs (embedded multi-die interconnection bridge) on a silicon substrate to each HBM inside the SiP with a core count of 56. When the transition is made to hyperscaled 10 nm, the core count grows to 78 (=56 X 1.4), requiring 2048-bit wide links to each HBM and the ability to connect to 1.4 times as many HBMs. This will require finer interconnection pitches in the EMIB or other enhancements that will require additional metal layers (beyond the 4 to 6 metal layers in use now on the silicon bridge) and additional vias in the EMIBs, or alternative on-package interconnection techniques. In general, the on-chip interconnection problem may be exacerbated when dies integrating general-purpose cores and accelerators are integrated with other components, as off-die connections may be grossly limited by the physical dimensions of the die. Table 2 includes commonly used PHY options and corresponding specifications.

Table 2: Die-to-Die PHY Options in Advanced Process Nodes (R. Horner, OPC-ODSA)

	Parallel Interface		Serial Interface (SerDes)
	Non-Memory	Memory	
Standards & Specifications	DARPA, OCP-ODSA AIB, HBI, BoW	JEDEC HBM2, HBM2E HBM3	IEEE 802.3, PCI-SIG, OIF 1G - 56G & 112G URS/XSR
Data Rate per Lane (Gbps)	1 to 2 → 2 to 4 → 4 to 6	2.4 → 3.2 to 3.6 → 6.4	1.25 to 112
I/O number per link	30 to 2000+	1024	2 pairs (4) ✓
Latency	Low ✓	Low ✓	High
Interconnect reach	Short	Short	Long ✓
Interconnect medium type need	High density routing (Silicon Interposer)		Low resistance (High Density Fan-Out)

3D integration can also be a promising solution if thermal, yield and reliability issues are addressed to permit stacking of memory dies and processor dies. This will be more realistic for stacking lower power, energy-efficient integer cores targeted to specific data center applications with DRAM memory/HBM dies.

A possible SiP solution in the HPC/Data Center market will be to use tiling to decompose a large/low-yield die, such as a multicore CPU, into smaller homogeneous dies (which will have a higher yield), and build appropriate interconnections among them to realize the same throughput as the larger die. Other advantages of this approach will include the ability to distribute the heat load, efficiently distribute power, do microarchitectural innovations, etc. The possible solutions for addressing these needs are as follows:

Short-term (0 - 5 years):

As mentioned above, driven by the demands of on-package performance and functionality scaling, high bandwidth interconnects are experiencing an explosive growth. Both I/O speed and density have resulted in a large variety of proprietary I/O standards. These short-term solutions will temporarily satisfy the emerging demands.

- **2.5D integration; Si-interposer and EMIB:** Embedded Multi-die Interconnect Bridge (EMIB) is an approach developed by Intel for in-package high density interconnect of heterogeneous chips. The industry refers to this application as 2.5D package integration. Instead of using a large Si interposer typically found in other 2.5D approaches (like TSMC's CoWoS and Unimicron's embedded interposer carrier), EMIB uses a very small bridge die with multiple routing layers, but without TSVs. This bridge die is embedded as part of Intel's substrate fabrication process. With further improvement and broader applications, EMIBs will continue to play a dominant role in the near future with enhancements in the choice of organic materials, number of metal layers, and improved driver/receiver circuitry for signal integrity enhancements.

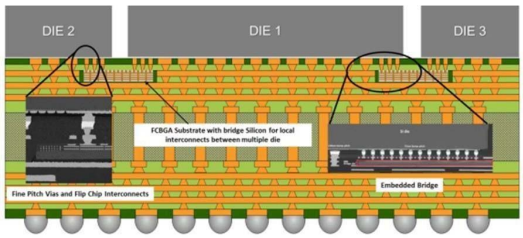


Figure 1: Embedded Multi-die Interconnect Bridge (EMIB) [1]

- **High-density organic substrate:** By combining with thin film processes, high-density flip-chip organic packaging is emerging as a potential integration carrier. 8/8um line/spacing and <50um via pitch will soon be commercially available at reasonably low cost, and 2/2um line/spacing is projected on the five-year roadmap. Various solutions are proposed, and there will be multiple options to choose as a substitution for a silicon interposer and/or EMIB-like hybrid. Even though there are still gaps – particularly line width/spacing – compared with silicon technologies, organic substrates are much easier to design and cost much less. On the other hand, fine lines may cause RC delay due to high line resistance, as on silicon chips, and therefore there is an optimal line width number, which is roughly between 2 and 5 um.

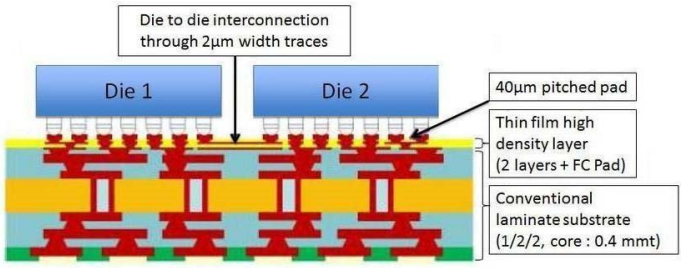


Figure 2: Emerging high-density organic substrate [2]

- **3D integration on the horizon:** At the end of 2018, Intel announced a 3D chip stacking technology, called FOVEROS. It utilizes a large silicon carrier to integrate multiple chips, and differs from a silicon interposer by incorporating active devices into the silicon carrier. This is a breakthrough, since the development of silicon-level 3D integration for high-performance systems slowed down due to thermal and power delivery issues.

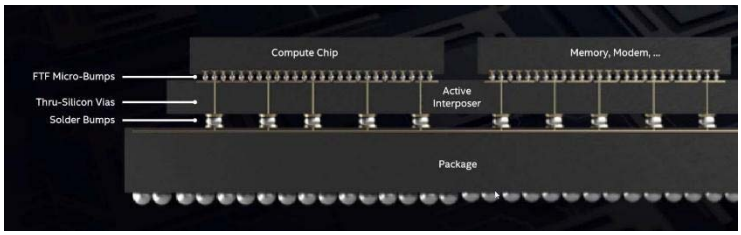


Figure 3: FOVEROS 3D chip stacking by Intel [3]

- **Ceramic-Based Heterogeneous Carrier (CBHC):** Ceramic substrates had been widely used to integrate multi-chip modules for decades until being gradually replaced by organic laminates due to continuous advancements of semiconductor technology. Now that system integration comes back to the package level, leading ceramic companies, such as NTK technologies, are developing a Ceramic-Based Heterogeneous Carrier (CBHC) by taking advantage of both ceramic and organic materials. Such low-cost, large dimension, low-CTE, BA friendly, reliable, and reworkable heterogeneous substrate technology is expected to become commercially available in the near future.

Longer term (5 – 15 years):

Package-level 3D integration will address the demands for performance and miniaturization, which will also be more effective in terms of scalability and cost, together with the following upcoming technologies:

- Integrated photonics with polymer waveguides and improved optical transceiver stability;
- Plasmonic interconnections;
- Optical vias for 3D integration;
- Other technologies, protocol-specific, such as embedded components.

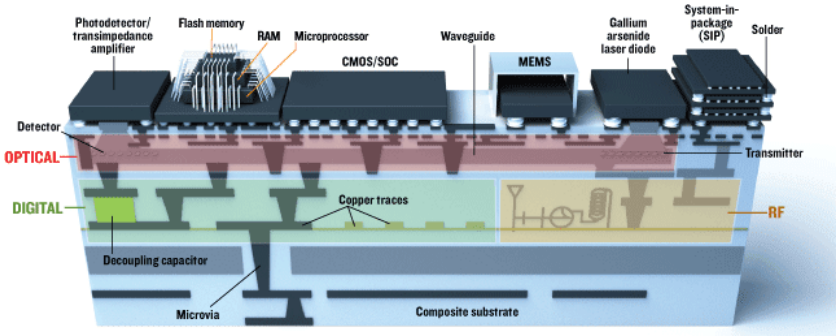


Figure 4: 3D heterogeneous integration

Very similar scaling rules apply to the point-to-point interconnections between GPU dies and stacked memory dies or between special function FPGA dies and stacked memory dies.

Connections to off-package interfaces and DRAM controllers on the SiP substrate can continue to rely on the PCIe standard, and the evolution path for multi-lane PCIe has been well-defined. The implementation of alternatives to direct links based on point-to-point interconnection technologies will require multiple metal layers in the silicon substrate and the exact topologies used are specific to the SiP architecture. Signal integrity needs for longer links in

the substrate, symbol encoding, and clock synchronization issues have to be addressed here. If higher speed serial links are used, the silicon-imposed limits on SERDES have to be observed. Photonics links will be a viable interconnection alternative for implementing high data rate, relatively longer links on the substrate, but this will require significant advances to be made for realizing low-power emitters whose wavelength drifts are limited with temperature variations, as well as the design of reliable detectors.

2. Off-package interconnections

As additional components are integrated within a single package, the demands on the off-chip interconnections go up commensurately with the number of processing elements that are integrated. The newer generation of PCIe links can possibly meet these needs, but the ultimate limitation will be imposed by the package pinout. As an example, when 1.4X more cores are accommodated on a multicore die, the off-package link count will need to go up commensurately. With a limit on the pin count, this need can be met by increasing the link data rate and multiplexing multiple logical links on a single physical link. Photonics links can be an alternative to copper links as techniques like wavelength division multiplexing can be used to implement several connections concurrently on a single photonic link. Here again, limiting wavelength drifts become critical.

Possible Solutions

A. Future-generation links:

Package-level system integration tends to blur the line between on-package and off-package I/O. Many I/O standards are commonly used for both I/O scenarios. PCIe is one of the most popular I/O standards, and it takes over four years for each generation evolution (doubling of the data-rate). However, as PCIe Gen4 had barely settling down in 2017, the industry had already started searching for solutions for PCIe Gen5, which is a clear indication of package-level system integration advancement. PCIe Gen5 is expected to carry 32Gbps per data channel without changing the TX/RX specifications. IBM and Amphenol Corporation jointly developed a new PCIe connector and demonstrated PCIe Gen5 bandwidth in early 2018, which significantly accelerates the availability of the new standard. The upcoming PCIe Gen6 will further double the data rate to 64Gbps by adopting PAM4 signaling.

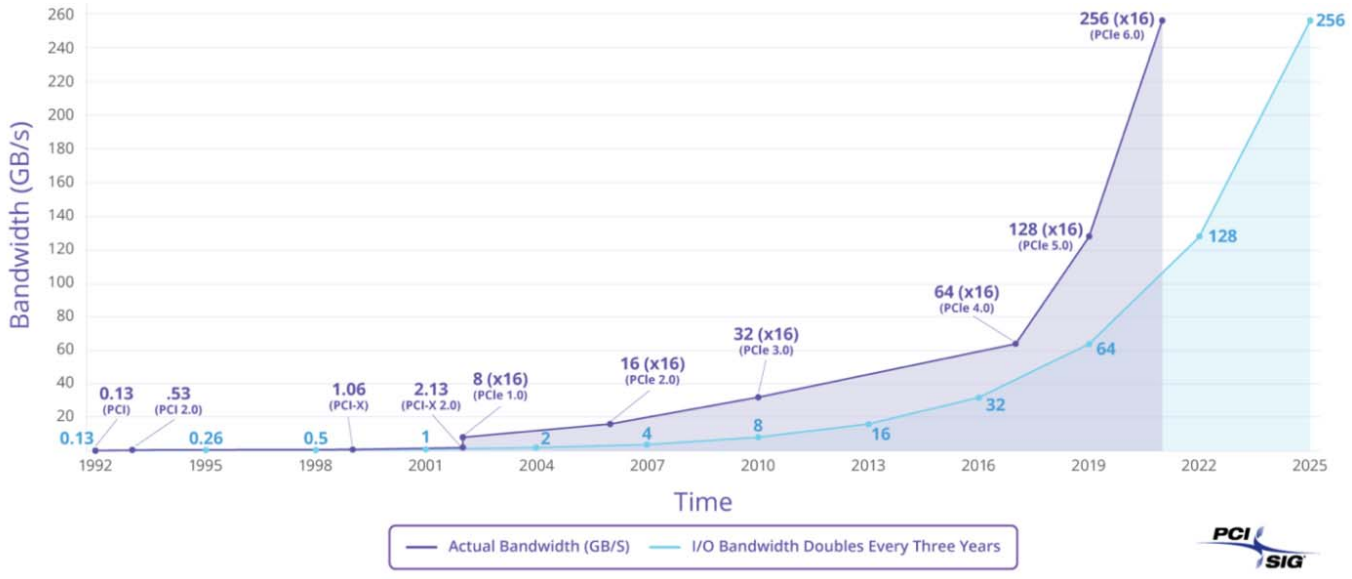


Figure 5: PCIe standard evolution [5]

Driven by package-level integration, numerous proprietary I/O standards have been emerging in recent years, such as GenZ, Omni-Path, NVLink, etc. Most are evolving towards 32Gbps in the next couple of years. Table 3 shows SERDES I/O speed, distance, and channel topologies. Off-package 56Gbps data-rate is expected by 2020 with PAM4 signaling.

Table 3: SERDES I/O speed, distance, and channel topologies (compiled from various sources)

Parameter	Ultra Short Range	Very Short Range	Short Range	Mid Range	Long Range
Data Rate, Gbps	32 – 112	32 – 112	32 – 112	32 – 112	32 – 112
Bit Error Rate	1E-15	1E-15	1E-15	1E-12	1E-9
Distance, cm	1	5	15	50	100
Interconnect	MCM	PCB + 0 connector	PCB + 1 connector	PCB + 1 connector	PCB + 2 connector
Insertion Loss, dB @ f_N	3	6 (PAM4), 14 (NRZ)	15 (PAM4), 35 (NRZ)	30 (PAM4)	27 (PAM8), 45 (PAM4)
Modulation	NRZ	NRZ or PAM4	NRZ or PAM4	PAM4	PAM4 or PAM8
Forward Error Correction	N	N	Y	Y	Y

B. Electrical/optical "flyover" cabling:

To mitigate the impairments of via and solder joint transitions in package and PCB, direct "flyover" cabling as shown in Figure 6 is another option for scaling channel data-rate from the current 28Gbps up to 56Gbps and 112Gbps. Compared with conventional interconnects, channel loss may be significantly reduced. However, the number of channels is limited by cable flexibility, and multiple cable arrangement is much more complex than wiring in PCBs. Therefore, it will likely be a supplemental solution for long-reach interconnections.

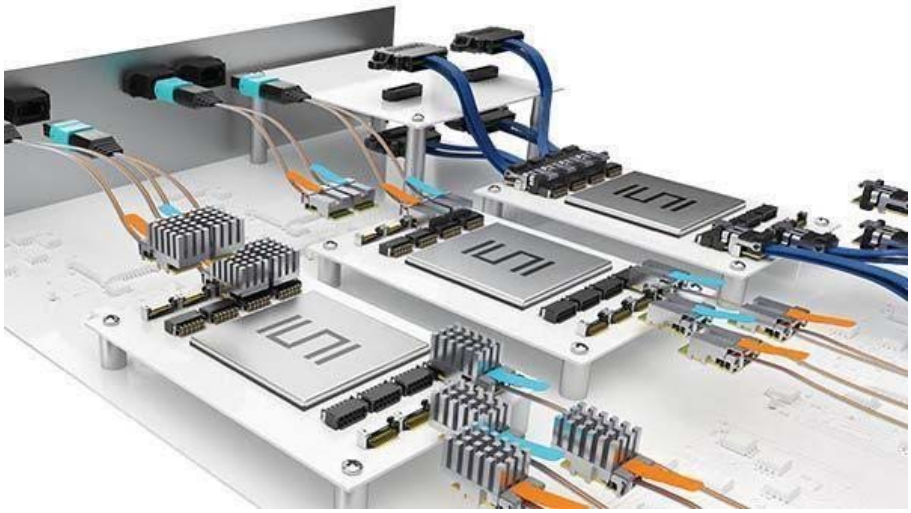


Figure 6: Electrical/optical "flyover" cabling [6]

1. Signal integrity issues

In general, to exploit the capabilities of a SiP without I/O bottlenecks, dense parallel connections need to be used on-package, and higher-bandwidth off-package connections operating at very high link rates become a necessity. These certainly introduce potential signal integrity problems that need to be dealt with adequately. Powerful error correction capability going beyond ECC will be necessary for critical on-package connections, and alternative symbol encoding and signal processing necessary for recovering data waveforms for off-chip links may well become the norm in very high-end, high-availability SiPs.

With growing data-rate, both loss and crosstalk increase significantly, and channel signal integrity can be compromised. Therefore, new materials, connectors/sockets, and via transitions are required to achieve link specifications. For dielectric materials, 3-4 times lower dielectric loss (compare with FR4, tanδ=0.22) will be widely available, combined with smooth copper foil to mitigate skin effects. Meanwhile, a low dielectric constant (<3.2) may help reduce within-layer channel-to-channel crosstalk. For via transitions, via-stub removal by using blind vias or backdrilling is critical, and a smaller via diameter may be needed for via impedance control and crosstalk reduction. Further, signal conditioning and equalization will be widely adopted to compensate for excessive loss, ISSI, and crosstalk. For data rates beyond 50Gbps, PAM4 signaling will prevail, for much lower Nyquist frequency.

2. Power integrity issues

Power distribution and power quality issues become dominant as more components that operate at lower voltages (sub one Volt or close to a Volt) are integrated. In the extreme case, assuming a 200 Watt package TDP, if these components dissipate 70% of the package power (that is, 140 Watts), the current draw from the regulated source will be around 140 Amps. With many components drawing high levels of current that are placed at different positions on the substrate, a larger number of pins needs to be devoted to the power connections. Worse, inductive noise on the power connections will be significant, affecting power quality and requiring additional decoupling capacitors. Additionally, Ohmic losses may be non-negligible, affecting the overall energy efficiency.

A potential remedy for these issues will be to incorporate local voltage regulators within the package itself as a separate integrated component, but adequate cooling will need to be provided. Inductorless integrated switched-capacitor regulator technologies have certainly evolved and can be operated in a distributed configuration to provide point-of-load regulation; these are a strong contender as the best solution, whether used intra-die or intra-package. Complementing these solutions, distributed point-of-load power regulators implemented in the mainstream CMOS process technologies, that enable DVFS control and have a low setting time, appear to be an attractive solution at the die level. The microprocessor industry has been using distributed regulators on the die for the past few years and SiP-level solutions extending these are thus viable for meeting short-term needs.

Advances in low-loss discrete switching devices for power electronics (such as SiC, GaN) are likely to permeate the SiP product spectrum and offer improved efficiency, reliability and availability in power distribution systems for emerging and future SiPs.

A final solution that has the potential for scaling well with SiP complexity will be to use distributed regulators within the package that operate at higher input DC voltage and regulate down in a distributed configuration to the sub one Volt or one Volt region as needed. This solution will certainly reduce Ohmic losses on the power connections but their benefit in terms of reducing inductive noise is not clear and may not be commensurate with the reduced current draw on the power lines to the package.

Issues & Challenges

- High-performance processing chipset power rating: 300W
- High-performance graphic chipset power rating: 400-500W
- Sub-volt power supply (0.85V), maximum switching current ~300A, requires >100uF on-chip capacitance for less than 10% voltage variation
- Bring regulator closer to the silicon die
- Multi-level decoupling
- On-package embedded capacitor/capacitance and inductor
- Operation coding for lower simultaneous switching current

3. Global power and thermal management

The various components integrated onto a single substrate in a SiP can each have their own power management strategy. A global power management scheme is essential to synergistically manage the power dissipation of all integrated components to not only stay within the package TDP but also to address any inevitable hot spots that may result. There are several ways to implement a global power management scheme, and all require the ability to sense temperature and the power dissipated within key blocks of the various dies. A dedicated controller for power management may be needed, similar to the PMU microcontrollers used in many multicore processor chips. Several power management policies are possible that use static or dynamically allocated power budgets. PMUs implementing machine learning-based global power and temperature management are also possible. This is an open area of research and may well dictate the standardization of sensor and actuator interfaces for each integrated component, including voltage regulators inside the package.

References

1. EETimes, "Intel Shows 2.5D FPGA at ISSCC", By Rick Merritt, February 2017
2. ARS Technica, A. Usman, et, al, "Interposer Technologies for High Performance Applications", IEEE Transactions on Components, Packaging, and Manufacturing Technology pp99, March 2017
3. "Intel introduces Foveros: 3D die stacking for more than just memory", by Peter Bright, December 2018
4. Tummala, R., “Moore’s Law Meets Its Match”, IEEE Spectrum, 43(6), June 2006.
5. Smith R., “PCI Express Bandwidth to Be Doubled Again: PCIe 6.0 Announced, Spec to Land in 2021”, in AnandTech, June 18, 2019
6. Samtec Micro FireFly Flyover System (samtec.com)

Section 3 Thermal Management

While there is a separate Thermal Management Technical Working Group (Chapter 20), this section refers specifically to thermal issues closely related to Single and Multichip Integration. We invite the readers to refer to Chapter 20 for further insight into the difficult challenges and potential solutions for thermal management in electronic systems and to explore the science and technology for the Heterogeneous Future.

The Longevity of Moore's law has been under constant debate [1]. Many packaging experts still believe Moore's law would hold good for at least the next 10 years due to recent technological efforts to combine functionalities horizontally and vertically into a single module, commonly referred to as heterogeneous integration (HI). Heterogeneous integration (HI) is the basis of future-generation computer systems and refers to the assembly and placing of multiple separately manufactured components onto a single module to improve functionality and enhance operating characteristics. In such applications, heat sinks are typically designed to cool multiple dies with different powers, sizes, heights, and allowable maximum temperatures in specified arrangements on electronic boards simultaneously. For several decades, electronic industries and relevant academic research communities globally have made constant efforts to develop and commercialize Moore's law[1,2], leading to many technological breakthroughs and a revolution in packaging technologies and skyrocketing thermal management challenges. The thermal management challenges are attributed to a decrease in device size and increased power consumption. With simultaneous shrinkage in size and more heat dissipation, the flux values have increased exponentially, yet the required temperature difference as the thermal budget for thermal management is also reduced or unchanged making it even more challenging for effective thermal management. Thermal management of electronic devices is important to improve the reliability of devices by maintaining silicon junction temperatures below critical safe operating temperatures. While the academia efforts on optimization of thermal solutions have been predominantly streamlined to theoretical studies, the constraints on manufacturability and reliability are not widely considered part of the optimization problem. However, incorporating commercially available cold plates into the optimization process makes the study more pragmatic.

A typical thermal solution is to have a remote separable cold plate/heat sink attached to the chip with a thermal interface material (TIM). The thermal solutions for heterogeneously integrated packages, in general, can be categorized into:

Technology 1 (T1)– Embedding the cooling solution onto the device (two phase)

Technology 2 (T2) – Embedding the cooling solution onto the chip lid/integrated heat spreader (single- and two-phase liquid cooling).

Technology 3 (T3) – remote separable heat sink/cold plate thermal solution (Air and single-phase liquid cooling).

The heat transfer limit of each of the stated technologies are as shown in Figure 1.

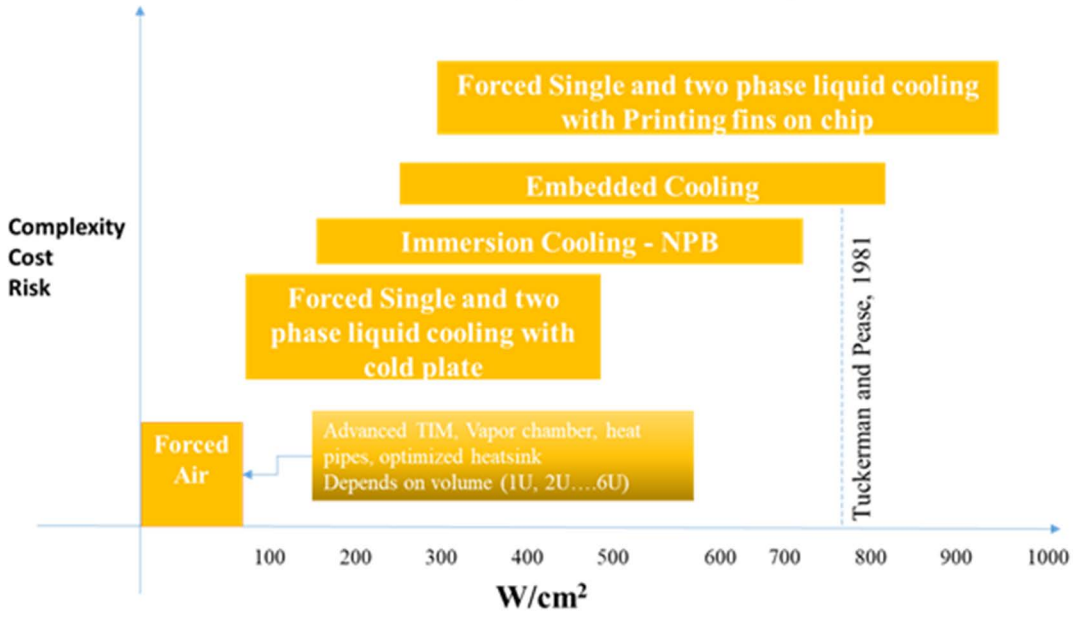
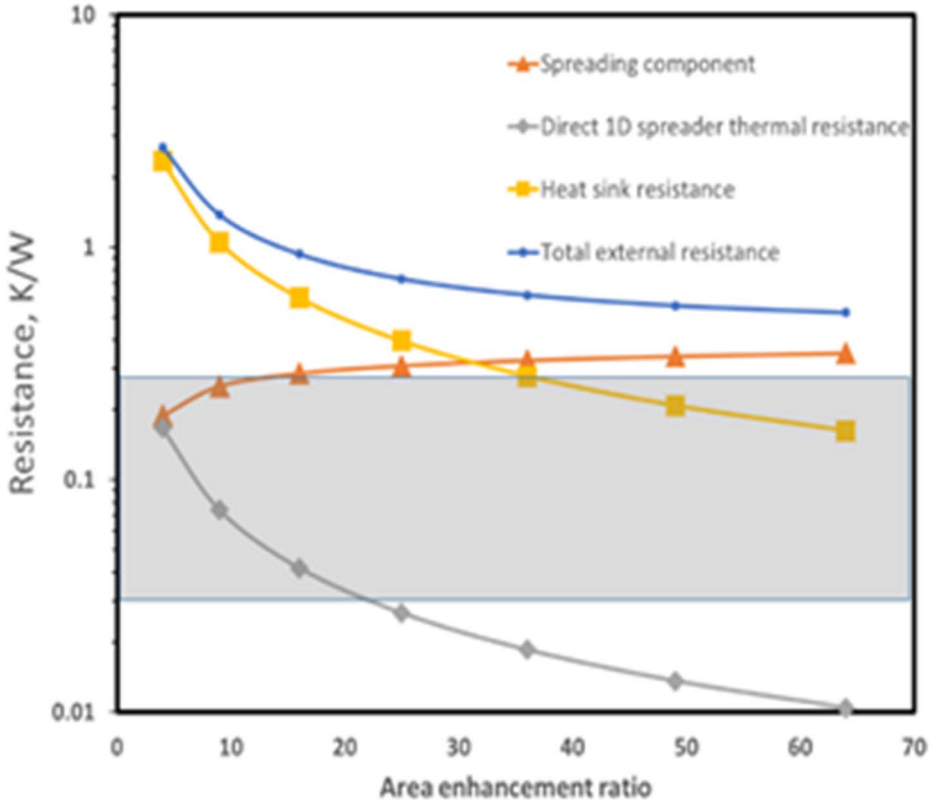
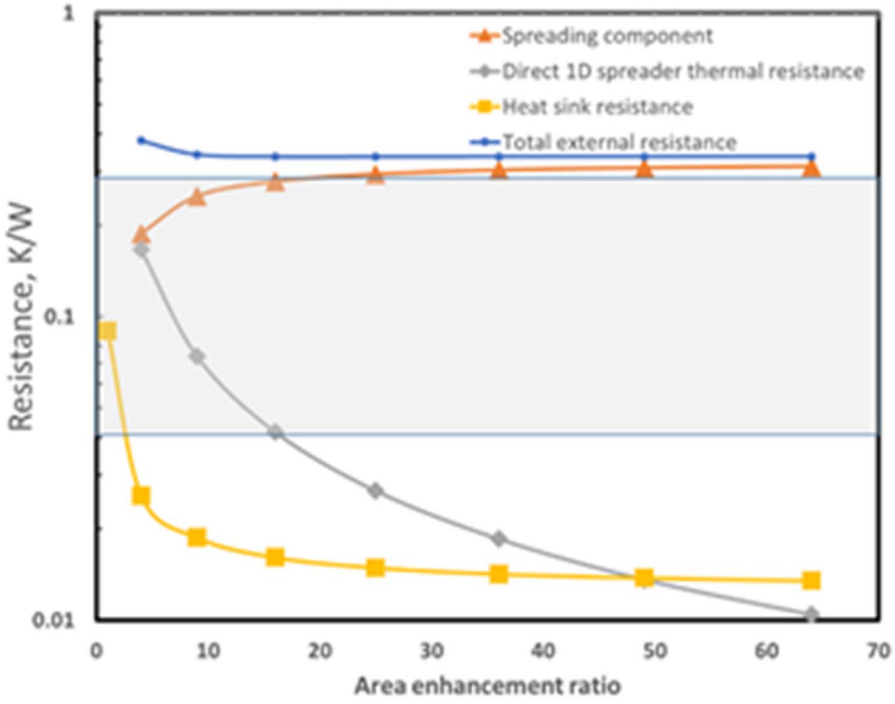


Figure 1: Heat transfer limits of different thermal management technologies

The associated complexity and risk in implementing each of these technologies are marked on the ordinate of Figure 1. In practice for today's ongoing data center power demand, the most common cooling method takes the form of air-cooling [3]. Air-cooling is often claimed to be the most reliable technology that is of practical interest. However, in practice air-cooling suffers from low heat transfer performance and acoustic-related problems. The low heat transfer performance is attributed to low heat transfer coefficient, high inlet air temperature and low thermal spreading owing the requirement for a larger heat sink base area. The noise issue is addressed by imposing an engineering constraint on the maximum fan air velocity [4]. To address the thermal spreading issue, technologies like vapor chambers integrated with heat pipes are employed. The acoustic constraints and the allowable pressure drop dictate the limit on the maximum airflow in the system. The constraint on minimum inlet air temperature is dictated by the Telcordia GR63 standard and ASHRAE for telecom application (55°C) and data center application (45°C) respectively. With regard to easier manufacturing, the plate fin parallel channel heat sinks are still the most widely adopted fin design. With most of the stated factors associated with air cooling having reached standard limits, only the geometric optimization of air-cooled heat sinks can help the research community define the limits of the technology under practice. Unlike air-cooled heat sink modules, liquid cooled heat sinks (commonly referred to as cold plates) can perform extremely well due to the superior thermal properties of liquid water. However, the superior thermal performance trades off with the reliability and technological risks associated with using water closer to the circuit (e.g. unfavorable dielectric strength). In a liquid-cooled module, the heat sink resistance is 10-fold lower than the spreading and the TIM counterparts. Tuckerman and Pease [5] dissipated heat flux up to 790 W/cm^2 for a substrate temperature rise of 71°C by etching a microchannel on the backside of the active device. However, there are practical limitations, challenges, and risks in implementing the work in practice that requires a massive change in the existing infrastructure. Understanding the scaling limits of air and liquid cooling is essential to understand better the potential of embedding the cooling solutions onto the silicon die. Figures 2(a) and 2(b) show the limit of air and liquid cooling with increasing area ratio (AR= Ratio of area of the spreader to the area of the chip). AR=1 translates to a scenario of embedded cooling. The results from the scaling analysis make it evident that the liquid cooling with AR=1 has the minimum total external resistance. Hence liquid cooling with printed fins on-chip with no additional spreader area ideally can extend the heat transfer limit of single-phase liquid cooling. However, air cooling requires a large area for effective heat dissipation and does not perform well when embedded.



(a)



(b)

Figure 2: (a) Tradeoff analysis for air cooling with variable area enhancement ratios, assuming the chip size Ac = 1cm^2, metal conductivity, k = 150W/mK, spreader thickness t = 10mm, wch = 1mm, hch=25mm, Vair= 10 m/s. This figure excludes the TIM resistance. (b) Tradeoff analysis for a single phase liquid cooling scenario for a chip size Ac = 1 cm^2, metal conductivity k = 150 W/mK, Spreader thickness t = 10 mm, wch = 0.05 mm, hch=0.3 mm, Q= 0.5 liters/min. The plot excludes the TIM resistance. [6]

With growing heterogeneous integration of multiple functionalities onto a single chip, there are growing regions of intense hotspots. 3D printing of fins onto the chip [6][7] is seen to be a potential solution to mitigate hotspots and minimize the maximum chip temperature. Vahideh et al [7] numerically demonstrated single-phase heat transfer enhancement of a liquid water-based cooling system employing printed fins directly on chip. The fin material was pure silver. The authors demonstrated that a careful optimization of the pin fin profile and shape could lead to a total chip to ambient thermal resistance of 0.26 °C-cm^2/W (compared to separable cold plate of 0.58 °C-cm^2/W) for a 4cm^2 uniformly heated chip under a constrained pressure drop [7]. The same technique was extended by Vahideh et al [8] to chips with simultaneous hotspots and background heating conditions. A minimum total case to fluid resistance of 0.21 °C-cm^2/W was achieved for four 4cm^2 chips under similar flow conditions. The rationale behind the low resistance was attributed to the fact that each core/hotspot had a dedicated inlet on top, thus minimizing the total resistance's caloric resistance component.

While recent efforts on exploring two-phase liquid cooling using separable cold plates for data center applications have shown remarkable performance [9] [10] in mitigating high heat fluxes with significant energy savings, two-phase cooling is essentially recommended for direct chip cooling using dielectric coolant (Technology 1) [11].

Two-phase immersion cooling could be a potential solution for highly intense and non-uniform heated devices. Boiling enhancement coatings on top of the heated surface is seen to increase the critical heat flux (CHF) and the boiling heat transfer coefficient (hb) simultaneously [12]. 3D printing of fin and microporous structures can be beneficial in promoting nucleation and enable rewetting of the heated surface. The combination of fin and microstructures has been highly beneficial in the simultaneous improvement of CHF and hb[13]. Wong et al [14] printed lattice structures on top of a heat surface, performed pool boiling experiments, and demonstrated a critical heat flux of 107 W/cm^2 corresponding to a boiling heat transfer coefficient of 1.5 W/cm^{2}K.

For all of the listed technologies, the optimization process of heat sinks/ cold plates is as shown in Figure 3.

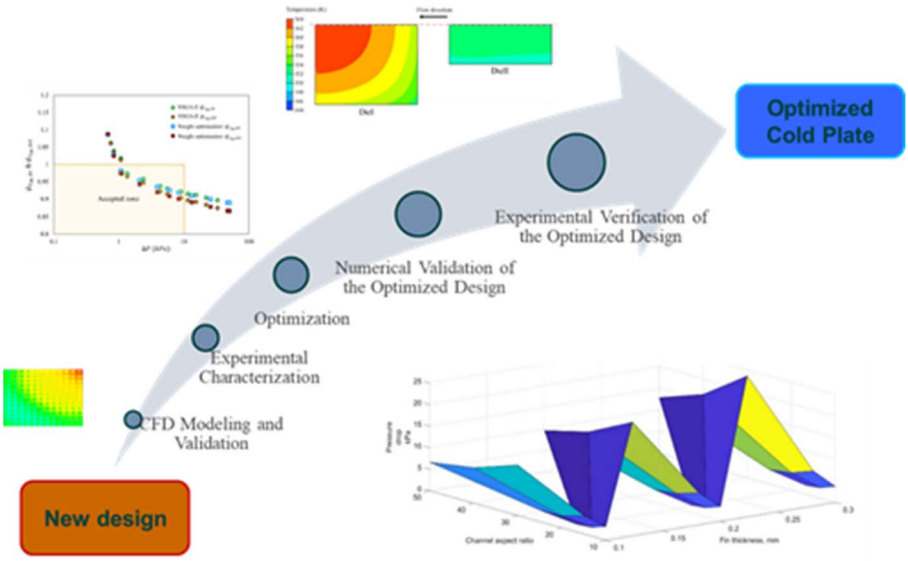


Figure 3: Optimization approach for the cold plate/ heat sink thermal solution

The new design or a working design is in general a commercially available cold plate. The working design is initially characterized numerically and experimentally. Once a benchmarked numerical model of the working design is developed, a robust multi-objective optimization is carried out employing a physics-informed machine learning process. The obtained optimal solutions are then verified numerically and experimentally.

References

1. Courtland, Rachel. "Transistors could stop shrinking in 2021." IEEE Spectrum 53, no. 9 (2016): 9-11.
2. ITRS, The International Technology Roadmap for Semiconductors, 2015, Semiconductor Industry Association,
3. Ellsworth Jr, Michael J., and R. E. Simons. "High powered chip cooling—Air and beyond." Electron. Cooling 11, no. 3 (2005): 14-22.
4. G. Refai-Ahmed et al., "Establishing Thermal Air-cooled Limit for High Performance Electronics Devices," 2020 IEEE 22nd Electronics Packaging Technology Conference (EPTC), Singapore, Singapore, 2020, pp. 347-354, doi: 10.1109/EPTC50525.2020.9315139
5. Tuckerman, David B., and Roger Fabian W. Pease. "High-performance heat sinking for VLSI." IEEE Electron device letters 2, no. 5 (1981): 126-129.
6. Rangarajan s, Schiffres S, Sammakia Bahgat, "Scaling limits, challenges, opportunities in embedded cooling," Encyclopedia of Thermal Packaging: Thermal Packaging Tools/Ed.-in-chief Avram Bar-Cohen, World Scientific, 2021.
7. Radmard, Vahideh, Arad Azizi, Srikanth Rangarajan, Najmeh, C. Hiep Hoang, Ghazal, Charles Arvin, Kamal Sikka, Scott N. Schiffres, and Bahgat Sammakia. " Performance Analysis of Impinging Chip-Attached Micro Pin Fin Direct Liquid Cooling Package for Hotspot Targeted Applications" In 2021 36th Semiconductor Thermal Measurement, Modeling & Management Symposium (SEMI-THERM), pp. 1-9. IEEE, 2020.
8. Radmard, Vahideh, Arad Azizi, Srikanth Rangarajan, Najmeh, C. Hiep Hoang, Ghazal, Charles Arvin, Kamal Sikka, Scott N. Schiffres, and Bahgat Sammakia. "Direct Micro-Pin Jet Impingement Cooling for High Heat Flux Applications." In 2020 The Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (IEEE ITherm) (To appear)
9. Hoang, Cong Hiep, Srikanth Rangarajan, Sadegh Khalili, Bharath Ramakrisnan, Vahideh Radmard, Yaser Hadad, Scott Schiffres, and Bahgat Sammakia. "Hybrid microchannel/multi-jet two-phase heat sink: A benchmark and geometry optimization study of commercial products." International Journal of Heat and Mass Transfer 169 (2021): 120920.
10. Khalili, Sadegh, Srikanth Rangarajan, Vadim Gektin, Husam Alissa, and Bahgat Sammakia. "An Experimental Investigation on the Fluid Distribution in a Two-Phase Cooled Rack Under Steady and Transient Information Technology Loads." Journal of Electronic Packaging 142, no. 4 (2020).
11. Drummond, Kevin P., Doosan Back, Michael D. Sinanis, David B. Janes, Dimitrios Peroulis, Justin A. Weibel, and Suresh V. Garimella. "A hierarchical manifold microchannel heat sink array for high-heat-flux two-phase cooling of electronics." International Journal of Heat and Mass Transfer 117 (2018): 319-330.
12. Liang, Gangtao, and Issam Mudawar. "Review of pool boiling enhancement by surface modification." International Journal of Heat and Mass Transfer 128 (2019): 892-933.
13. Patil, Chinmay M., and Satish G. Kandlikar. "Pool boiling enhancement through microporous coatings selectively electrodeposited on fin tops of open microchannels." International Journal of Heat and Mass Transfer 79 (2014): 816-828.
14. Wong, Kin Keong, and Kai Choong Leong. "Saturated pool boiling enhancement using porous lattice structures produced by Selective Laser Melting." International Journal of Heat and Mass Transfer 121 (2018): 46-63.

Section 4: Mechanical Requirements

Mechanical stresses stem from differential thermal expansion of different materials within a package built-in during assembly and fabrication. Warpage is a manifestation of physical deformation and built-in stresses. Warpage engineering is thus an essential part of mechanical stress management and is essential to packaging to ensure that packages can be assembled into subsequent applications.

Warpage Engineering and Stress Management

Warpage engineering has long been an integral part of electronics packaging in both reliability and manufacturing for high performance package fabrication and its assemblies as well as for mobile and consumer products and their assemblies. As the form factors get thinner and smaller, yet higher in power consumption, understanding warpage behaviors during fabrication processes and service life became essential for successful product engineering from design to development qualification and volume manufacturing.

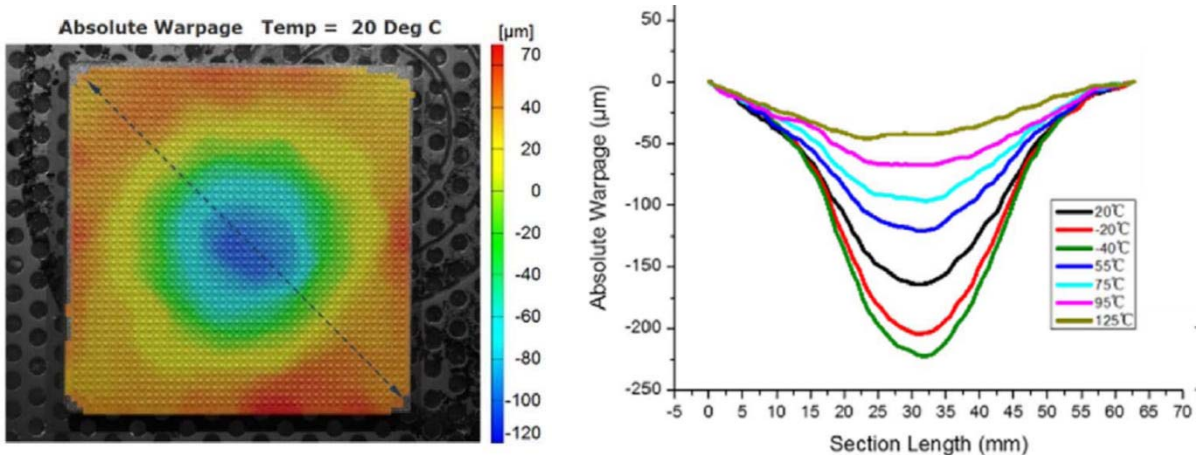


Figure 1. Thermal deformation (warpage) of a high-performance flip chip PBGA package

For System-in-Packages (SiP) through Heterogeneous Integration, the characterization of warpage is complicated, since the package includes multiple components and materials within a package. The package may include thinned dies, passives (capacitors/resistors/inductors), MEMS sensors, and stacked memories.

A good example is the FoWLP where it is important to understand the coplanarity of the reconstituted wafer after the molding process. What are the warpage characteristics and built-in stress of the reconstituted wafer as well as the package parts after singulation? Warpage engineering is a manufacturing issue as well as a product robustness issue. An important question is how to measure warpage. What are the modelling and simulation tools? What are the metrology tools?

Finite Element Analysis tools have been well developed to address mechanical stress requirements. Coupled with accurate metrology tools such as warpage metrology, there is great potential to bring greater accuracy and deeper insight in simulation of stresses and deformation in the package.

Warpage Metrology tools available today:

- Shadow Moiré
- Projection Moiré
- 3D Digital Image Correlation
- Confocal Displacement Metrology

Each tool has its advantages and disadvantages. Proper methods should be selected for different applications such as manufacturing line monitoring or in-depth laboratory study. Following is a wish list:

- Optical non-contact method
- No surface treatment 3D metrology tool
- Measurement accuracy: < 1 um
- Special resolution as < 20 um
- Capable of measuring in a mixed-surface condition (specular as well as diffusive surface)
- Measurement speed: as fast as under one second in 15x15mm FOV
- Measurement in heating and cooling environment in -55~250 C

Package stress and contribution to warpage

Following is an example of a flip chip package which lies flat entering the reflow oven. As the parts exit the reflow oven, all the parts start to shrink. The differential expansion and contraction of the die and substrate during temperature cycling results in warpage of the package, and built-in stress in the package.

Package stacking and die stacking add layers to the assembly to shorten interconnect and reduce size. Package-on-package and other stacked structures would require insight into their warpage behavior and stress management. Figure 3 shows the assembly process for a 2.5D device, where the package will undergo reflow 3 times with associated warpage phenomenon.

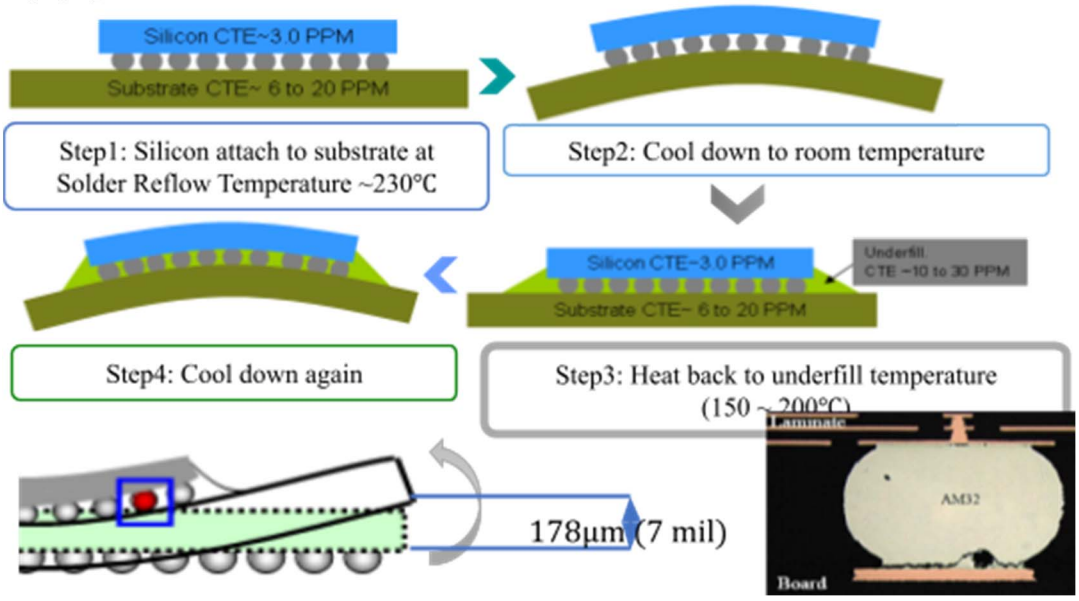


Figure 2. Reflow of a flipchip package

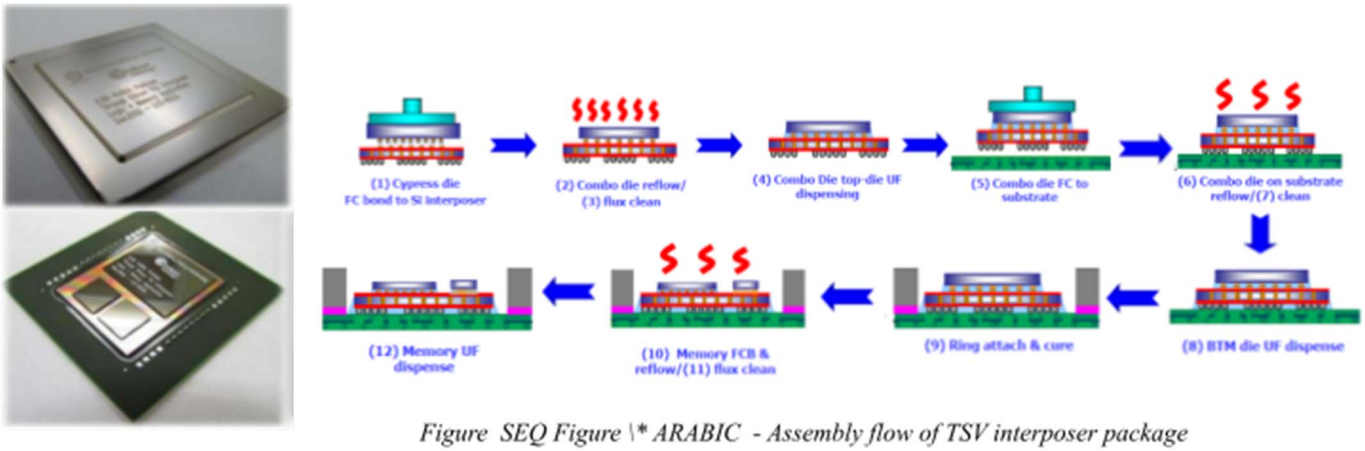


*Figure SEQ Figure * ARABIC - Assembly flow of TSV interposer package*

A method to help visualize the warpage during reflow is called Digital Image Correlation (DIC). This tool uses stereo cameras pointed into a chamber where the part to be studied is placed. The chamber goes through a temperature cycle from 25°C to 245°C. This data is captured and a deformation map is created to show the movement of the points that were captured, as shown in Figure 4.

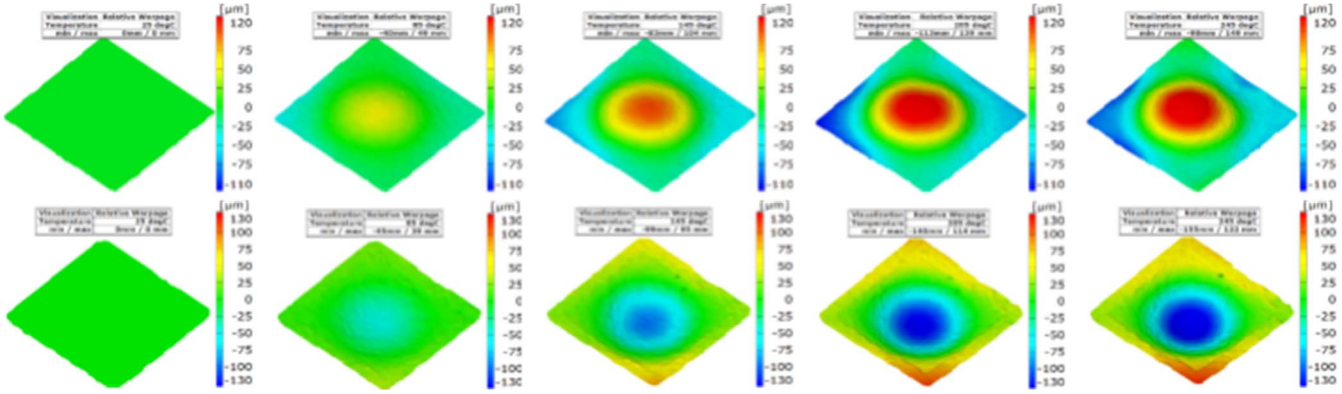


Figure 4. Warpage progression through thermal cycling

Stress Management through FEA and Warpage Metrology

Understanding the stresses and where they occur allows the engineer to make tradeoffs – underfill, stiffener, adhesive, substrate and copper loading. FEA tools and full field metrology imaging analysis through Artificial Intelligence have great potential for insight to assembly yield and product robustness.

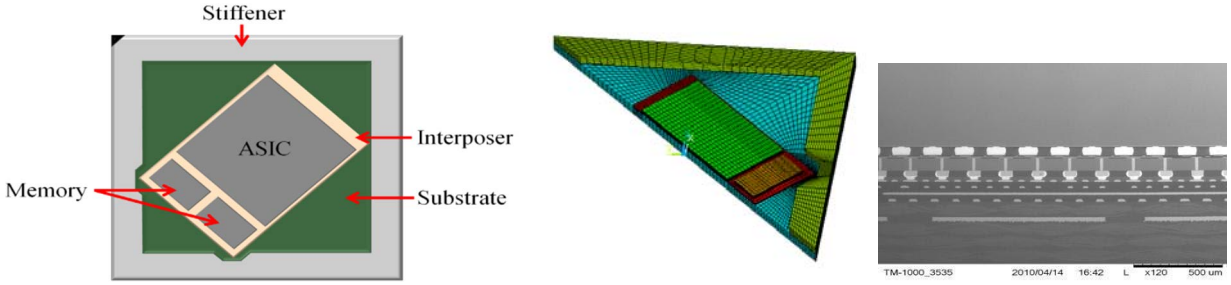


Figure 5. FEA model of 2.5D package and micro bump cross section

Stress Management and Chip Package Interaction

Chip package interaction (CPI) refers to failure modes such as delamination in low k dielectric or in solder bump failure due to mechanical stress from temperature excursions during assembly or product usage. Mechanical CPI is well known and rigorously managed by the packaging community. Shown in Figures 6 through 8 are examples of three different packages: the well-established Flip Chip CSP Package, and the more recent WLCSP and FOWLP packages. As the industry continues towards advanced IC nodes, and develops new package types, stress management and CPI will continue to occupy an essential part of the engineering community tool kit.

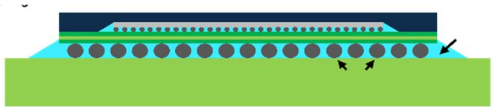


Figure 6. Structure of a FCCSP Package

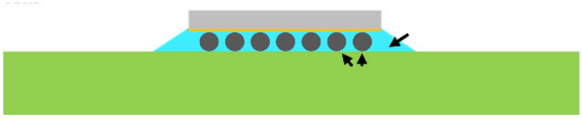


Figure 7. Structure of a WLP Package

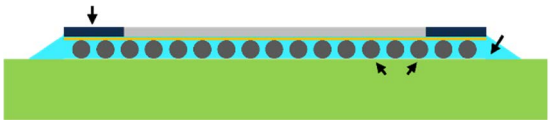


Figure 8. Structure of a FOWLP Package

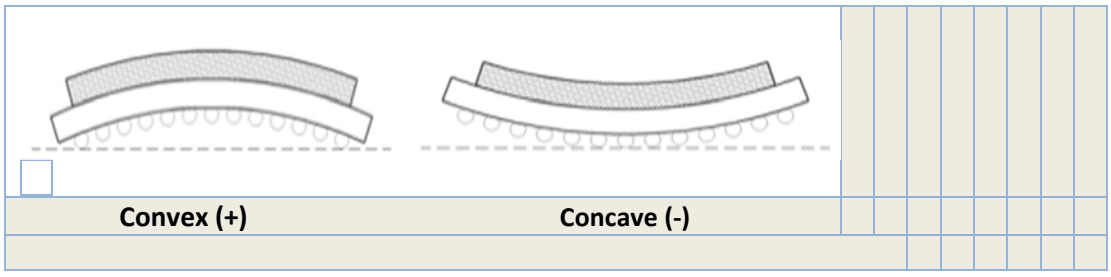
Table 1 shows the warpage allowance vs interconnect pitch. The table is split to cover both HPC and consumer (mobile) packages. HPC applies for devices 25mm and larger and consumer applies to packages smaller than 25mm. The current state-of-the-art mobile device is the advanced processor with a 0.35mm interconnect pitch and a body size of approximately 14mm x 13mm. HPC processors will be heading to body sizes larger than 60mm x 60mm with interconnect pitches remaining at 1.00mm to 0.8mm. The interconnect pitch is being driven by the complexity of escaping these large packages.

The warpage allowance is governed by interconnect pitch-reduction and package size. The smaller the pitch, the lower the warpage allowance. As the pitch gets smaller, the interconnect will also get smaller, so for a solder ball interconnect, the ball size will be smaller so the allowable warpage and co-planarity will be reduced to ensure proper

assembly yields. The package size has the same effect: as package sizes get larger, the warpage will increase so more consideration will be made to materials and structures to reduce the warpage during reflow to ensure manufacturing yields.

Table 1: Warpage Allowance across two market segments

	Year of Production	*2018*	*2019*	*2020*	*2023*	*2026*	*2029*	*2030*
	Pitch (mm)							
HPC	**1.0**	-0.13, +0.21	-0.11, +0.18	-0.11, +0.18	-0.11, +0.18	-0.10, +0.16	-0.08,+0.16	-0.08,+0.17
		-0.13, +0.20	-0.11, +0.18	-0.11, +0.18	-0.11, +0.18	-0.10, +0.15	-0.08,+0.15	-0.08,+0.16
	0.8	-0.13, +0.21	-0.11, +0.18	-0.11, +0.18	-0.11, +0.18	-0.10, +0.16	-0.08,+0.16	-0.08,+0.17
		-0.10, +0.10	-0.09, +0.09	-0.09, +0.09	-0.09, +0.09	-0.08, +0.08	-0.07, +0.07	-0.07, +0.07
	0.65	-0.10, +0.10	-0.09, +0.09	-0.09, +0.09	-0.09, +0.09	-0.08, +0.08	-0.07, +0.07	-0.07, +0.07
		-0.09, +0.09	-0.08, +0.08	-0.08, +0.08	-0.08, +0.08	-0.07, +0.07	-0.065, +0.065	-0.065, +0.065
	0.5	-0.09, +0.09	-0.08, +0.08	-0.08, +0.08	-0.08, +0.08	-0.07, +0.07	-0.065, +0.065	-0.065, +0.065
		-0.08, +0.08	-0.07, +0.07	-0.07, +0.07	-0.07, +0.07	-0.065, +0.065	-0.06, +0.06	-0.06, +0.06
	0.4	-0.08, +0.08	-0.07, +0.07	-0.07, +0.07	-0.07, +0.07	-0.065, +0.065	-0.06, +0.06	-0.06, +0.06
		-0.07, +0.07	-0.065, +0.065	-0.065, +0.065	-0.065, +0.065	-0.06, +0.06	-0.055, +0.055	-0.055, +0.055
Mobile	**0.3**	-0.07, +0.07	-0.065, +0.065	-0.065, +0.065	-0.065, +0.065	-0.06, +0.06	-0.055, +0.055	-0.055, +0.055
		-0.06, +0.06	-0.055, +0.055	-0.055, +0.055	-0.055, +0.055	-0.05, +0.05	-0.045,+0.045	-0.045,+0.045
	0.25		-0.055, +0.055	-0.055, +0.055	-0.055, +0.055	-0.05, +0.05	-0.045,+0.045	-0.045,+0.045
				-0.055, +0.055	-0.055, +0.055	-0.05, +0.05	-0.045,+0.045	-0.045,+0.045
	0.2			-0.055, +0.055	-0.055, +0.055	-0.05, +0.05	-0.045,+0.045	-0.045,+0.045
					-0.045,+0.045	-0.045,+0.045	-0.045,+0.045	-0.045,+0.045
	0.15				-0.045,+0.045	-0.045,+0.045	-0.045,+0.045	-0.045,+0.045
					-0.025,+0.025	-0.025,+0.025	-0.025,+0.025	-0.025,+0.025
	0.1				-0.025,+0.025	-0.025,+0.025	-0.025,+0.025	-0.025,+0.025
						-0.020,+0.020	-0.020,+0.020	-0.020,+0.020



Future Challenges and Opportunities

For the 2020 update, there has been a trend for HPC packages to use larger and larger substrates with multiple dies. The current packages being developed for use in the upcoming year range from 75mm to 100mm square. These packages are close to basically a large SIP with multiple ASIC/CPU, memory and IO devices. Putting these functions together onto a substrate reduces latencies and allows for more efficiencies in the system. The challenge comes when this large-substrate package gets integrated into the next-level assembly. As the package gets larger, warpage control gets more important. If the package warpage exceeds a total of 0.006" (per JEDEC JESD22 – B112B) the instances of opens and head-in-pillow (HiP) will increase. Figure 9 shows how warpage causes such a failure.

The larger the package the less warpage is allowed per unit area in order for a successful joining to the next-level assembly. More focus will need to be placed during co-design to meet not only electrical performance but also assembly yields and reliability targets. New materials will be needed with a better CTE match of the various components in the package. Mechanical modeling and verification can provide a structure that will have the least amount of warpage. These are tough challenges that the industry has to face as we move to higher performance, more demanding packages.

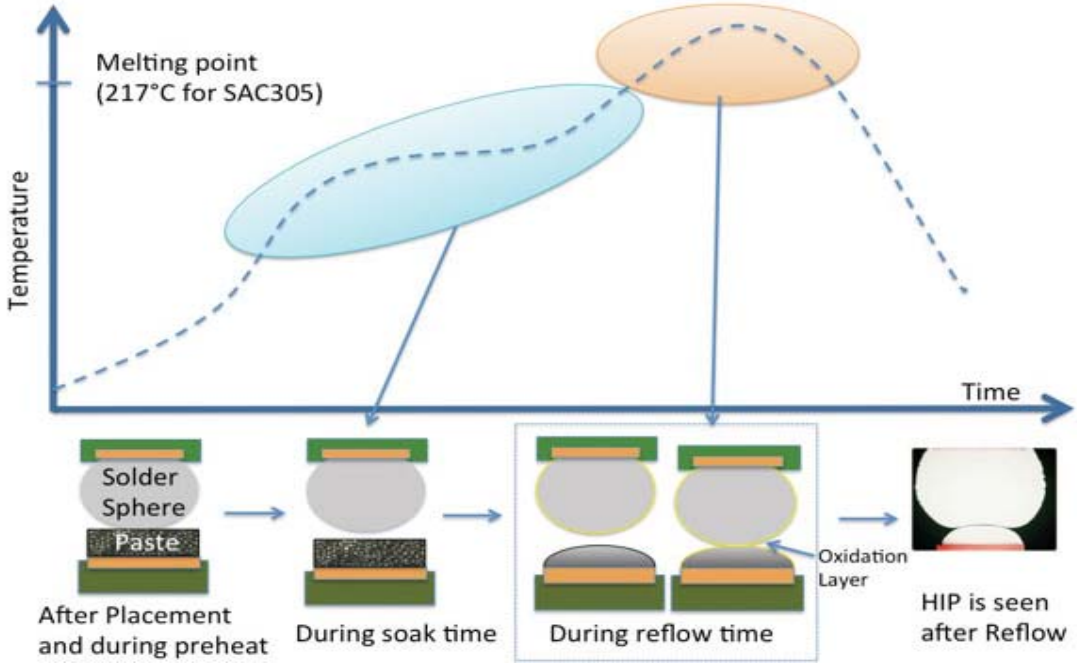


Figure 9 - Head in Pillow failure mechanism (source: Circuit Assembly 19 May 2017)

The trend towards smaller and lighter packages has created a "perfect storm" of possible **chip-packaging interaction** (CPI) catastrophes, and stimulated the industry to stay ahead through innovation, forward-looking research and smart, rational thinking. With artificial intelligence and development of advanced DIC such as full-field imaging capabilities, we see potential for high accuracy feedback on the manufacturing line for high yield and a high-quality product.

References

1. G. Wang, P. S. Ho, and S. Groothuis, "Chip-packaging interaction: a critical concern for Cu/low k packaging," Microelectronics Reliability 45, pp. 1079–1093, Jan. 2005.
2. S. E. Thompson, G. Sun, Y. S. Choi, and T. Nishida, "Uniaxial-process-induced strained-Si: extending the CMOS roadmap," IEEE Trans. on Electron Devices, vol. 53, no. 5, pp. 1010–1020, May 2006.
3. M. Nakamoto, R. Radojcic, W. Zhao, V. K. Dasarapu, A. P. Karmarkar and X. Xu "Simulation methodology and flow integration for 3D IC stress management", Custom Integrated Circuits Conference (CICC), San Jose, CA, Sept. 19-22, 2010.
4. K. Dhandapani, A. Syed, W. Lin, M. Nakamoto, W. Zhao and R. Radojcic, "A methodology for Chip - Package Interaction (CPI) modeling in 3D IC structures," IMAPS 8th International Conference and Exhibition on Device Packaging (DPC), Fountain Hills, AZ, Mar. 6-8, 2012.
5. W. Zhao, M. Nakamoto, V. Ramachandran and R. Radojcic "Mechanical stress management for electrical chip-package interaction (e-CPI)", Electronic Components and Technology Conference (ECTC), 2014.
6. Sentaurus Band Structure, version H-2013.03.

Section 5: Thinning and Singulation

1. Introduction to Thinning

All silicon wafers start out at the foundry at between 0.7 to 0.8mm in thickness. To fit inside today's low profile single and multichip integrated packages, wafers are thinned by processes where the wafer backsides are removed leaving the active frontside at a fraction of the original thickness. Thinning semiconductor wafers is a widespread process using abrasive rotary grinding and polishing wheels, reaching 50um thicknesses with very good total thickness wafer variation across the entire 300mm wafer. Mechanical abrasive grinding of silicon wafers reaches a limit at around 50um and a gentler chemical removal is required to thin below 50um in the ultrathin realm. Chemical mechanical polishing, wet etching and dry etching are used as very smooth, stress-free and defect free surfaces become imperative. Traditional thinning using grind wheels has been covered in previous editions of this chapter. This edition will talk more about foundry-type thinning which is more chemical and less mechanical to achieve these extremely thin dimensions. The same thinning techniques are used for compound semiconductor wafers including silicon carbide, sapphire, InP, GaAs and others. According to Yole, the wafer thinning market activity is $100M in 2019 and is expected to increase to $135M in 2025[1].

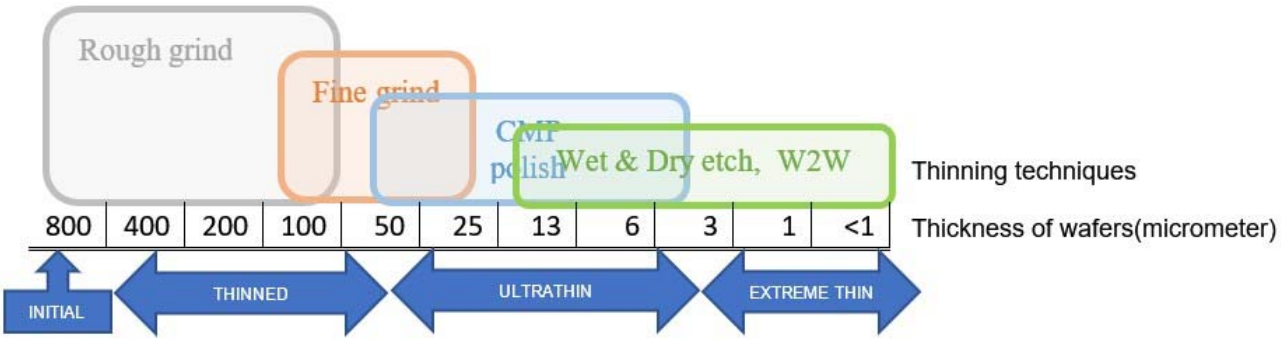


Figure 1. Thinning techniques vs. wafer thickness

Wafer thinning allows for more devices and functions to be fitted into a tight space to create smart lightweight and portable devices that are commonplace now. Additionally, thinning and miniaturization can improve and speed up device performance as well as improve thermal and power management. For wearables, the ultrathin stress-free dies at 30um and below allow for the devices to behave like fabric. For memory dies, multiples of 50 um die are interlayered and stacked allowing astonishing levels of vertical integration. For image sensors used in cameras, making dies extremely thin by flipping wafer backside to become frontside creates smaller pixel size and better camera performance. Sony and other suppliers have been making ultrathin image sensors in the 3-5um range for over a decade based on concepts patented by Kodak and Tower Semiconductor[2]. Figure 2 lists the technologies and thicknesses of the respective devices courtesy of Yole[1].

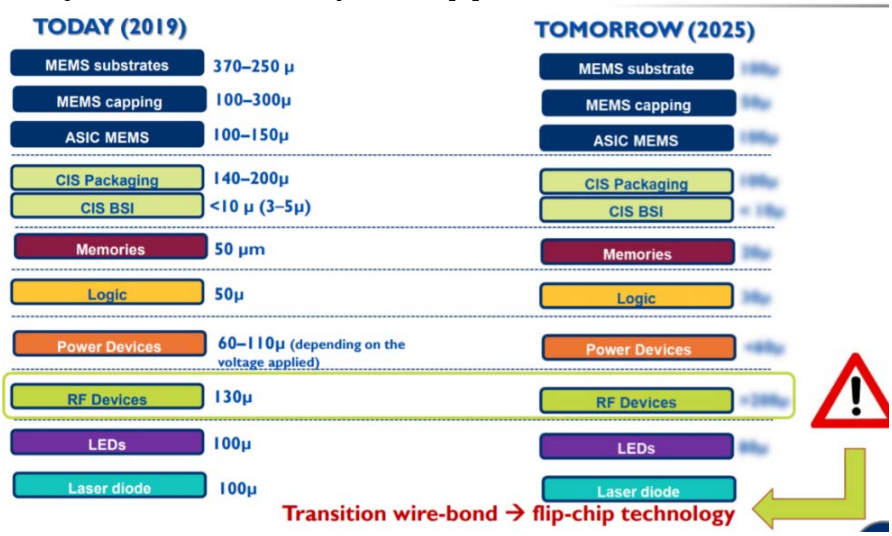


Figure 2. Various final thicknesses of die by application [1]

2. Extremely thin dielets

To transition from thinned to extreme thinning, wafer level processing (WLP) techniques are used such as carrier wafer bonding and debonding. A. Jourdain and an imec team published the fabrication of extremely thin dies at the 0.5um level at ECTC 2020[3]. To do this, they started with a specially prepared wafer with etch stop layer separating the top submicron epitaxial device layer and bottom substrate. After front-end processing is done, the bottom substrate is ground and etched off, exposing the etch stop layer which is SiGe. This is subsequently removed with selective etch leaving a submicron wafer which at this point is supported by a carrier wafer. The flipping of backside to become front side with the use of temporary support wafers is a proven technique for making extremely thin wafers/dies. This is the same concept used by backside illuminated (BSI) camera manufacturers to achieve 3-5um thin dies where the backside is etched away after bonding to the lens wafer[2]. So the challenge to keep these wafers flat has been solved by permanent or temporary carrier wafer bonding providing support to the ultrathin wafer throughout the RDL and BOEL processes of chip fabrication. The result is a hybrid stackup which is mechanically robust. Another challenge is the control of total thickness variation (TTV) to near zero across the entire wafer, and wafer-edge geometry control[4].

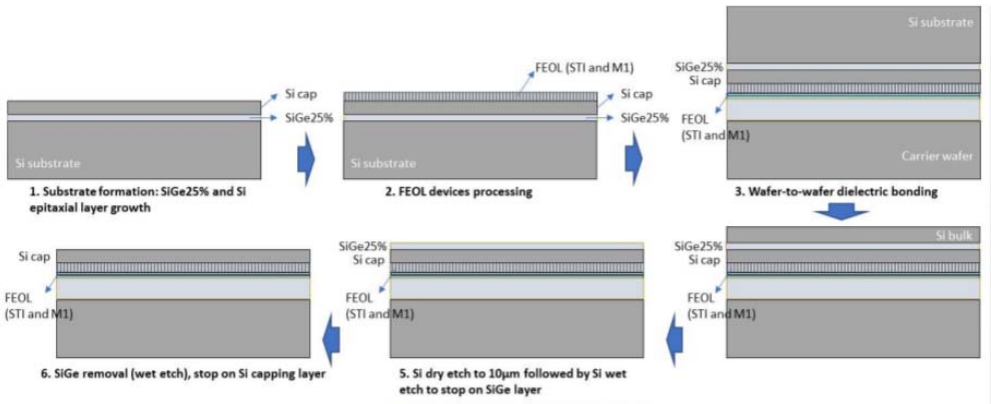


Figure 3. Process flow for making extreme thin dies, courtesy of A. Jourdain et al, imec.[3]

The approach by C. Bower and team at X-Display does not require back-side flipping to front side, but instead extremely thin dielets are fabricated by undercutting the die by chemical etching [5]. Individual die mesas on wafers are formed first and, as shown in Figure 4, the method of fabricating an extremely thin dielet in the 0.5 um thickness range is developed. The wet etching is both the thinning and singulating process combined to one process step.

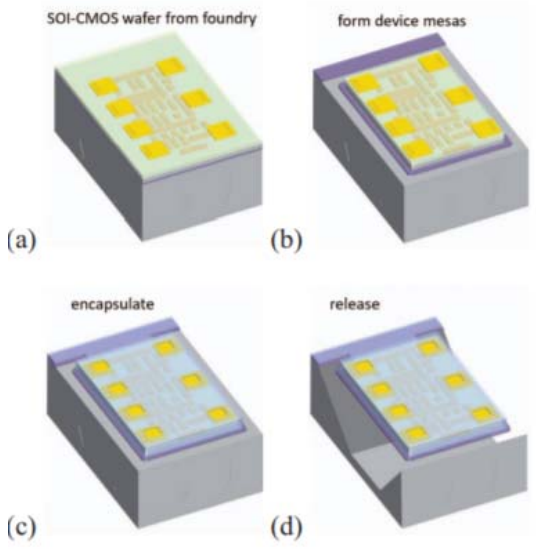


Figure 7: process flow for making print-ready microICs.

Figure 4: Process flow for making print-ready microICs, courtesy of C.Bower of X-Display[5]

3. Singulation of dielets

Figure 4. Singulation methods for silicon wafers

Similar to backgrinding, the conventional mechanical technique with rotary abrasion blades is the workhorse for singulation of dies from the wafer. Even for ultrathin wafers, ultrathin blades of 10um and less are available for singulation. However, the non-saw method of laser dicing has gained ground, driven by the growth of compound semiconductor dies, MEMs and other applications where dies are delicate. Various laser methods (stealth, ablation, thermal laser separation) do not expose dies to as much vibration, so mechanical stresses and structural defects are minimized. For non-silicon wafers, such as InP and GaAs where the material is very brittle, laser dicing including stealth laser is the go-to method for singulation. These described methods are based on a wafer sitting on a support tape or a carrier wafer consisting of streets where the blade or laser beam will cause a separation. Laser dicing consists of multiple steps including a final mechanical stretching or cleaving step to separate individual dies. This separation step may introduce stress with the exception of through-cut wafers. Both laser and saw techniques result in stress and are non-batch compared to the chemical etching technique which removes the streets in a batch step. This has driven wet etch, plasma and dry etch to be the chosen method for dielets and chiplets. Dry etching is especially advantageous in reducing singulation time for wafers with tens of thousands or even millions of dielets.

A new approach is revealed by C. Bower, et al [5], to undercut material under the die surface to fabricate dielets without grinding. A die mesa is formed and, after undercutting, each dielet is held by a tether anchored to the mother wafer and ready to be snapped off by the picking action.

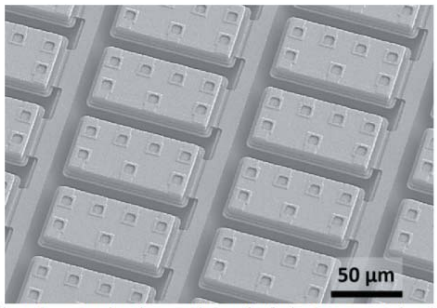


Figure 8: a scanning electron micrograph of microIC source wafer with undercut, print-ready ~90x50 µm² microICs.

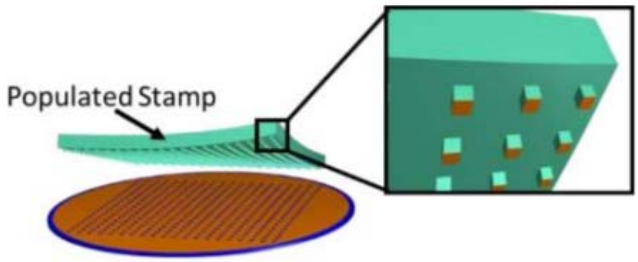


Figure 6. A scanning electron micrograph of 3.5 µm thick 50x90um silicon CMOS dielets [5] ready for picking with an elastomeric array tool [6]

4. Dielet Pick

Die pick is a challenging step for fragile ultrathin dies that are susceptible to mechanical damage. Picking with a traditional vacuum pick tool on the top of die with ejector pins underneath the taped die is no longer viable for ultrathin dies. Consequently, the "push and pick" action has been replaced by "peel and pick" action with the use of timed and graduated pushing and vacuuming. This is particularly necessary for larger dies. The need for large quantities of MicroLEDs to be efficiently placed on large displays have established the micro transfer printing (µTP) process. The pick tool is made of the elastomeric material PDMS with properties of softness and temporary stiction. Preparation at the wafer level through WLP must be developed and implemented to allow wafers to be µTP. This includes patterning tethers and anchors to hold dies in place after undercutting. The tether of each dielet breaks off during pick and sticks to the tool. After transfer, it is released to stick to the receiving substrate, followed by interconnection to pads. Contact sites can multiply on the tool to pick multiples of dielets, at once tremendously

increasing efficiency to >10000 per minute.[5] µTP scaling to larger dies for the future is a possibility. It is also possible that dielets can be harvested from a wafer by a liquid medium. E. Chow and team at PARC are developing a microassembler by making dielets suspended in liquid to self-sort, transport and self-orient to form programmed patterns. The arrangements are then transferred to a final substrate with contact stamping or an electrostatic roller belt.[7]

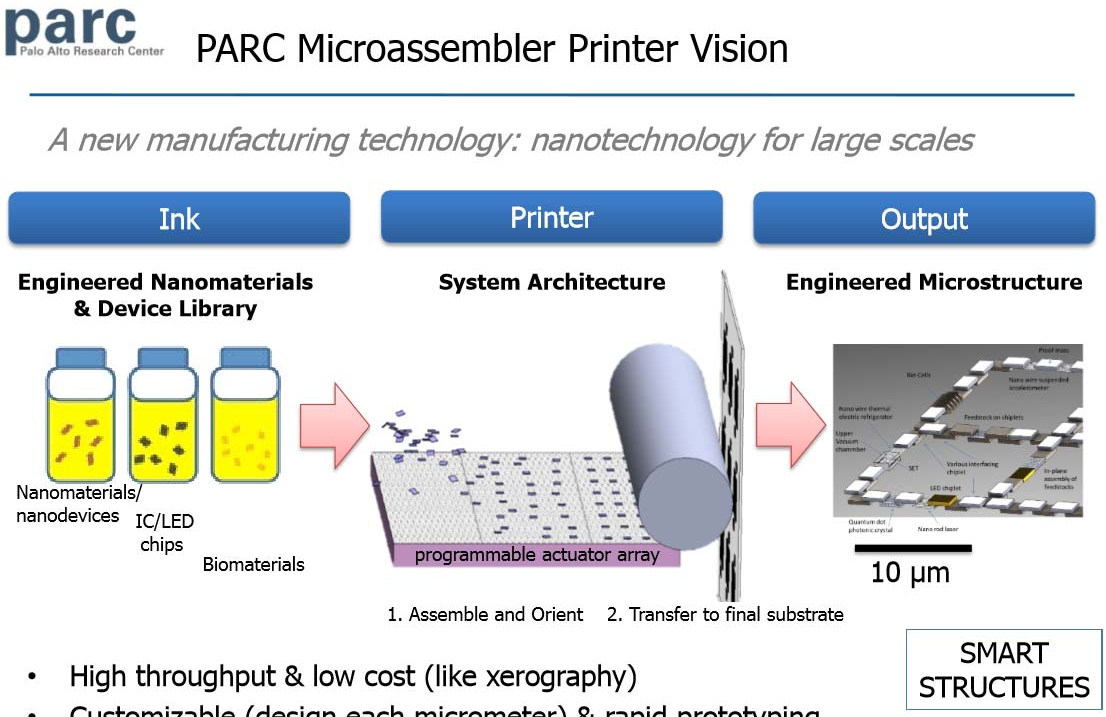


Figure 7. PARC Microassembler Printing concept for dielets; courtesy of E. Chow [7]

The push to thinner and smaller dies or dielets is revealing that wafer level processing can use some highly controlled processing of etching and W2W bonding. New concepts for making and picking dielets are cropping up and the landscape is constantly changing, while conventional thinning, dicing and picking will still be the mainstay for established devices.

References

1. YOLE Report "Thinning Equipment Technology and Market Trends for Semiconductor Device, 2020 Market and Technology Report 2020"
2. US 6,168,965 B1 Date of Patent: Jan. 2, 2001 METHOD FOR MAKING BACKSIDE ILLUMINATED IMAGE SENSOR Inventors: Yacov Malinovich, Tivon; Ephie cited by examiner Koltin, Givat Elah, both of (IL)
3. A. Jourdain et al., "Extreme Wafer Thinning and nano-TSV processing for 3D Heterogeneous Integration", 2020 IEEE 70th ECTC
4. A. Jourdain et al., "Extreme thinning of Si wafers for via-last and multiwafer stacking applications", 2018 IEEE 68th Electronic Components and Technology Conference (ECTC), pp1523-1530 May 2018
5. C. A. Bower et al., "High-brightness displays made with micro-transfer printed flip-chip microLEDs", 2020 IEEE 70th Electronic Components and Technology Conference (ECTC), June 2020.
6. D. Gomez, K. Ghosal, M.A. Meitl, S. Bonafede, C. Prevatte, T. Moore, B. Raymond, D. Kneeburg, A. Fecioru, A.J. Trindade and C.A. Bower, "Process Capability and Elastomer Stamp Lifetime in Micro Transfer Printing," 2016 IEEE 66th Electronic Components and Technology Conference (ECTC), pp. 680-687 (2016)
7. E. Chow, "Chiplet Micro-Assembly Printer" 2019 IEEE 69th Electronic Components and Technology Conference (ECTC). pp1312-1315 May 2019

Section 6: Wire Bonding for Multi-Chip and System-in-Package Devices

Wire bonding continues to be the most dominant interconnection technology used in the electronic packaging industry due to its low cost, high yield rate, design flexibility and proven reliability.

Wire Bonding Market: Trend and Challenges

Wire bonding makes more than 10 trillion bonds each year. Innovations continue to extend the life of wire bonding by further reducing packaging cost and providing more capability. Figure 1 shows semiconductor package growth using different interconnect methods. It shows that currently (2018) wire bonded packages made up about 77% including wire-bonded single die and SiP. The major growth in wire bonding is from SiP packages. From 2018 to 2023, wire-bonded SiP will grow from 25 to 38 billion units, while wire-bonded single chip will only grow from 176.0 to 181.8 billion (Prismark). Wire bonding continues to drive toward lower cost, improved productivity, increased interconnect density and improved process monitor, real time control, and defect and factory management. We will look at these four key trends next.

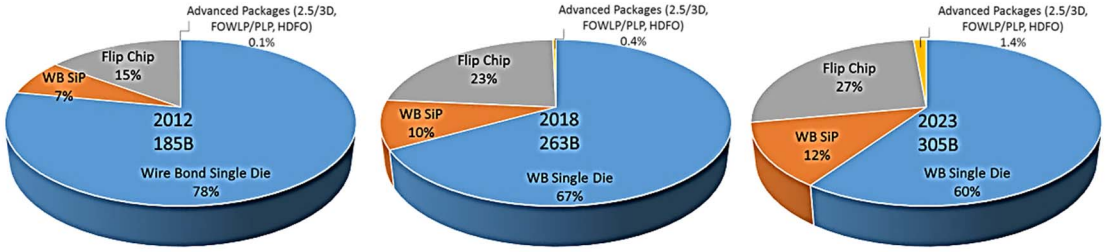


Figure 1: Semiconductor package unit growth by interconnect type. IC Package Shipment excluding Discrete, LED and Opto. Source: Prismark May 2019.

Lower cost

One of the biggest trends in wire bonding is replacing Au wire with lower-cost Cu and Ag wire. Figure 2 shows a wire cost comparison. By switching to a lower-cost wire, package cost can be reduced by 20% [1]. The combination of Cu and Pd-coated Cu wire (PdCu) has overtaken Au wire as the most popular wire used for semiconductor packaging. Cu processes are much more complicated. By leveraging R&D development, process models were developed and implemented with simplified 'response-based' inputs such as desired bonded ball diameter. These new advanced processes have demonstrated improvements in yield performance and throughput and cost savings [1, 2, 3, 4].

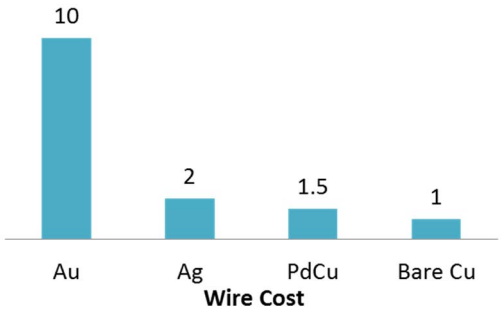


Figure 2: Relative cost for different type of wires, based on 25um diameter wire (Source: Heraeus and Tanaka). 94% Ag alloy is used in the comparison.

The memory device is one area in packaging that has not adopted Cu wire bonding due to a few challenges including thin bond pads, overhang configuration and multiple stack dies. An attractive alternative for these devices is Ag wire [5, 6].

Another material cost reduction is the development of lower cost substrates and leadframes such as PPF (Pre-Plated Frame) QFN. PPF QFN offers reduced cost and simplified assembly by eliminating deflash and Sn-alloy plating steps. Wire bonding to these cost-reduced materials is often more challenging and requires more advanced processes [1, 3].

Higher productivity

High wire bonder throughput is the key to supporting the volume requirement of our industry. Through the years, wire bonders went through many technology advances to improve the speed of wire bonding. In the last 20 years, wire bonder throughput has more than doubled [7]. This trend will continue. More and more of the recent speed

improvement is coming from improving the bonding and looping process instead of increasing the motor speed of the bonder's XYZ system [1, 3, 5].

Another productivity factor to consider is time to market – the key to survival in our industry. New products need to be developed and produced in a short period of time to capture market share. This desire drives a faster design and production cycle. 3D loop design and clearance-check software, along with the wire loop model on the wire bonder side, reduce loop optimization time significantly. Figure 3 shows multi-chip devices with highly complex wire bond looping configurations and Figure 4 shows the 3D loop design tool that can help shorten the overall time to market [8].

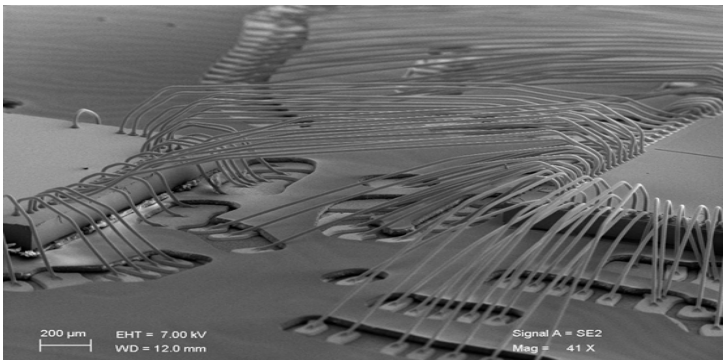


Figure 3. Advanced multiple tier looping and die-to-die bonding loops (K&S)

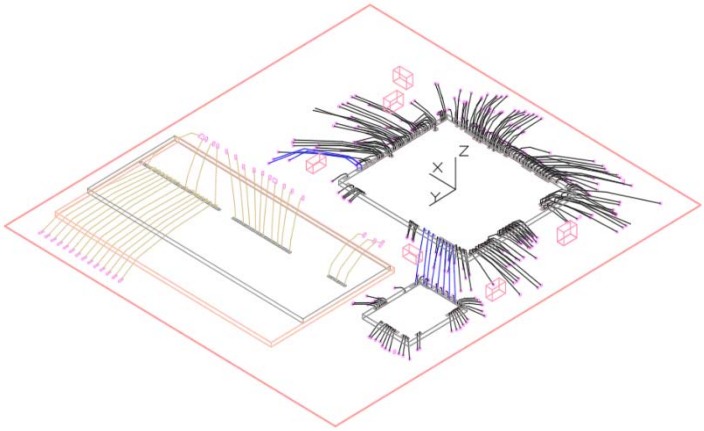


Figure 4. 3D loop design and clearance check software (K&S)

Increase Interconnect Density

One of the key wire bonding roadmap drivers has been the fine pitch capability requirement. The current finest pitch capability is 35um in-line bond pad pitch [7]. Other approaches to increase packaging density include multi-tier looping, multi-chip modules and vertically stacked packages. We will look at each of these packaging approaches next.

Multi-tier pad design is a common solution for increasing I/O counts. For example, a 50um bond pad pitch device with 4 pad rows is a common configuration for packages over 1000 I/Os (Figure 5). Wire bonding technology is an intrinsic "fan out" technology making the package design more forgiving and allows more flexibility of the pad and substrate layout [8].

Multi-chip module and System in Package (SIP) are used to increase package density and functionality. Die-to-die wire bonding is often required in these types of packages. For die-to-die bonding, a type of wire bonding called Stand-Off-Stitch Bond (SSB) is widely used. The SSB process starts with a flat-topped bump bonding on one die, followed by the formation of a new ball bond (1st bond) on the substrate or on a 2nd die. Finally, the stitch bond (2nd bond) of that wire is bonded on top of the initial bump (Figure 6). Due to multiple bonds placed on the bond pad, the pitch capability for the SSB process is a few microns larger than the regular forward bonding process [9]. The industry is driving toward finer SSB pitch capability.

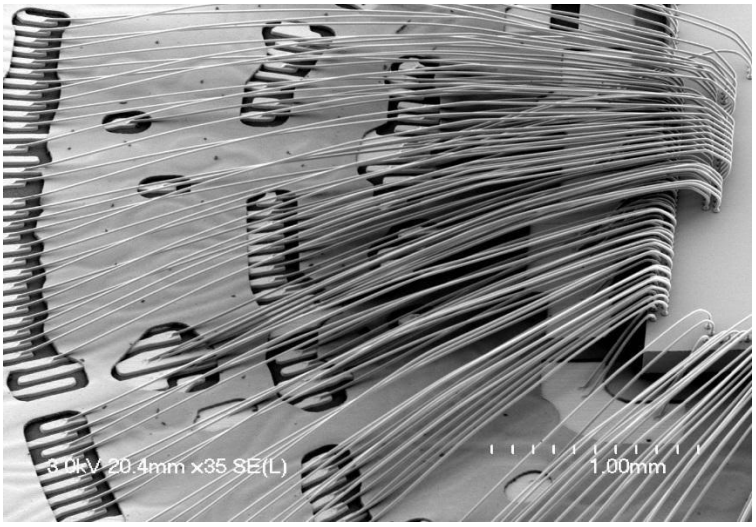


Figure 5. High density multi-tier package (K&S)

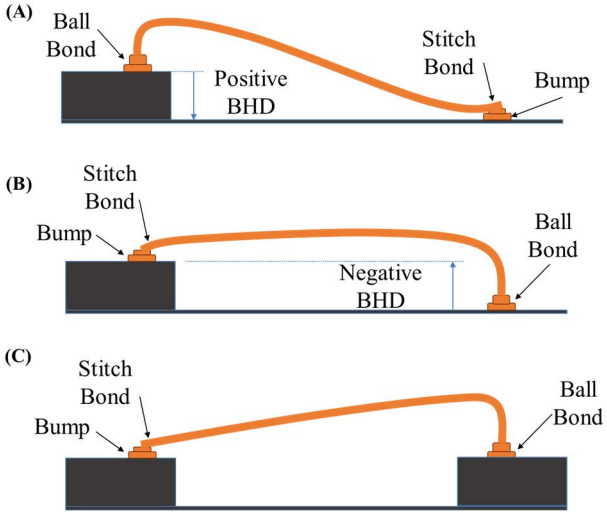


Figure 6. Stand-Off-Stitch Bond (SSB) process is more and more common on devices such as multi-chip modules, stack die and LED devices (K&S)

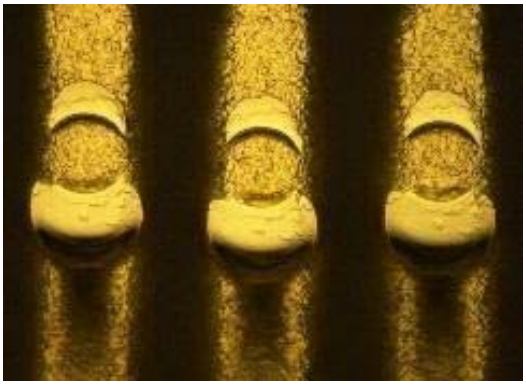


Figure 7. Bonding on 20um width lead fingers (K&S)

As bond pad pitch shrinks, lead finger pitch and width is also reducing, especially for stack die applications. Another driver for finer pitch lead finger is the shrinking of the package size. The lead fingers are closer to the die edge and have less space to spread out. Currently, 20um lead fingers are used successfully in mass production.

Vertically stacking semiconductor devices can effectively integrate more functionality in the same footprint. Stacked dies, such as NAND and DRAM, can minimize the packaging footprint by more than 200% [10,11]. Stacks of 4, 8 and 16 dies are common. Figure 8 shows an 8-stack die device. Higher numbers of stacked die such as 32 and 64 stacks are in the R&D phase and low volume production.

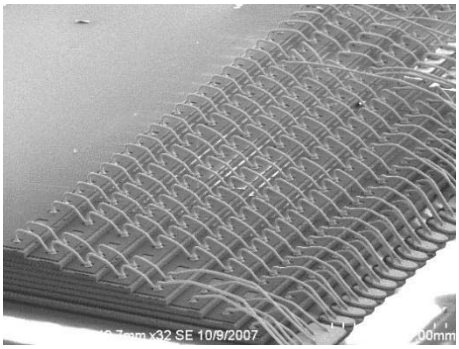


Figure 8. Example of a stacked-memory device (K&S)

In stack die packages, an 'overhang' configuration is very common. In overhang configurations, one or more dies may be unsupported (Figure 9). An optimization software feature provides accurate measurement of die deflection and optimization of the bonding parameters [12]. The wire bonding roadmap calls for more overhang bonding capability, including the ability of bonding on thinner dies with longer overhang distances.

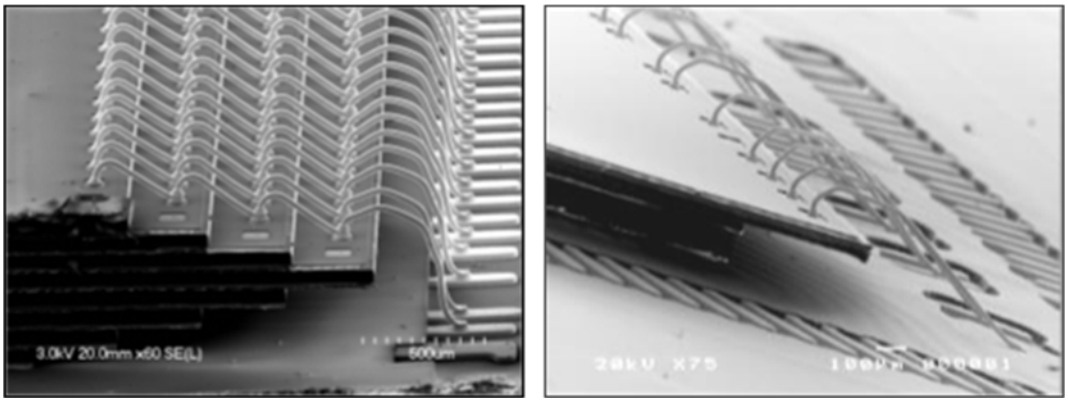
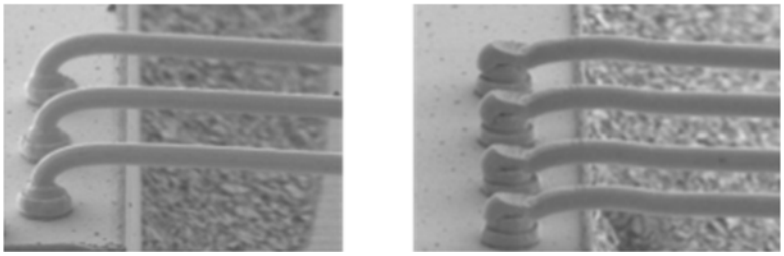


Figure 9. Examples of bonding on overhang die (K&S)

In order to achieve overall package height requirements for multi-stack memory packages, low loop heights of 100 um or les are often required. A long loop span with a bend near the 2nd bond is sometimes required to clear the lower tier dies in the stacked-die package. Due to the low loop height and die edge clearance requirements, loop formation needs to be carefully optimized. Examples of two loop types are given in Figure 10. A normal loop is faster, easier to optimize and has higher pull strength; however, the loop height is normally limited to 3x wire diameter or higher. In order to achieve a lower loop, a compressed loop is needed for loop height of 2x wire diameter or lower [6]. The wire bonding roadmap calls for improved low loop capability.



Normal Loop Compressed loop

Figure 10. Examples of low loop bonding with Ag wire using a normal loop and a compressed loop (K&S)

Future Challenges and Roadmap towards the Smart Wire Bonder

A key driver for the wire bonder roadmap is to support smart factory and industry 4.0 initiatives. In the semiconductor industry, automotive applications are leading the way to achieve "Zero Defects". Reliability becomes the top priority for automotive applications, with emphasis on process control, defect detection and traceability [13]. A new generation of smart bonding and looping processes have been developed in response to the new challenges facing the industry including the transition to Cu wire bonding, and the increasing need for real-time monitoring and closed-loop control. New machine functionalities have been added to meet the desire for factory automation, real-

time monitoring, closed-loop optimization and traceability [14]. The new smart equipment and functionalities reduce time-to-market and improve yield and throughput. The autonomous wire bonder is a real driver for advanced wire bonding technologies in the next five years.

Table 1: Wire Bond Interconnects

Year	2018	2019	2020	2021	2022	2025	2028	2031	2034
Wire Type									
Au	38%	35%	31%	28%	24%	23%	22%	21%	20%
Ag	3%	4%	5%	6%	7%	8%	8%	9%	10%
PdCu	30%	32%	34%	36%	38%	38%	38%	38%	38%
Cu	29%	29%	30%	30%	31%	31%	32%	32%	32%
Single In Line Pitch (µm)									
Au (Forward)	40	35	35	30	30	30	30	30	30
Au (SSB*)	45	40	38	33	33	33	33	33	33
Ag (Forward)	40	35	35	30	30	30	30	30	30
Ag (SSB)	45	40	38	33	33	33	33	33	33
PdCu (Forward)	40	35	35	30	30	30	30	30	30
PdCu (SSB)	45	40	38	33	33	33	33	33	33
Number of Pad/Loop Tiers (in HVM)									
Au	3	4	4	5	5	5	5	5	5
Cu	6	7	8	9	10	10	10	10	10
Cu (SSB)	4	5	6	7	8	9	10	10	10
Overhang Capability									
30um Die	0.5	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.0
20um Die	0.15	0.2	0.25	0.3	0.35	0.35	0.35	0.35	0.35
Low Loop Capability for Forward Bonding Wires (µm)									
Au and Ag wire	45	40	35	30	28	28	26	26	25

* -- SSB stands for Stand-Off Stitch Bond. It is widely used in SiP and Stack Die packages where die-to-die bonding is required.

Heavy Wire Bonding for Power Electronic Devices

While the traditional Al-based wedge bonding innovation continues to extend the life of wire bonding by further reducing packaging cost and providing more capability, technology development drives the need for new materials like Cu wire and ribbon. Wedge bonding stays competitive by reducing cost of ownership, increasing the level of automation, and real-time process monitoring and control. Additionally, wedge bonding is utilized in automotive applications that are growing due to the increasing electrification.

Higher Temperature Capability

One of the biggest trends in wedge bonding technology is the development of Cu wire and Cu ribbon bonding driven by changing semiconductor materials towards SiC and GaN (Figure 11), both of which operate at temperatures up to 400 °C. At such temperatures Al as interconnect material is no longer feasible and Cu must be used. Cu has its own unique set of challenges for the wedge bonding interconnect process. Its higher density means that spools with the same wire and ribbon length have a higher mass that must be accommodated. Cu wire and ribbon is stiffer and more abrasive on the consumables it comes in contact with. To maintain or extend consumable life, it is necessary to research and develop new materials and geometries.

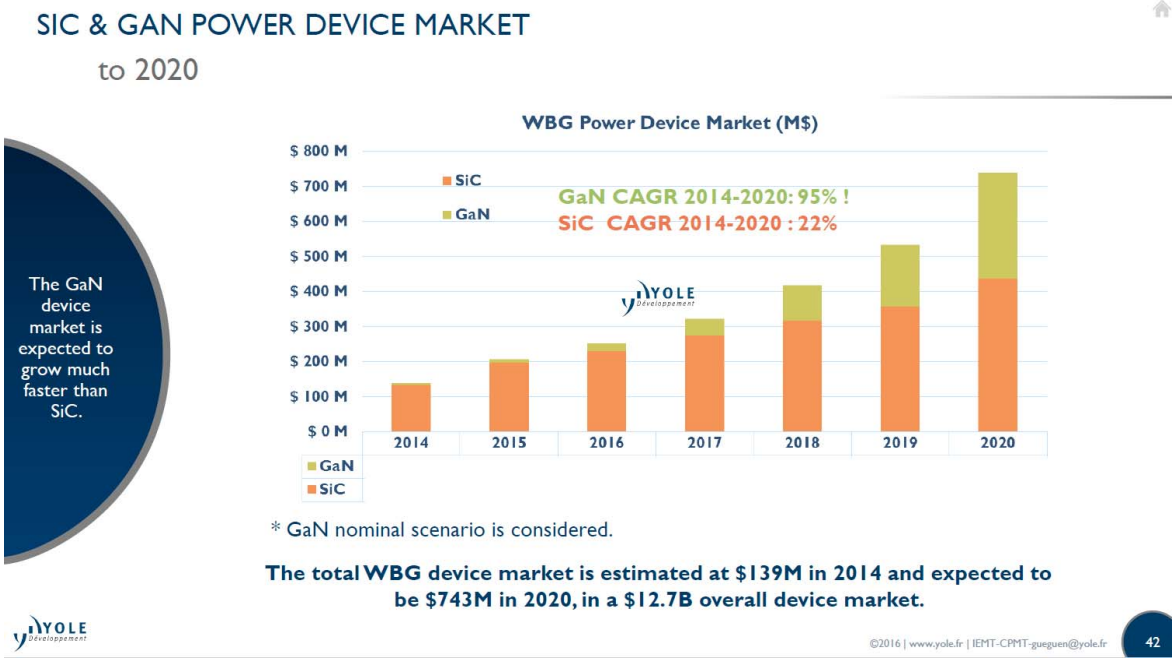


Figure 11: SiC and GaN Power Device Market. Source Yole Research 2016

Material Thickness Increase

To account for higher-power applications in automotive and power electronics, material thicknesses in both Al and Cu material need to be increased. These thicker materials require higher bond forces and ultrasonic power for a successful connection than is available with traditional wedge bonders. Therefore, higher-power ultrasonic systems with increased bond force capability are being developed to handle such applications. Currently these are niche applications, but the market for these applications is expected to grow at a moderate pace.

Automotive Electrification

A strongly growing driver for wedge bonding is increasing automotive electrification. This applies not only to electric vehicles but also conventional cars that become more electrified with sensors to accommodate smart controls and developments toward autonomous driving (Figure 12).

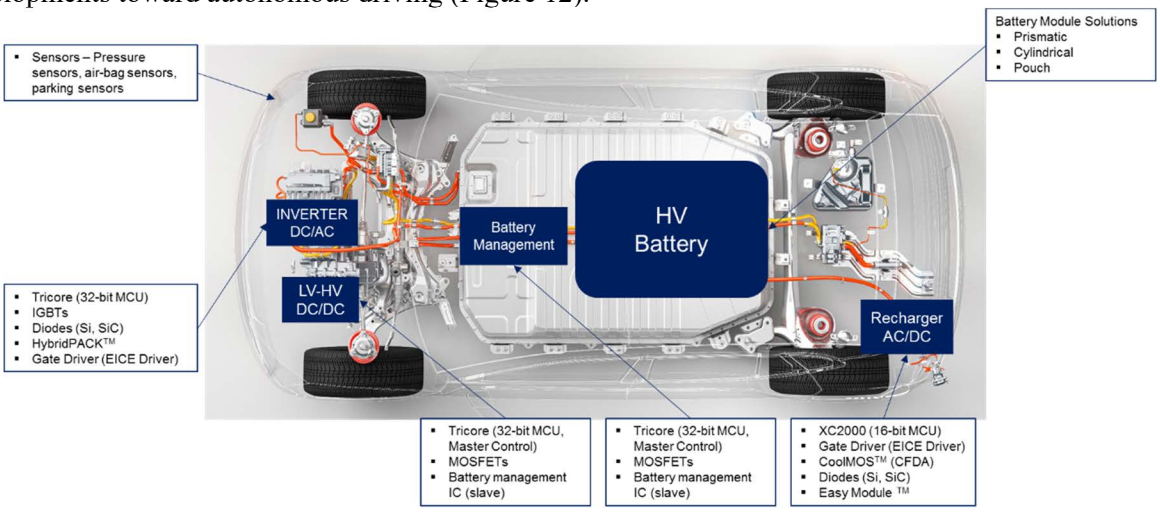


Figure 12: Examples of Automotive Wedge Bond applications [15]

Future Challenges and Roadmap towards Autonomous Wedge Bonder

High-volume manufacturing in battery and other automotive applications will remain a key driver for continued innovation for high volume production. The wedge bonder roadmap needs to support smart factory and industry 4.0 initiatives for product traceability. At the same time, wedge bonding must compete with other interconnect technologies that are also viable for high-volume manufacturing, such as laser welding or resistance welding [16].

New consumables and expanded capabilities are being developed to enable the transition to Cu wire bonding. Features enabling more and more autonomy are being deployed on wedge bonders to meet the desire for factory automation towards 'lights-out' factories.

Other Emerging Wire Bonding Applications

Many breakthroughs in packaging solutions were enabled recently through Fan-out Wafer Level Packaging (FOWLP). A cost-effective process has been developed using wire bonding to form vertical free-standing copper wires as interconnections. This vertical interconnect process may be performed on existing wafer-level wire bonders. It has been demonstrated as feasible by IME (Figure 13).

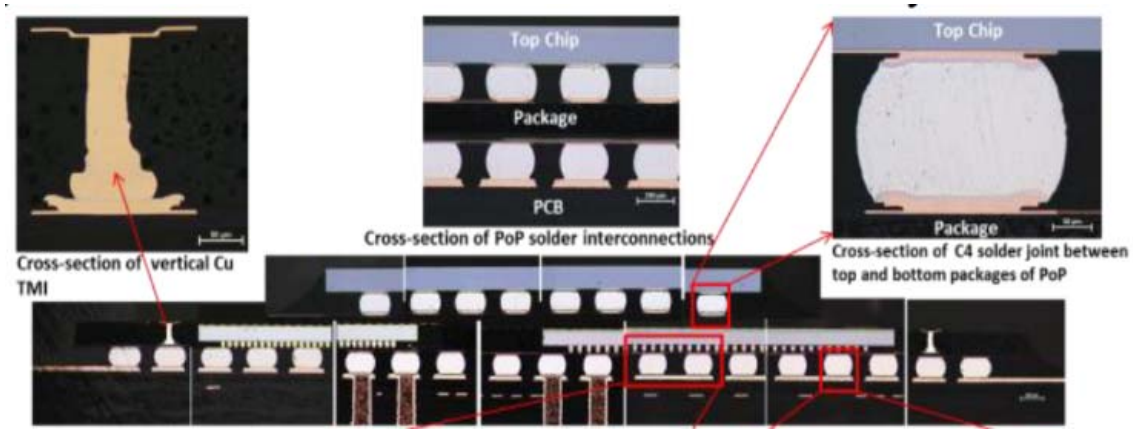


Figure 13. Cross section of assembled FOWLP package with wire bonded Vertical Interconnects [17] (Source: IME)

There is also growing demand to use wire bonded wire and loops as EMI shielding. A few different methods are being implemented including vertical wires, square loop across the device and U-shaped wire loop as shown in Figure 14.

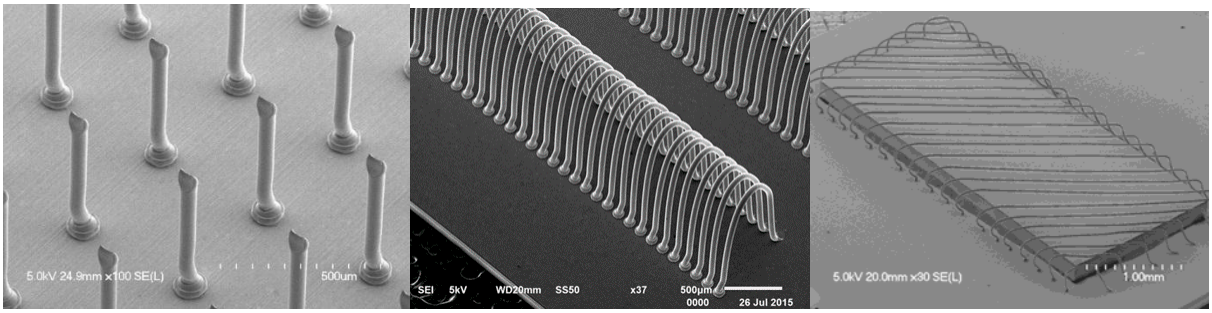


Figure 14. Wire bonded loops using for EMI shielding (Source: K&S).

REFERENCES

[1] I. Qin, J. Foley, B. Chylak, N. Wong. "Advances in wire bonding to lower the cost of interconnect", Chip Scale Review, Nov/Dec Issue, 2015

[2] Inderjit Singh et al., "Pd-coated Cu Wire Bonding Reliability Requirement for Device Design, Process Optimization and Testing", IMAPS Conference, September, 2012

[3] Hui Xu et al., "Development of Advanced Wire Bonding Technology for QFN Devices", 65th Electronic Components and Technology Conference, 2015

[4] Rich Rice, "Wire Bond Interconnect: Will It Stand the Test of Time in Era of Shrinking Technology Nodes and Increasing Technology Complexity?", IMAPS Workshop, January, 2013

[5] A. Shah et al., "Advanced Wire Bonding Technology for Ag Wire", Electronic Packaging Technology Conference, December, 2015

[6] I. Qin et al., "Optimizing Ag Wire Bonding for Memory Devices", Electronic Packaging Technology Conference, December, 2017

[7] I. Qin, "Bonding Fundamentals", Chapter 19, Semiconductor Manufacturing Handbook, Second Edition, pp 205-208, McGraw Hill, 2018.

[8] I. Qin, B. Milton, G. Schulze, C. Huynh, B. Chylak, N. Wong. "Wire Bonding Looping Solutions for Advanced High Pin Count Devices." Electronic Components and Technology Conference (ECTC), 2016 IEEE 66th, (pp. 614-621). IEEE, 2016.

[9] H. Xu, A. Shah, B. Milton, and I. Qin, "Wire Bonding Advances for Multi-Chip and System in Package Devices", IMAPS Conference, Oct., 2018

[10] F. Carson and M. Karnezos, "Advances in Stacked –Die Packaging", Global SMT and Packaging, March 2003. www.globalsmt.net.
[11] M. Karnezos, "3-D Packaging: Where All Technologies Come Together", Proceedings of 2004 IEEE/SEMI International Electronics Manufacturing Symposium, 2004.
[12] A. Shah et al., "Advances in Wire Bonding Technology for Overhang Applications", IMAPS Conference, Oct., 2016
[13] Hoet, J, and Hiebert S, "Automotive Apps are Driving Inspection Requirements for Advanced Nodes," Chip Scale Review, Volume 22, Number 4, 2018, pp. 8-9.
[14] I. Qin et al., "Smart Wire Bonding Processes for Smart Factories", Electronic Packaging Technology Conference, December, 2018
[15] G. Prasad, P. Klaerner, J. Fu "High Volume Manufacturing Requirements For EV Battery Interconnect And Power Management Applications", Charged Electrical Vehicles Magazine, Sept. 2nd Conference, 2020
[16] P. Klaerner "The advantages and limitations of wire bonding in electric vehicle applications", Charged Electrical Vehicles Magazine, Jul. 16th, 2019
[17] V. S. Rao et. al., "Development of High Density Fan Out Wafer Level Package (HD FOWLP) with multi-layer Fine Pitch RDK for Mobile Applications", Proceedings of IEEE Electronic Components and Technology Conference (ECTC), Las Vegas, NV, Jun 2016, pp. 1522-29

Section 8: Substrates

Scope

In this substrate section we shall focus on substrates used in key components in four market application areas: (a) High Performance Computing, (b) Mobile and Wearables, (c) Network Connectivity and (d) Automotive. As we address leading edge organic substrate technology for FC-BGA and for FC-CSP, we have included wafer & panel fanout as potential alternatives to conventional laminate substrate technologies.

High performance computing

FC-BGA has been the leading package for high performance computing. The key metric is defined by signal transfer rate per transfer lane and by bandwidth. Both present substantial challenges to materials and processes for the substrate. High bandwidth can be enabled through high interconnect density in the substrate. Significant advances in interconnect technologies are needed to meet the bandwidth requirement without increasing the substrate form factor and layer count.

Table1 Substrate interconnect scale roadmap

Materials	Application	Features	2018	2019	2020	2021	2022	2025	2028	2031	2034
Organic laminate	FC-BGA	Min. Bump Pitch (um), Full grid array	110	100	100	90	90	80	80	70	70
		Min. Bump Pitch (um), Periphery Staggered)	40/80	30/60	30/60	30/60	30/60	20/40	20/40	20/40	20/40
		Min. Line width/space (um)	9/12	9/12	9/12	8/8	8/8	5/5	5/5	5/5	5/5
		Min. uVia diameter (um)	50	50	50	40	40	30	30	20	20
	Chiplet (Fan-out, Organic interposer)	Min. Bump Pitch (um)	50	50	50	45	45	40	40	30	30
		Min. Line width/space (um)	2/2	2/2	2/2	1.5/1.5	1.5/1.5	1/1	1/1	0.5/0.5	0.5/0.5
		Min. uVia diameter (um)	30	30	30	20	20	10	10	5	5
Silicon	Chiplet (Si Interposer, 3D)	Min. Bump Pitch (um)	40	40	40	35	35	30	30	20	20
		Min. Line width/space (um)	0.6/0.6	0.6/0.6	0.6/0.6	0.6/0.6	0.6/0.6	0.5/0.5	0.4/0.4	0.3/0.3	0.2/0.2
		Min. uVia diameter (um)	0.6	0.6	0.6	0.6	0.6	0.5	0.4	0.3	0.2

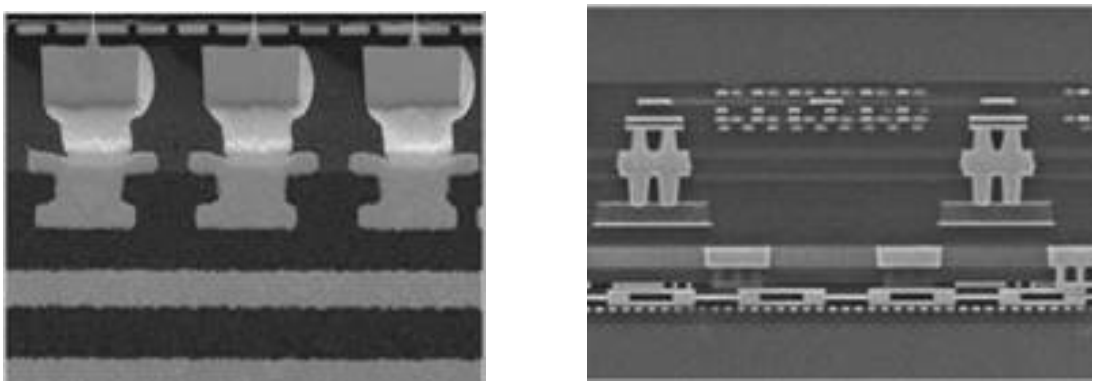
Industry has developed a system-on-a-chip (SoC), where different functions at each technology node are placed onto a monolithic die. While this is ideal in terms of performance, it becomes more costly as process technologies continue to enable higher integration with more complex integrated circuits. Some industries continue to follow this path for certain applications, but many are looking for alternatives. One way is to assemble dies with different functions at various technology nodes in an advanced chiplet package. A monolithic SOC chip can be broken down into smaller dies and mixed and matched depending on application and technology readiness and time to market. With the higher cost to yield good dies and the increased demand for high performance, chiplet technology continues to gain traction in the market. The industry has been developing a number of new and advanced packaging technologies to enable chiplets. The choice of these platforms includes silicon interposer, organic fan-out, silicon bridge embedding technologies, and silicon die stacks.

Leading-edge FC-BGA build-up substrates at 14/14um L/S are in production. Advanced laminate for EMIB is reported at 9/12um L/S. [1]. A silicon-based chiplet substrate can scale min line width to the sub-micron level without much engineering effort since it can take advantage of the existing fab backend infrastructure. However, organic-based substrates require a whole new set of materials and process development at the panel level. A potential solution would require thinner seed metals below 0.5um, high-resolution resist and lithography equipment at the panel level, in conjunction with clean room environment below Class 100K. With the need to reduce Cu roughness below 200nm of Rq, there is a need for chemical-assisted adhesion promoter development between resin and copper. Flash etching chemistry improvement to minimize roughness increase during seed metal etching may also be needed.

uVia scaling for organic laminates is challenging below around 40um using CO_2 laser and a conventional wet metallization process (desmear, electroless Cu seeding, electrolytic Cu plating). Potential solutions for continued uVia scaling include alternate uVia drilling laser sources (UV, Excimer), photoimagible dielectric materials, dry desmear, PVD seeding processes, or damascene-like via-reveal processing.

Products with bump pitch below 150um have transitioned from solder-paste printing to ball placement or electrolytic plating technologies. It is expected that ball placement technology can be extended to about 70um bump pitch, below which electroplating technology is expected to be a potential solution. The coreless substrate with embedded traces can enable much tighter bump pitch with periphery staggered bump layout. This would be an alternate way to tighten the bump pitch and reduce layer count by scaling the critical dimension only on the surface layer.

One of the challenges in fine pitch bumping is to enable mixed bump pitch with different pad sizes and uniform bump height. However, due to the very fine pitch, the solder bump stand-off height must be low and its variation has to be very tight. As the bump pitch scales aggressively, conventional solder-based flip chip packages are on the verge of migrating to advanced diffusion bonding using Cu-Cu and Au-Au. [9]



(a) Flip Chip Solder Based Interconnect (1995+) (b) Cu-Cu Diffusion Based Interconnect (2021+)

Figure 1 First Level Interconnect roadmap

With the full-grid-array fine-pitch products moving from mass reflow to thermal compression bonding (TCB) for the die assembly process, a critical parameter for assembly will be the substrate thickness uniformity within the die. The challenge is more crucial for larger form factors and higher stack-up organic-based substrates. Process optimizations are needed to ensure substrate thickness uniformity is within the die assembly process window.

Test and visual inspection is also challenging for substrates with these tighter pitches and larger number of bumps. A scalable and cost-effective test solution is required in terms of design, probe technology, and probe materials development.

As power density increases for high-performance computing applications and the vertical interconnect feature size scales, the demand for current-carrying capability of the vertical interconnect will continue to increase. This requires materials and process development to improve the current-carrying capability envelope for the first-level interconnect (FLI), uVia, plated-through hole, and second-level interconnect.

High performance computing including chiplets must do more, be faster, and burn less energy. given that it has an added layer of interconnect interface for chip-to-chip communication compared to the monolithic SOC system. High density interconnect and high-speed signal transfer rate are key enablers to match the performance of the monolithic SOC by increasing the bandwidth. However, electrical losses are becoming predominant as data transfer rate increases. Insertion loss consists of four components of loss: dielectric, conductor, leakage and radiation. Radiation loss and leakage loss are not an issue when properly designed. Dielectric loss is mostly related to the loss tangent of the laminate build-up material, while conductor loss is primarily from conductor surface roughness.

Low dielectric loss material typically has a low polar molecular resin system which is resistant to desmear process post-laser drilling and hence provides weak adhesion to the conductor. In order to extend the current cost-effective substrate manufacturing infrastructure, it is essential to develop a build-up material (resin and filler) that is compatible with current desmear and electroless Cu processing and good adhesion to the conductor. For low dielectric loss build-up material (at <0.02), it is imperative to develop cost effective dry processes (dry desmear, PVD seeding process) at the panel level.

Table 2 Substrate materials and conductor roughness requirement for high-performance computing

	2019	**2023**	**2027**
Bandwidth_Package to board (Gpbs)	45	55	65
BU dielectrics loss, Df	0.007	0.004	0.002
BU dielectric roughness (Rq)	300~400 nm	150~200 nm	100~150 nm
Cu roughness (Rq)	350~400 nm	200~250 nm	50~100 nm

Traditional power delivery systems use off-chip voltage regulators to supply the required voltages and currents, but this lengthy power delivery network (PDN) that connects them to the chip results in significant parasitic effects. This becomes a significant issue, especially for the high-performance computing system with higher total design power (TDP). Recently, fully integrated voltage regulators (IVRs) are gaining traction since this enables dedicated voltage domains. However, IVRs are typically much less power-efficient and quite sensitive to the package dimensions and process variation. Embedding passive components inside the package is an efficient way to reduce the ohmic resistance to improve the power delivery due to the shorter interconnect length between passive components to die compared to the discrete component attached on the land side or die side. Along with the development of highly efficient passive components, it is essential to develop an innovative way to embed the larger number of passive components (capacitors and inductors) inside the package.

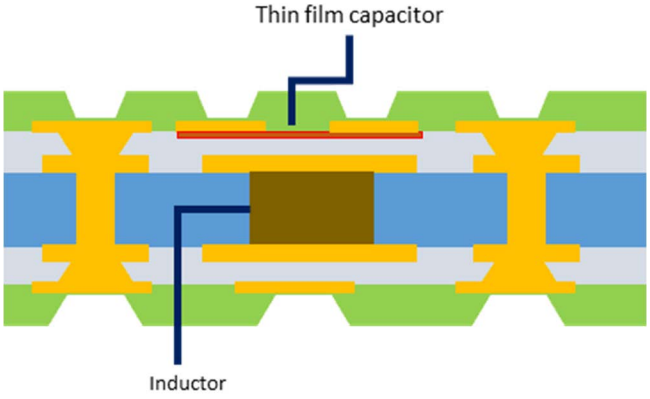


Figure 2. A schematic to demonstrate thin film capacitor and inductor embedding inside the substrate

Typically, an inductor is too bulky to embed inside the buildup. The core may provide enough space to embed the inductor, but this requires innovative inductor materials that are compatible with the substrate manufacturing process. Embedding a few capacitors inside the core has been commonly used in industry, especially for mobile applications where it requires a thin profile for the package. However, it is quite challenging to embed hundreds of

passive components in the core for high-performance computing. High density thin film capacitor embedding can be an alternate way to avoid the core space issue, as illustrated in Figure 2.

Mobile and Wearables

The Smart Phone has become the major consumer electronics product across the globe. According to the International Monetary Fund (IMF) [2] 1 billion units were sold in 2017, one for every fifth person on earth. The main trends from smartphone consumers are the continuous demand for more usability, longer battery life, and affordability. At the same time the smartphone industry rolls out new models every year incorporating new functions, better performance within the same form factor and faster processing speed. The strategy has been to develop different sets of different functional System-in-Packages (SiPs) or modules, each miniaturized and qualified. Going from one smartphone product generation to the next, the product design and release cycle would be contained within each of these functional SiPs. Smartphones today may have twenty or more SiPs. Perhaps the most important SiP is the Package-on-Package (PoP) technology. In this PoP approach, two packages are stacked on top of each other. Each package is fully assembled on its own substrate. The PoP design, with electrical and mechanical interconnect platform built in, enables integration of ASIC/logic with memory into a single PoP package. PoP approaches are most common with the baseband or application processor assembled in FC-CSP format and memory mounted on the top package. A major advantage is that the devices can be fully tested before assembly. This is critical since the processors are often at the newest nodes where the wafer yield needs to be closely watched as the foundry product ramps up.

For FC-CSP in the PoP application, a coreless substrate is one of the most cost-affordable low z-height packaging options, since it can still leverage most of the existing processes and materials that are available in substrate manufacturing. The major challenges are (a) it needs to increase I/O density by decreasing the line and space below 15/15 um L/S (as far as 10/10 um), and a blind via diameter below 60um. The second challenge is package warpage control for robust package-to-package assembly.

Shown in Figure 3 is the Qualcomm 855 package in the Samsung Galaxy 10 Smart Phone based upon a 12.5 x 12.4mm MCeP style PoP. It has a 100 um thick die with 25 um Cu Pillar at < 100 um pitch. The top substrate is for mounting the memory stack. The processor die, 100 um thick, with 25 um Cu pillar at < 100 um pitch is assembled on a lower FC-CSP substrate. The embedded trace substrate (ETS) is 10 um L/S, 130 um thick with 55 um diameter via. [3]

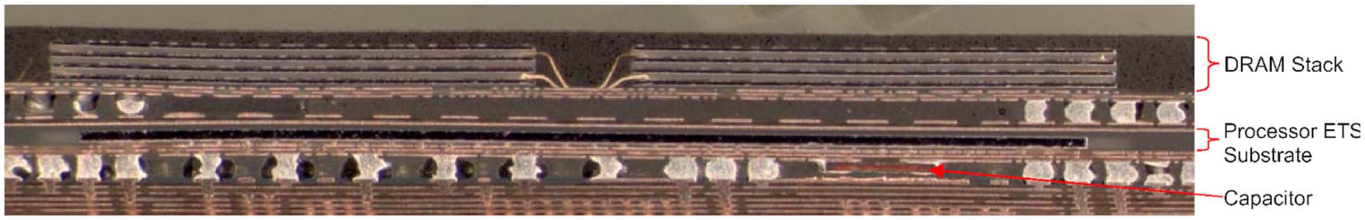
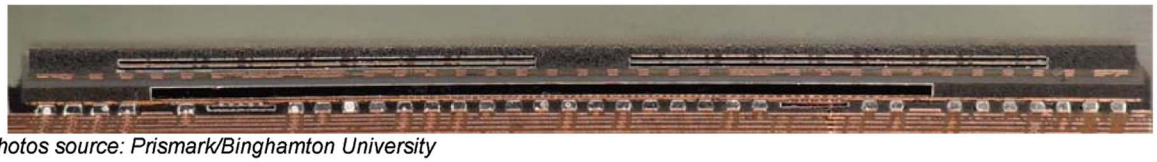


Photo source: Prismark/Binghamton University

Figure 3. PoP cross-section for Qualcomm 855 processor in Samsung Galaxy 10 Smartphone (source Prismark Partner)

Fan-out technology is another effective z-height reduction solution for semiconductor devices. This could provide a smaller package footprint with higher I/O density along with improved thermal and electrical performance due to reduced z-height and short circuit distance from motherboard to die. For in-depth review in wafer level and panel level fan-out roadmap, please refer to the WLP – Fan-in and Fan-out Chapter of the HIR Roadmap.

In the Apple iPhone, the A12 processor is packaged using PoP architecture in TSMC INFO Fanout technology [3,4] as shown below.



Photos source: Prismark/Binghamton University

Figure 4. Cross-section of Apple iPhone A12 processor. source Prismark Partners

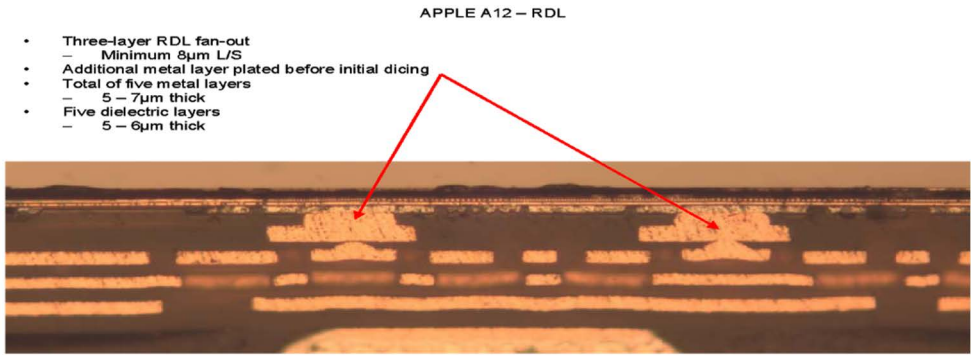


Figure 5. Expanded Fanout RDL cross-section. source Prismark Partners

TSMC's Fanout (InFO) process has been implemented in the Apple iPhone for its application processors since the adoption for the A10 processor in 2016, the A11 in 2017, and the A12 in 2018, and also used for the application processor in Apple S4 watch in 2018 . Samsung has implemented its Panel Level Fanout technology [5] in its Galaxy phone and smartwatch since 2018. Readers are encouraged to read the WLP Fan-in and Fan-out Chapter of the roadmap for in-depth discussion.

Network Connectivity

With the exponential growth of data traffic in the global digital economy and the arrival of 5G [6], network bandwidth and latency are becoming critical factors in most connectivity applications. Applications such as video streaming/downloading and AR/VR applications will drive increases in data, likely increasing bitrate requirements towards 10 Gbps, which can only be enabled through 5G mmWave. This is described in depth in the 5G Chapter of the Roadmap.

Table 4. 5G/mmWave bandwidth and substrate materials requirements

	2015	2019	2023
Frequency (GHz)	2.4/5.0	28	28/60 (mmWave)
Dielectric loss	0.01	0.006	0.002
Conductor/Build-up roughness, Rz (um)	7	3	1.5
Build up roughness, Rq (nm)	500~600	400~300	100~200

Network systems applications require a large FC-BGA package for the coming 5G mobile broadband, and wider spectrum for higher bandwidth data communication. This requires higher I/O counts and lower signal transfer loss. High I/O counts require substrates with fine line and space and fine vertical interconnect pitches (uVia and plated through hole). Low loss signal transfer rates require low loss dielectric materials and smoother conductor surfaces. These substrate requirements are similar to those required for High Performance Computing and Data Center in the previous section. The network systems would have an operating environment and operating life requirements different from the High Performance and Data Center needs.

Autonomous car

With the increasing capabilities in Advanced Driver Assist System (ADAS) from Level 1 to Level 5 and fully electric vehicles, automotive electronic systems are increasing in computing power and connectivity. The Automotive Electronics Council (AEC) [7] has published standards for verification of electronic components reliability requirements: AEC – Q100 Rev H (ICs), AEC – Q101 (Discrete), AEC – Q102 (Discrete Opto), & AEC Q104 (Multichip Module). AEC–Q100 Rev H published 9-14-2017 "Failure Mechanism Based Stress Terst Qualification for Integrated Circuits" provided requirements for Grade 0, Grade 1, Grade 2, Grade 3 based upon their function. AEC-Q104 published 9-14-2017 "Failure Mechanism based stress test for Multichip Modules" to address increasingly complex heterogeneous integration packages such as SiPs, and MCMs which includes components that passed AEC-Q100, 101 and 200 before being integrated into a package. It includes board level reliability (BLR) and ESD requirements.

ADAS and autonomous vehicles require a high level of computing power, in-vehicle networking and sensor integration, requiring multi-die large form-factor package integration with advanced packaging. Substrate process

design and materials are at the front line for innovation, including core materials, buildup materials, and solder masks. Materials properties such as elastic modulus, glass transition temperature, thermal expansion coefficients (CTE1 and CTE2) and their interlayer adhesion need to be carefully examined to meet Grades 1 and 0 requirements. [8]

Development of new core, build up, and passivation layer materials is needed to meet the stringent reliability and durability requirement of automotive components in heterogeneous integration. With fast transients in voltage and current causing noise which impacts nearby electronic devices, cost-effective and reliable EMI shielding technology is needed for the automotive space.

Difficult Challenges

In this substrate section we have reviewed leading-edge organic substrates for FC-BGA and FC-CSP used in four major market application areas. We have shown that wafer and panel fanout are potential alternatives, displacing conventional laminate substrate technologies in future leading-edge applications so that difficult challenges can be addressed to retain the technology edge in this highly competitive and innovative industry.

References

1. "Flip Chip & WLP: Trend & Market Projection", TechSearch International, May 2019
2. "World Economy Outlook" International Monetary Fund, (IMF), April 2018
3. "Semiconductor Packaging Report" Prismark Partners, December 2018
4. "INFO (Wafer Level Integrated Fan-Out) Technology" Chien-Fu Tseng et al ECTC 2016
5. "Study of Advanced Fan-Out Package for Mobile Application" Taejoo Hwang et al ECTC 2018
6. "The Mobile Economy 2019", GSM Association, February 2019.
7. Automotive Electronics Council (AEC) standards publications: AEC – Q100 Rev H (ICs), AEC – Q101 (Discrete), AEC – Q102 (Discrete Opto), & AEC Q104 (Multichip Module). AEC–Q100 Rev H published 9-14-2017 "Failure Mechanism Based Stress Terst Qualification for Integrated Circuits"
8. "Challenges and Approaches for Developing Automotive Grade 1/0 FCBGA Package Capability" Rajan Dias et al, ECTC 2019.
9. "Heterogenous Integration An Advanced Packaging View" Hamid Azimi, ISES 2020.

Section 9: Board Assembly

Board Assembly Overview

Board-level assembly encompasses those assembly operations required to manufacture the final functional electronic sub-assembly to be incorporated into the end product, including rigid circuit board SMT assembly as well as assembly to flexible and non-planar (e.g., 3D printed) structures.

9.1 Board Level Interconnect Density

The density of package interconnects to the printed circuit board varies by industry sector. Package size and density will vary for specific applications, but common package attributes within industry sectors are listed in Table 1. It is expected that heterogeneous integration will require the mixing of varying I/O pitches and interconnect dimensions at various levels in the package and manufacturing sequence.

Table 1. Typical attributes of area array packages by industry sector with expected I/O increases.

	Package Body (mm)				Package I/O pitch (mm)			
Year of Production	**2018**	**2023**	**2028**	**2033**	**2018**	**2023**	**2028**	**2033**
Server/Data Centers	65	70	70	75	1.0	1.0	0.8	0.8
Smart Mobile	32.5	35	37.5	40	0.8	0.8	0.65	0.65
Aerospace/Defense	27.5	33	33	35	0.65	0.65	0.5	0.5
Automotive	27.5	31	31	33	0.8	0.8	0.65	0.65
Wearables/Health					0.4	0.35	0.3	0.3

In every industry sector, board interconnect densities will increase with increased package integration levels. Some anticipated changes are also included in Table 1. Finer pitches and increased package body sizes will be introduced in every market sector. Increased package I/O density will drive higher board layer counts and increased use of stacked microvia structures to escape the larger I/O arrays, raising challenges of assembly solder defects and board-level reliability.

The attributes of leading-edge packages drive the assembly innovations required for next generation product. Leading interconnect pitches anticipated in each of the industry sectors are tabulated in Table 2. Use of reduced package pitches are most often limited by PCB wiring escape or substrate warpage concerns, both of which are primarily associated with area array packages. Table 2 therefore lists only area array package pitches.

Table 2. Board Level Interconnect Pitch (area array packages - leading edge capability)

Year of Production	2018	2019	2020	2021	2022	2025	2028	2031	2034
BGA/LGA Solder Ball Pitch (mm) Conventional system board									
IoT	0.4	0.4	0.4	0.4	0.4	0.35	0.35	0.35	0.35
Autonomous Vehicles	0.65	0.65	0.65	0.65	0.65	0.5	0.5	0.5	0.5
Smart Mobile	0.35	0.35	0.35	0.35	0.35	0.3	0.3	0.27	0.27
High Performance/Data Center	0.8	0.8	0.8	0.65	0.65	0.65	0.5	0.5	0.5
Aerospace/Defense	0.65	0.65	0.65	0.65	0.65	0.5	0.5	0.4	0.4
Wearables & Health	0.4	0.4	0.4	0.35	0.35	0.3	0.3	0.27	0.27

9.2 Difficult Challenges

Board assembly invariably involves elevated-temperature processing. The most ubiquitous problems are those associated with heating the heterogeneous material sets comprising electronic packaging structures to the peak temperatures required for lead-free solder reflow processes (245°C).

9.2.1 Temperature Induced Distortion

Complex Module Warpage

The asymmetric structures inherent in heterogeneous packages produce complex warpage shapes during board assembly, raising the risk of soldering defects. Finer interconnect pitch means smaller solder bumps and reduced tolerance to warpage or out-of-plane distortion. Warpage engineering considerations are reviewed in more detail in the Mechanical Requirements section.

Head-on-Pillow and Non-Wet Open Soldering Defects Risk

A well-known consequence of module and board warpage is the formation of Head-on-Pillow (HoP) or Non-Wet Open (NWO) solder defects in BGA solder joints. Package distortions at peak reflow temperatures separate package solder bumps from board surface solder paste. Increasingly complex heterogeneous packaging structures increase the severity and unpredictability of soldering challenges posed by temperature-induced distortions.

Via-in-Pad, Plated-Over BGA Pads

To accommodate increased levels of integration, circuit board wiring designs require increasing use of Via-in-Pad, Plated Over (VIPPO) structures. With mixed VIPPO ball grid footprint design, localized differential thermal expansion poses the risk of separation of the VIPPO joint, typically at the component-side intermetallic interface, after repeated reflows.

Potential Solutions for Temperature-Induced Challenges

Reduced reflow temperature minimizes solder problems with warpage and distortion at high temperature simply by avoiding the high temperature. Various approaches to package interconnect formation at reduced temperatures are being explored. These include:

- low melting temperature solder alloys (e.g., eutectic-based BiSn)
- nano- and micro-particle sintering pastes (e.g., Ag or Cu)
- supercooled molten solder beads (e.g., SAFI-Tech)
- conductive adhesives, especially with liquid metal fillers (e.g., Sekisui Self Assembly Paste)

Significant research and development effort will be needed to bring these approaches to practice.

9.2.2 Diversity of Feature Sizes

Integration of heterogeneous packaging technologies may require combining die-level interconnects on the scale of tens of micrometers, packaging interconnects at tens of millimeters, and board assembly operations at tens of centimeters. Designs requiring device-level interconnects developed for semiconductor packaging technologies for assembly directly onto board-level structures will be particularly challenging.

Driven by the demand for mobile electronic products, wafer level packages, passive devices and memory packages are being introduced in smaller formats. Such smaller formats can be placed in closer proximity to the processor function, further enhancing electrical performance. However, these design advantages pose challenges for conventional SMT tools and processes.

Broadband Solder Paste Printing

Heterogeneous structures requiring diverse solder interconnect feature sizes, placed in proximity, require some joints to be processed under non-optimum conditions. The paste print resolution required for fine pitch passives or CSP memory is difficult to achieve in proximity to large-body BGA SiP footprints.

High Accuracy Device Placement

Die stacking technologies and other semiconductor integration implemented at the device packaging level routinely require placement accuracies of 5µm or better. Routine board assembly manufacturing placement operations occur over tens of centimeters with an accuracy ~15µm. Heterogeneous integration of fine-pitch devices directly onto large-scale board assemblies encounters a severe capability mismatch of accuracy and speed. System in Package (SiP) designs forestall this heterogenous accuracy challenge by constraining high-accuracy placements to the package assembly, which can then be placed as a conventional large body component.

Extreme Proximity Rework

Close proximity device placement can only be tolerated in large, expensive system boards if methods exist for manufacturing rework of defective devices. Current hot gas rework tools are limited to removing and replacing devices with >1.8mm spacing to adjacent components.

Selected-area laser reflow methods are now in development as potential means to rework close-proximity devices. While promising for passives and small active devices, laser rework will be challenging for reworking thermally massive components, and further innovations will be needed.

9.2.3 Reflow Cool-down

Laminate pad cratering from reflow cool-down has been a known issue for designs coupling thick, stiff circuit boards with large PBGA components. High levels of integration requiring board attachment of large body SiP sub-assemblies will meet with similar pad-cratering challenges.

9.2.4 Flexible Substrate Assembly

Substrate Temperature Capability

Polyimide and liquid crystal polymer based flexible circuits, currently used in defense and aerospace, are too costly for many wearable and health monitor applications. Flexible circuitry in the wearable and health monitor sector are therefore often based on low-cost polymers having limited elevated temperature capability. Pb-free solder connections are not viable for these materials.

Printed Ink Joint Integrity

Printed electronics are being actively pursued as a means to produce high-volume, low-cost circuitry for disposable applications such as wearable health monitoring devices. Significant development is required to bring printed ink circuitry to sufficient physical integrity and chip-joining quality for general use.

9.2.5 Additive Manufacturing of Electronics

The infrastructure for assembling various electronic devices to 3D printed electronic structures does not yet exist. Since 3D printed electronics are not limited to conventional planar packaging structures, the resulting component mounting surfaces may be in any orientation. The required component assembly tools must therefore be able to place and join both functional and passive devices at arbitrary orientations in three-dimensional space. A full description of additive electronics will be found in a separate section of this chapter.

9.3 Board Assembly Supply Chain Requirements

9.3.1 Assembly Materials

Solder Pastes and Fluxes

Broadband Printing: Solder pastes capable of printing with high transfer efficiency over a wide range of stencil aperture sizes would alleviate many of the challenges with diversity of feature sizes.

Low Melt Soldering: Pastes with low melting point solder alloys that can be reliably used for the attachment of various SnAgCu-based solder-preformed components are necessary for establishing a sequential hierarchy of solder melting point through the final board assembly operation.

Laser Reflow Soldering: Fast acting solder pastes and fluxes optimized for the extreme soldering rates of laser reflow soldering are a prerequisite for wider industry adoption of this rapid local joining method, anticipated to enable attachment to arbitrary bonding surfaces posed by 3D printed electronic structures.

Vacuum Reflow Soldering: Other flux carrier formulations optimized for low-pressure operation may well be advantageous for high-yield vacuum reflow operations.

Metal Sintering Pastes

Metal sintering pastes that sinter at relatively low temperatures (<200C) without the need for applied pressure during the sintering process are critical for packaging of high-power devices, including light emitting diodes for lighting applications.

9.3.2 Assembly Manufacturing Tool Requirements

Component Placement Tools

Currently available board assembly placement tools will require additional capabilities to adequately address the needs of heterogeneous packaging integration manufacturing. These include:

- Wafer feeder and die ejection tooling for picking ultrathin die (<50μm) directly from dicing tape for placement on boards or flexible substrates
- Higher accuracy placement (<5μm) of fine pitch devices in the board assembly process
- Placement with heated spindles and preheated stage for tacking sintered metal joints

Solder Reflow Tools

Selective Area Laser Reflow: Selective area laser soldering tools would be invaluable for localized device attachment in close proximity to highly temperature-sensitive components such as optical transceivers, as well as providing solutions for temperature-sensitive assembly challenges such as the solder attachment of sensors and electronic controls to Li-ion battery systems.

SMT Rework Tools

Complex heterogeneous assemblies will require novel rework methods and rework tooling with the ability to apply local heat with extreme precision and to handle fine passives. Reattachment methods will need to permit finer pitch interconnects including fine pitch control of replacement interconnect material.

9.4 Summary

Challenges posed by heterogeneous package integration to conventional board assembly operations include those arising from:

- Higher I/O package requirements driving added structural complexity into the supporting circuit boards leading in turn to various assembly-induced yield and reliability problems,
- Complex temperature-induced warpage behavior of heterogeneous package structures producing unpredictable and unreliable solder joint formation, and
- Immature materials and manufacturing infrastructure to support flexible circuit assembly.

For those designs requiring heterogeneous integration on the PCB rather than on package, the diverse scales of interconnect dimensions will pose significant challenges to the board assembly process to reliably form large numbers of interconnects over relatively narrow ranges of feature sizes. Repeatedly manufacturing fine interconnections positioned over large distances, or joining fine features in close proximity to coarse features, will require improvements in tools, materials and manufacturing practices.

Section 10 Additive Manufacturing

A. Background and Overview

Additive manufacturing (AM) refers to a wide class of 3D printing technologies ranging from laser-based metal printing approaches to the jetting of photocurable resins. Many of these technologies could be applied towards advancing SiP technology (i.e. creating complex geometry encapsulation packaging), but one AM segment stands out as a significant potential contributor: Additive Electronics (AE). Also referred to as 3D Printed Electronics, AE itself represents a class of additive manufacturing technologies. AE generally refers to systems which can print dielectric material and conductive material selectively within a volume, with these volumes being uniquely defined for each print. Some AE technologies can only do this by slowly building up material microns at a time or only in

stacked layer geometries (no overhangs or internal structuring). Such approaches could be considered as 2.5D printed electronics and will not be the focus of this section. Also not covered are progressions in flexible electronics and 2D printed electronics, which both have significant current and future offerings for SiP integration. As all these fields develop, a combined solution will likely be found as optimal, incorporating aspects of flexible electronics, 2D printed electronics and AE in concert with traditional SiP manufacturing methods.

By comparison to AM, AE is a newer field with a growing body of academic research and few examples of commercial realization. The current classes of AM which have been leveraged for AE include Fused Deposition Modeling (FDM), Direct Write (DW), Powder Bed Fusion (PBF), Inkjet Printing (IJ) and Photo Resin Jetting (PRJ) (Figure 2). Combinations of one or more of these AM approaches are used to create AE methods. A critical development for AE in order to provide a robust SiP solution is the integration of lumped components into printed parts. This can be done by the incorporation of Pick and Place (P&P) capabilities within an AE apparatus, or by utilizing a print pause + resume (PP/R) approach, wherein part printing is paused, the partial part is moved to another manufacturing apparatus for component placement and is then moved back to the AE apparatus for print resumption. The PP/R approach is especially interesting as other electronics manufacturing methods (i.e. wire-bonding) could possibly be used on the printed partial part in between the pause and resume steps. AE solutions like this are under development both in the literature and commercially.[1-3]

As applied to SiP fabrication, AE could reduce costs of and time to part, eliminate currently necessary manufacturing steps, and allow the SiP form factor to be equivalent to or near that of the final product form factor. Reduced costs and time to part could come from replacing traditional multi-stack PCB manufacturing, which currently can take on the order of multiple weeks for manufacturing and shipping, which significantly impedes time-to-part and ease of design iteration. An AE solution with layer times of seconds could allow for fabrication of prototypes and final designs at a fraction of the time and cost. Regarding the elimination of currently necessary manufacturing steps, an optimized AE technology with P&P could allow for the concurrent fabrication of a multi-stack PCB, embedded components, printed or placed passives, placed dies, printed interconnects, printed encapsulation, high aspect ratio vias, printed antennas, and other necessary SiP components (Figure 1). A fully developed AE technology could therefore allow for a highly streamlined SiP manufacturing solution which eliminates multiple manufacturing steps. Finally, given the geometric freedom when additively manufacturing objects, AE could allow for the final product form factor to be created during SiP fabrication, eliminating the need for numerous fabrication steps. A vision of how AE could revolutionize SiP fabrication is presented in Figure 1, with approximate timeframes until AE matures to the point where SiP manufacturing is feasible. Additional development time is assumed past these timeframes before traditional fabrication methods might be replaced by AE methods.

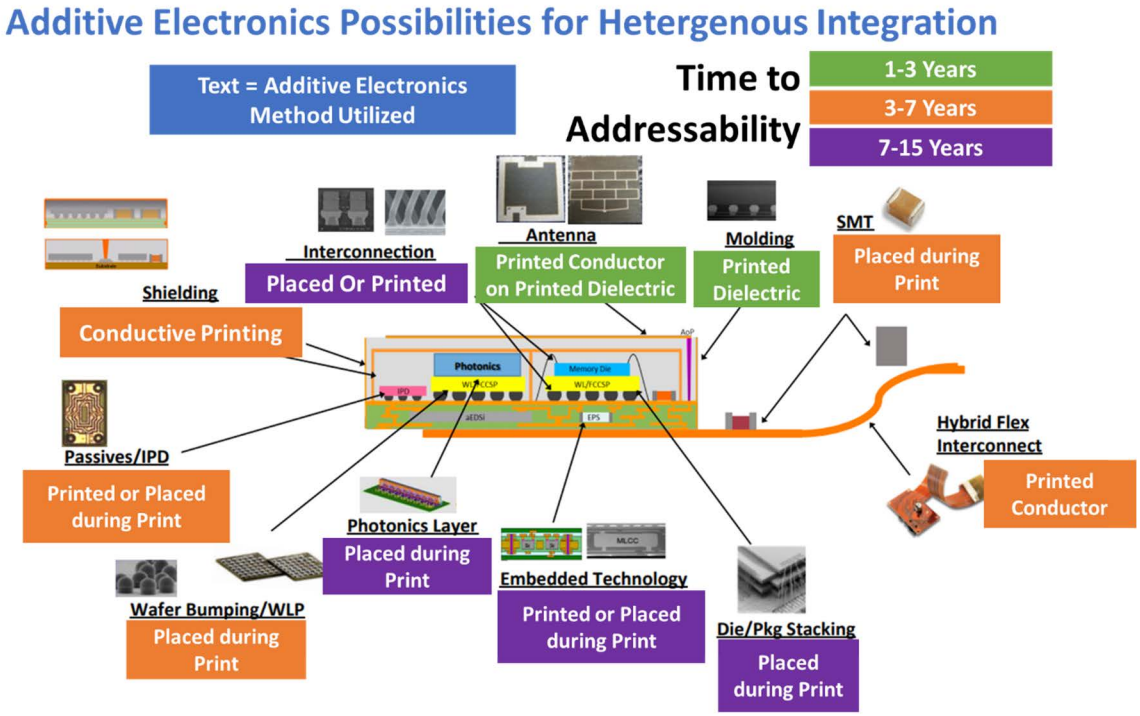


Figure 1. Possible uses for AE within SiP applications for heterogenous integration with estimated time horizons for development.

However, currently AE technologies are largely under-developed for SiP applications, making them more suitable for low-complexity consumer electronics. Herein, we will overview the current state of AE, highlight the gaps necessary for AE to offer solutions to SiP applications, and outline a roadmap for research and development necessary to fill those gaps.

B. Current State of AE Technologies

A number of AM technologies have generally been applied towards AE, with a few reviews being available.[3-6] The main AM technologies are outlined in Figure 2, with AE approaches often combining basic AM printing methods. We will briefly describe basic AM methods and then cover associated AE methods which rely upon these basic methods.

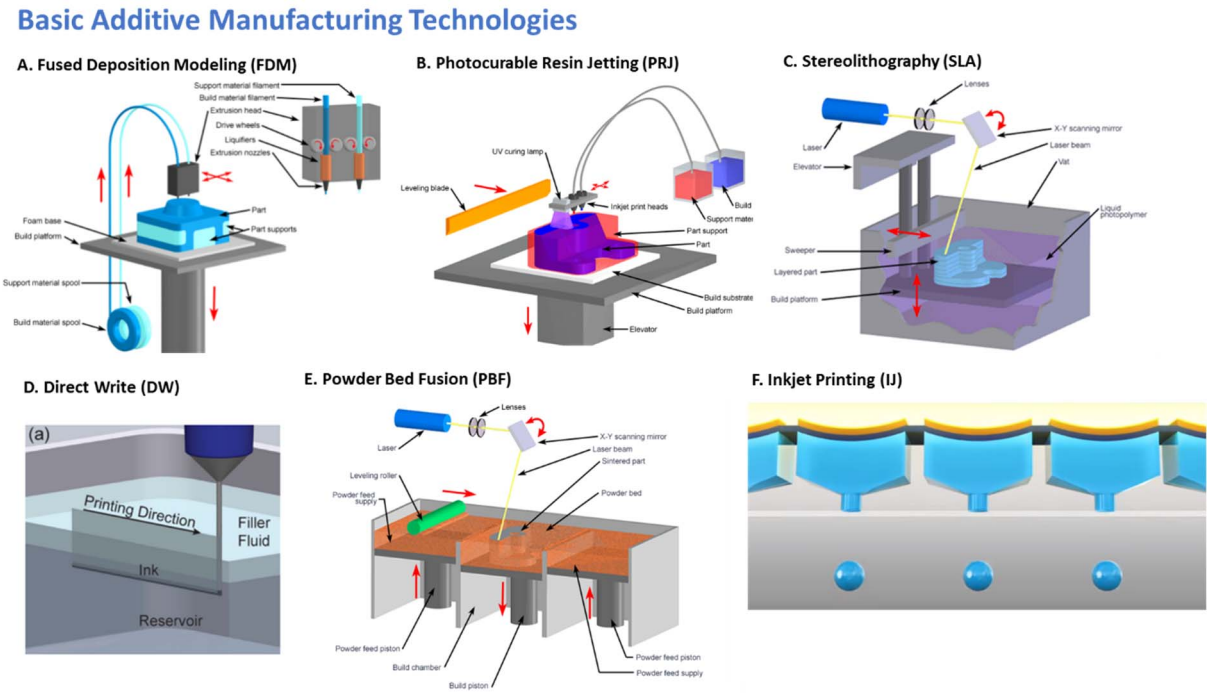


Figure 2. Basic AM technologies, which form the basis for AE [7,8]

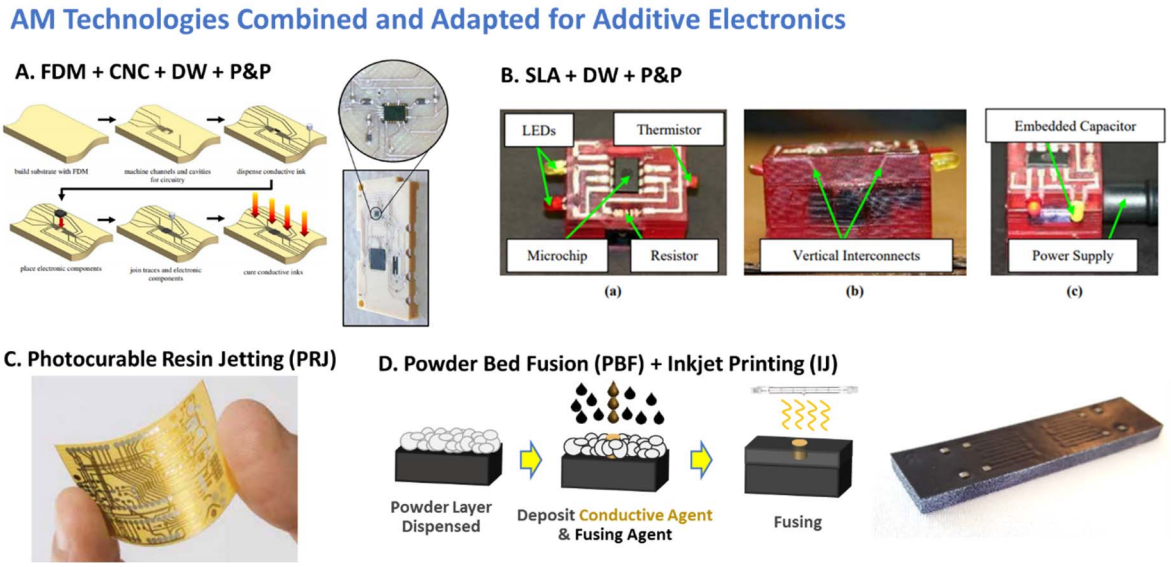


Figure 3 AE technologies, which often combine one or more AM approaches and conventional manufacturing approaches. [A,3 B,2 C,8 D,8]

FDM consists of one or more spools of polymeric-based material being melt extruded through a nozzle with parts being built in a layer-wise fashion and with each region having a specific material being selectively deposited (Figure 2). For AE, one of the materials is loaded with a highly conductive composite (such as graphene-filled polymer) for

the selective deposition of conductive features.[8] Alternatively, FDM can be combined with DW with highly conductive pastes (usually silver-based) being extruded during the FDM print (Figure 3A). The combined approach can be complemented by P&P and CNC, with components added during or after the print.[3] FDM benefits include high mechanical properties and multi-material printing, with detriments being build speed (<10 mm/hr typical), low resolution (>100 microns typical), support materials being required, and singular parts being built at a time.

PRJ jets droplets of UV curable resin followed by curing with UV lamps (Figure 2B). PRJ is also amenable to multiple materials. For AE, a photo resin with silver is jetted along with a dielectric support material (Figure 3C).[9] Benefits include high resolution (<20 microns typical) and good surface finish, with detriments being slow build speed (<5 mm/hr) and difficulties in creating wide build beds.

SLA uses a photocurable resin which is selectively cured by a laser to create a layer, with more photo-resin being deposited for each subsequent layer (Figure 2C). For AE, researchers have taken advantage of the ability to stop and restart the SLA, combining it with DW of silver pastes for conductive traces as well as P&P for component integration (Figure 3B).[1-3] Benefits include a high build rate (>15 mm/hr) and good surface finish, with detriments including the need for supports and incompatibility with multiple materials (unless PP/R is implemented).

DW extrudes high viscosity fluids in a desired pattern (Figure 2D), which can then be cured (laser, UV, etc.). For AE, DW is used in combination with other AM methods which create the object while DW creates the conductive features.[1,7,10-13] Benefits include high conductivity traces (up to 11.8E-8 Ωm) [3] and adaptability into other AM methods.

PBF consists of spreading thin layers of a powdered polymeric material which is selectively fused either through the use of a laser or in a combined approach using inkjet printing and a light source (Figure 2E). For AE, conductive agents (3D inks) can be used to create conductive features selectively within a fused dielectric material. Benefits include high mechanical properties, high build rates (>25 mm/hr), moderate resolution (~100 microns), no need for supports, and printing of multiple parts at a time, with detriments being that combined approaches using DW and P&P are difficult.

Commercial AE examples exist for some of the described methods. nScrypt has a FDM + DW commercial printer which also has P&P capabilities.[8] Nanodimension has a PRJ AE commercial printer, specialized for the fabrication of multi-stack PCBs.[14] HP has presented a pre-commercial, research-level PBF + inkjet technology capable of producing highly conductive traces within a dielectric polymer material.[15] Another common AE commercial tool is Optomec's Aerosol Jet technology, based upon Aerosol Jet Printing (AJP). This approach can be utilized for printing conductive features onto 3D parts,[16] or used to eventually build up material in a full 3-dimensional form factor.[17]

C. Technology Gaps and Research Needs for AE to offer SiP solutions

None of these AE approaches currently provides an optimized SiP solution, with the main developmental needs falling within a few areas: printing characteristics, substrate characteristics, and additional process integration (like P&P or PP/R), summarized in Table 1.

Table 1. Important developmental areas for AE to provide SiP solutions

Development Area	Current Best State	AE Approach for Current Best	Desired State (Depends on use-case)	Developmental Challenges and Suggested Research Areas (Depends on use-case)
Printing Attributes				
Line Width	>40 µm	PRJ, IJ	≤40 µm	Making robust to all print conditions and geometries. Larger line width approaches (DW) brought to inkjet resolution. Higher resolution on inkjet.
Space Width	>100 µm	PRJ, IJ	≤150 µm	Making robust to all print conditions and geometries. DW not at inkjet levels.
Trace Conductivity	12E-8 Ωm	DW, Aerosol Jet	≤10E-8 Ωm	Making robust to all print conditions and geometries and at above width and pitch. PRJ and IJ + PBF need improvements.
Build Speed, Parts per Build	>15 mm/hr, multiple parts	IJ + PBF	Maximize for optimal utility	Improvements to build speed and number of parts/build generally difficult for PRJ and FDM + DW, but likely necessary
Substrate Attributes				
Dielectric Strength	~10 kV/mm	FDM	>15 kV/mm	High dielectric strength materials available, incorporate into AE approaches
HDT	189 °C	FDM (PPSF)	>220 °C	High temperature polymer available, needs development for AE.
Tensile Strength	70 MPa	FDM (ULTEM)	~70 MPa	Highly rigid polymers available for AE
Additional Process Integration				
Component Attachment	Amenable to P&P	FDM, SLA, PRJ	Optimized with P&P	Processes incorporating P&P not optimized: speed, interconnects, in-situ testing, resumption of printing processes, etc.
Print Pausing/Resume	Amenable to PP/R	SLA, FDM	Optimized with PP/R	Processes incorporating P/R not optimized: system integration, workflow optimization, interface mechanical integrity, etc.

For printing attributes, line width, space width and conductivity are of primary importance. Resolutions down to 20 μm and conductivities of up to 11.8E-8 Ωm have been reported using SLA + DW,[3] and space width is nominally ~2X line width. For integration with SiP components, line width should be around 40 μm with space width around 150 microns, matching well with current AE capabilities. These results still need to be demonstrated for a variety of print conditions and geometries and need development for replication using other AE methods.

For trace conductivity, AE inks are not purely metallic, so conductivities are below bulk material properties, at best 2-3X bulk resistivities. For some SiP applications, this may be acceptable, with other applications requiring improvements.

Other important attributes include high build speed and printing of different electronic-type materials for printed components. These are compelling features of AE, which could drive further adoption of AE if basic SiP requirements are met.

For substrate attributes, high dielectric strength, high heat deflection temperature (HDT) and high tensile strength are needed. Dielectric strength is needed to have good device performance and high trace densities. High HDT is required for good high temperature operation and for a solder reflow step if components are attached during or after printing. FDM shows good promise for meeting these needs with other methods needing further development.

For additional process integration into AE, significant development is needed to incorporate into AE apparatus additional tools like P&P, PP/R, wire-bond, or others such that components can be embedded into parts while they are being fabricated. Without this development, the benefits of AE will be largely diminished compared to conventional SiP fabrication methods. AE approaches more amenable to pausing and resumption of printing (like FDM, PRJ, and SLA) are more suitable for incorporation of P&P and/or PP/R approaches, but seamless integration of these features needs significant further development. Development and optimization of such hybrid systems which incorporate additional electronics manufacturing technologies (P&P, in-situ testing and characterization, interconnect technology, selective thermal treatments, wire-bonding, pressure assisted processes, etc.) either directly into an AE apparatus or within a larger manufacturing system arguably presents the most significant yet necessary developmental challenge for AE in order to offer a robust SiP solution.

D. Printed Passive Components within Additive Electronics

Passive components are broadly defined as those electronic components that do not require energy from a source for their basic operation. Therefore, by definition, they cannot amplify or provide gain, which fundamentally differentiates them from active devices. Passive components can, however, attenuate, modulate, sense, and monitor currents. By doing so, they perform various important functions such as noise suppression, energy storage and release, filtering, signal and power transfer, feedback, and terminations in electronic circuitry. A typical smartphone has approximately 1000 passives mounted on the board, with many of these passives embedded in one or more packages and in ICs. This count is 20 times higher than the number of ICs (~50) in a system. The number of passive components dominate in any consumer system, even as the electronics industry is driving relentlessly toward integrated and miniaturized systems such as smartphones. The number of passive components in iPhone 4s accounts for 79 percent of the total number of its components. On the other hand, the cost of passives account for only about 2 percent of all the components.

All passive components can be classified into discretes, integrated passive devices (IPDs), or embedded passives. Traditionally, passives have been manufactured as dscretes. Performance and miniaturization eventually led to their evolution into IPDs and embedded passives. Discrete passives are individual components that are separately manufactured, packaged with terminations and encapsulation, and surface-mounted onto the package or the board. The barriers from pick-and-place and other routing issues can be avoided by combining multiple passives on the same devices. They are broadly termed as integrated passive devices (IPDs), though they are also traditionally referred to as passive arrays or passive networks.

One approach to streamline the manufacture of SiP and associated devices would be to directly incorporate passives (resistors, inductors, and capacitors) into the substrate during fabrication. Embedded thin-film components are formed in the build-up or redistribution layers (RDL) of the package, as a part of the wafer, package, or printed wiring board fabrication processes. Integrating these passive components as thin-film components into the package leads to significant miniaturization and also improves the electrical performance by means of shorter interconnections between the chip and the package. It can also result in reduced packaging costs, low power loss, low volume, low weight, and low profile. This could be done through a combined P&P and PP/R approach, but actively printing passives through an AE process would reduce the number of steps and required tools. Indeed, there have been developments around printing passives using AE methods,[18,19] again often through the combination of multiple approaches or steps (Figure 4).

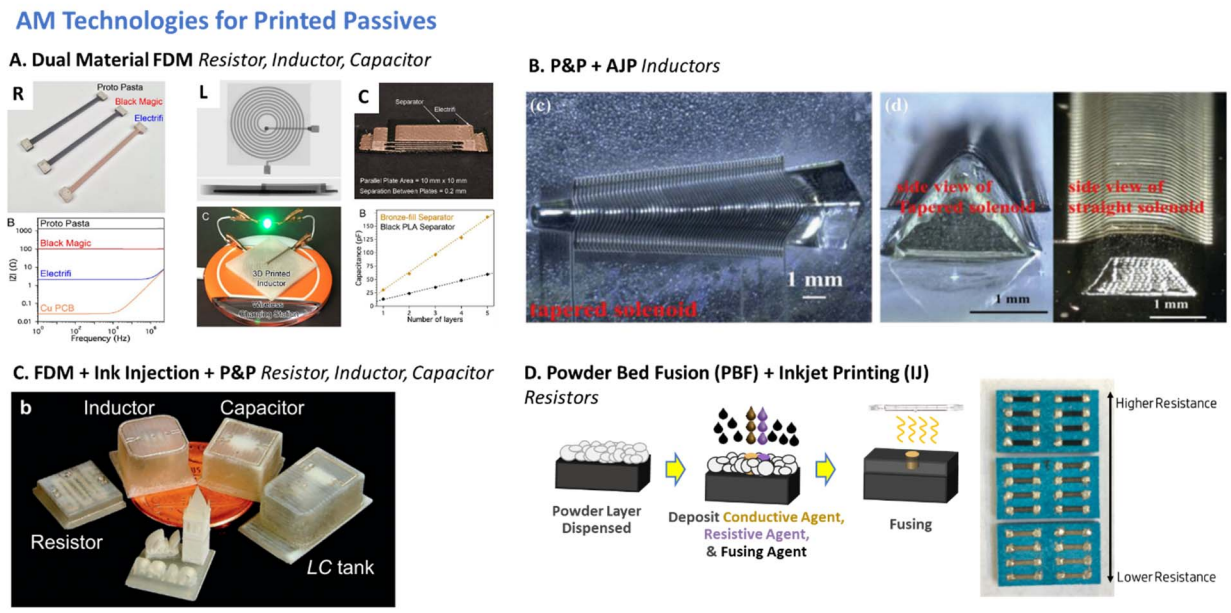


Figure 4. AE approaches for fabricating passives (R,L,C) using the AE methods indicated. [A,20 B,17,21 C,22,23 D15]

Passives in RF modules:

Realizing high-performance and miniaturized passives is the most demanding challenge for RF module miniaturization. RF passives typically include transmission lines, matching networks, couplers, filters, diplexers, and electromagnetic shields and are necessary to support the active components. Typical requirements of specific inductance, capacitance, and sheet resistance are in the range of 10–20 nH/mm², 20–500 pF/mm², and 5–50 ohm/square, respectively. Tight tolerance of passive components is required in RF applications. For RF passive components, the key metrics are Q (50-100 at 1-5 GHz for lumped RF components), insertion loss (0.1-0.2 dB), and out-of-band rejection (>25 dB). Higher performance, size reduction, and proximity to active devices have been the primary drivers for passive component evolution. AE of passives has advantages as the fabrication process does not have to be disrupted with the insertion of IPDs or discrete passives.

Advances in manufacturing tools for precision fabrication and line definition, combined with new composite ink material systems that can achieve higher permittivity or controlled resistivity, have led to additive manufacturing as a viable approach to passives. With the development of multi-material FDM printing, different filament types can be used to create resistors by using different bulk resistivity materials, with capacitors and inductors being fabricated by using dielectric and conductive materials and precise geometry control . Reported ranges (Figure 4A) include 10 Ω - 100 kΩ for resistors and 10-120 pF for capacitors (parallel plate, 5-layers).[20] Alternatively, FDM can be used to create void spaces within which conductive ink can be injected, with the possibility of post-print placing a magnet into a printed void to increase the inductance of fabricated inductors (Figure 4C).[22,23] Reported ranges utilizing this approach include bulk resistivity values of 5E-5 Ωm - 100 Ωm for resistors, 23-350 nH for inductors (spiral air-core), and 3-24 pF for capacitors (cylindrical and parallel plate, 1-layer). Another method which has been utilized for AE fabricated passives is Aerosol Jet Printing (AJP).[24] Although this method only deposits through a single nozzle a few microns of a single material at a time, multiple materials can be used asynchronously to create 3-dimensional features with separate conductive and insulating volumes.

Another method for creating printed resistors includes utilizing inkjet printing during powder bed fusion (Figure 4D), wherein different bulk resistivity agents can be jetted into the powder bed during part formation, with 9 orders of magnitude of bulk resistivity being achievable.[15] Each of these methods utilize AE methods, but all of the above mentioned approaches have focused largely upon just creating the printed passive. In order to utilize these approaches for fabricating more complex SiP-type devices, further development of both the AE approach and integration with other manufacturing methods, like P&P and PP/R, is needed. Finally, more conventional 2D printing methods can be utilized to create R,L,C circuit elements[25,26] with the possibility of utilizing these methods through a PP/R approach to create a 3-dimensional structure.

Some additional considerations for the AE fabrication of passive includes considering their potential for creating commercially relevant R, L, and C values, and doing so at a relevant size. Printed resistors look to be the most promising, with 9 orders of magnitude of resistivities reported at size scale similar to small form-factor SMD

passives.[15] However, thermal dissipation for these resistors may prove a challenge if embedding them into a substrate. Specialized passive or active cooling approaches may be needed, with the possibility of printing these features into the substrate as well.

Passives in Power modules:

Power modules are used to transform the source voltage to the load voltage levels. For example, a buck converter steps down the battery voltage to a lower voltage level acceptable by a load such as a microprocessor; or a boost converter steps up the battery voltage to a higher voltage required by the disk drives. Power modules consist of a switch, often a metal oxide semiconductor field effect transistor (MOSFET), integrated with microcontroller and driver, input and output capacitors (C), and an output inductor (L). Passive components such as inductors and capacitors are the critical energy storage components that determine the size and performance of voltage regulators. For power components, the performance metrics typically are volumetric density, efficiency (inverse of losses), and power handling. The capacitors are usually in the range of 0.1–1 μF, whereas the value for the inductors vary from 0.01–1 μH. Emerging needs dictate capacitors with densities of 1-2 μF/mm^2 for 0.1-0.2 mm thickness, to supply 1-2 A/mm^2 pulses to the load (such as a processor) with lowest inductances for MHz operation. Such densities need nanoscale dielectrics on 3D nanostructured electrode structures. Capacitors using AE methods achieve thicker dielectrics, indicating significant progress needed to get to nF or μF ranges, especially at a suitable volume. The printing of high-K dielectric materials[26] could assist in improving printed capacitor performance. Although AE techniques may not achieve the capacitor component metrics, they can aid in the insertion of IPDs or pre-fabricated capacitor devices within the package build-up to form seamless 3D connectivity with the rest of the system components. For example, nScrypt's microdispensing tools can pick-and-place passive devices while printing the interconnect layers.

Inductors should continuously have a high current of 1-2 A/mm^2, making low resistance an imperative need. Hence, superior conductivity is critical so that the inductor efficiency is not degraded from DC current (I^2R losses). Processors need at least 5-10 nH/mΩ to meet the efficiency needs. To reduce the conductor trace length, it is critical to insert magnetic cores, where AE techniques are beneficial by introducing heterogeneous materials such as conductors, dielectrics and magnetics. Particularly, AE can bring advanced low-loss nanomagnetic composites and inserted low-loss ferrites to suppress the core high-frequency hysteresis losses[1]. AJP has been shown to be especially useful when creating straight-solenoid inductors with placed ferrite cores, with reported ranges of 0.33 – 400 μH (Figure 4B).[17,21] FDM with different filament types has been used to create 0.2-2 μH for inductors (spiral air-core). Inductors have been demonstrated up to 400 μH[17] at an area of 23.6 mm^2, putting them in the range of SMD inductors, albeit at a currently larger footprint. Currently, however, placed magnetic cores are seen to be needed to yield these higher inductance values.

E. Future Directions and AM/AE Challenges

As AM and AE further develop, their utility and integration within a broader manufacturing framework will likely expand, yet these methods will still have difficulty in directly competing with conventional manufacturing methods. The expansion in use will therefore mostly be predicated upon leveraging the unique capabilities offered by AM and AE, as opposed to using them as a replacement method for fabricating traditional designs. To provide competitive solutions within SiP development, AM and AE will need to be targeted at specific limitations within heterogeneous integration, finding new ways to leverage the geometric freedom offered by AM and AE to provide a superior solution compared to conventional manufacturing approaches. For instance, in the area of thermal management, additive manufacturing will have a difficult time competing with traditionally manufactured and designed heat sinks or heat pipes. However, if high thermal conductivity materials are printed near heat generating structures while additively manufacturing the SiP, with heat pipes printed or placed to connect these thermal channels, which lead to printed fins with an organic structure optimized for air flow and higher surface area compared to tradition fins, one could imagine devices with higher levels of heat flux compared to what traditional methods allow. A current example of taking advantage of the unique capabilities of AE can be seen with the use of AJP[16] and PRJ[27] to print 3D-antennas, yielding superior RF performance compared to 2D antenna geometries. However, many refinements are needed such that these solutions meet the full suite of rigorous product demands of commercial communications devices.

One of the greatest challenges of AM and AE will be to first meet performance metrics of traditional methods before the unique geometric freedom can be employed for superior solutions. These include improvements in feature resolution, dimensional accuracy, material breadth, final material properties, print speed, time-to-part, repeatability and scalability (see Table 1). Given the disparate methods within AM and AE, each method will need development within its own technological framework, with most solutions not being directly transferable to other methods. To be

useful in creating conductive features within a printed structure, AE also needs to develop along additional vectors including trace line and space, trace conductivity, via line and space, via conductivity, number of conductive layers, space between conductive layers, contact resistance, pad-pad pitch, dielectric strength, flame retardancy, and others (see Table 1). Although there are AM/AE methods which are able to currently meet or exceed some performance metrics (for instance, micro-stereolithography[28] has resolution down to 1 μm), no current AM/AE method meets all of them (micro-stereolithography has limitations on scalability and time-to-part), and methods will need to be further developed with each metric taken into account and balanced.

However, it is unlikely that one singular AM or AE method will achieve optimal performance among all these vectors, so AM/AE systems with P&P and PP/R enablement will likely be needed. This will allow for hybrid systems with the potential for multiple AM/AE methods and multiple conventional manufacturing methods which are all combined into a larger manufacturing system. Another major challenge, therefore, is the development of these hybrid manufacturing systems with both AM/AE and traditional methods. Additionally, the development materials and strategies for optimal characteristics at the interface between printed and prefabricated structures is needed. Especially in the area of interface materials between a printed material and a conventional substrate or device, little work has been done to date looking at optimizing characteristics at the interface.

A final major challenge for AM/AE methods is the use of build materials which can match the high-performance metrics (mechanical, thermal, chemical resistance, and other properties) typical for conventional manufacturing materials (like SU-8, PI or PBO). This may require the printing of composite materials to improve the properties of typical AM materials, or the development of new materials which are both able to be directly printed and achieve the desired properties. Some AM methods have unique challenges printing composites (SLA and PBF+IJ), whereas others have less difficulty (FDM), which may lead to quicker advancement of particular approaches to yield commercially viable SiP products compared to others.

References

1. Macdonald E, Salas R, Espalin D, et al. 3D Printing for the Rapid Prototyping of Structural Electronics. IEEE Access. 2014;2:234-242. doi:10.1109/ACCESS.2014.2311810
2. Joe Lopes A, MacDonald E, Wicker RB. Integrating stereolithography and direct print technologies for 3D structural electronics fabrication. Rapid Prototyp J. 2012;18(2):129-143. doi:10.1108/13552541211212113
3. Espalin D, Muse DW, MacDonald E, Wicker RB. 3D Printing multifunctionality: structures with electronics. Int J Adv Manuf Technol. 2014;72(5-8):963-978. doi:10.1007/s00170-014-5717-7
4. Rim YS, Bae S-H, Chen H, De Marco N, Yang Y. Recent Progress in Materials and Devices toward Printable and Flexible Sensors. Adv Mater. 2016;28(22):4415-4440. doi:10.1002/adma.201505118
5. Wang X, Jiang M, Zhou Z, Gou J, Hui D. 3D printing of polymer matrix composites: A review and prospective. Compos Part B. 2017;110:442-458. doi:10.1016/j.compositesb.2016.11.034
6. A Deloitte Series on Additive Manufacturing 3D Opportunity for Electronics. https://www2.deloitte.com/content/dam/insights/us/articles/3837_3D-opportunity-for-electronics/DUP_3D-Opportunity-Electronics.pdf. Accessed August 5, 2018.
7. Muth JT, Vogt DM, Truby RL, et al. Embedded 3D Printing of Strain Sensors within Highly Stretchable Elastomers. 2014. doi:10.1002/adma.201400334
8. nScrypt. https://www.nscrypt.com/.
9. Layani M, Cooperstein I, Magdassi S. UV crosslinkable emulsions with silver nanoparticles for inkjet printing of conductive 3D structures †. doi:10.1039/c3tc30253a
10. Mannoor MS, Jiang Z, James T, et al. 3D Printed Bionic Ears. 2013. doi:10.1021/nl4007744
11. Ladd C, So J-H, Muth J, Dickey MD. Microstructures: 3D Printing of Free Standing Liquid Metal Microstructures (Adv. Mater. 36/2013). Adv Mater. 2013;25(36):4953-4953. doi:10.1002/adma.201370225
12. Guo S-Z, Qiu K, Meng F, Park SH, McAlpine MC. 3D Printed Stretchable Tactile Sensors. Adv Mater. 2017;29(27):1701218. doi:10.1002/adma.201701218
13. Hoerber J, Glasschroeder J, Pfeffer M, Schilp J, Zaeh M, Franke J. Approaches for Additive Manufacturing of 3D Electronic Applications. Procedia CIRP. 2014;17:806-811. doi:10.1016/J.PROCIR.2014.01.090
14. NanoDimension. https://www.nano-di.com/.
15. Wittkopf, Jarrid A.; Erickson, Kris; Olumbummo, Paul; Hartman, Aja; Tom, Howard; Zhao L. 3D Printed Electronics with Multi Jet Fusion. NIP Digit Fabr Conf Print Fabr. 2019;5:29-33.
16. Aerosol Jet. https://optomec.com/printed-electronics/aerosol-jet-technology/.
17. Gu Y, Park D, Bowen D, Das S, Hines DR. Direct-Write Printed, Solid-Core Solenoid Inductors with Commercially Relevant Inductances. Adv Mater Technol. 2019;4(1):1-7. doi:10.1002/admt.201800312
18. O'Donnell J, Kim M, Yoon HS. A Review on electromechanical devices fabricated by additive manufacturing. J Manuf Sci Eng Trans ASME. 2017;139(1):1-10. doi:10.1115/1.4033758

19. Lu B-H, Lan H-B, Liu H-Z. Additive manufacturing frontier: 3D printing electronics. Opto-Electronic Adv. 2018;1(1):17000401-17000410. doi:10.29026/oea.2018.170004
20. Flowers PF, Reyes C, Ye S, Kim MJ, Wiley BJ. 3D printing electronic components and circuits with conductive thermoplastic filament. Addit Manuf. 2017;18(2017):156-163. doi:10.1016/j.addma.2017.10.002
21. Gu Y, Park D, Gonya S, Jendrisak J, Das S, Hines DR. Direct-write printed broadband inductors. Addit Manuf. 2019;30(July):100843. doi:10.1016/j.addma.2019.100843
22. Ota H, Emaminejad S, Gao Y, et al. Application of 3D Printing for Smart Objects with Embedded Electronic Sensors and Systems. Adv Mater Technol. 2016;1(1):1-8. doi:10.1002/admt.201600013
23. Wu S-Y, Yang C, Hsu W, Lin L. 3D-printed microelectronics for integrated circuitry and passive wireless sensors. Microsystems Nanoeng. 2015;1(1):15013. doi:10.1038/micronano.2015.13
24. Secor EB. Principles of aerosol jet printing. Flex Print Electron. 2018;3(3):35002. doi:10.1088/2058-8585/aace28
25. McKerricher G, Perez JG, Shamim A. Fully inkjet printed RF inductors and capacitors using polymer dielectric and silver conductive ink with through vias. IEEE Trans Electron Devices. 2015;62(3):1002-1009. doi:10.1109/TED.2015.2396004
26. Kang BJ, Lee CK, Oh JH. All-inkjet-printed electrical components and circuit fabrication on a plastic substrate. Microelectron Eng. 2012;97(4023):251-254. doi:10.1016/j.mee.2012.03.032
27. He X, Tehrani BK, Bahr R, Su W, Tentzeris MM. Additively Manufactured mm-Wave Multichip Modules With Fully Printed “Smart” Encapsulation Structures. IEEE Trans Microw Theory Tech. 2020;68(7):2716-2724. doi:10.1109/TMTT.2019.2956934
28. Behroodi E, Latifi H, Najafi F. A compact LED-based projection microstereolithography for producing 3D microstructures. Sci Rep. 2019;9(1):19692. doi:10.1038/s41598-019-56044-3
29 P. M. Raj, H. Sharma, S. Sitaraman, D. Mishra, and R. Tummala, "System scaling with nanostructured power and RF components," Proceedings of the IEEE, vol. 105, pp. 2330-2346, 2017.

Section 11: Electromigration

Perspective and General Trends

Electromigration (EM) is a major reliability concern for interconnect structures due to aggressive dimensional scaling and ever-increasing current density. Heterogeneous integration (HI) of advanced packaging technologies brings together novel interconnect structural components such as micro-bumps/pillars, hybrid bonding, RDL and TSV, all of which are subjected to EM-induced failure [1,2]. The interconnects in HIR can have distinct EM characteristics due to the parallel network configuration, where the standard weakest-link approximation used to evaluate EM lifetime would not be applicable and require new EM criteria for network systems. EM in this section is closely related to Section 12 Reliability and further linked as reliability prerequisites to Section 2 System Requirements, Section 3 Thermal Requirements and Section 4 Mechanical Requirements.

Some general trends have emerged from EM reliability studies:

- Apart from niche applications, the power/current density requirements for FOWLP and 2.5D/3D packaging are similar to flip chip with Cu Pillar products, but with reduced power due to reductions in package dimensions and interconnect lengths.
- Joule heating has become an important issue for EM and in consequence, its dissipation is critical.
- For power grids used in packaging applications, EM reliability has emerged to become an important issue due to the increase in current density requirements. The EM characteristics for power grids are different from individual conductors due to the parallel network configuration but is specific to each application.

This section is organized into 5 topical areas:

- Cu redistribution layer (RDL)
- TSV for high density integration
- Hybrid bonding (HB) structures
- Solders, Cu pillars and micro-bumps
- Power grid systems

1. EM in Redistribution Layer (RDL)

Wafer-level chip-scale packages (WLCSPs) and Fan-Out packages are subjected to the same drive for miniaturization as all electronic packages. The I/O count is increasing while ball pitch is shrinking at the expense of trace pitch, and in turn, current densities are increasing. This leads to current crowding and Joule heating near the solder joints and under-bump metallurgy (UBM) structures with significant resistance increases. These phenomena

are responsible for structural damage of redistribution line (RDL)/UBM and UBM/solder interconnects due to ionic diffusion or electromigration.

Al and Cu RDLs have been examined with different process/integration schemes using NIST [3, 4] or equivalent via-line structures [5, 6]. In addition to study of EM-related failure of the RDL, the impact on the surrounding RDL and passivation materials has been studied [3,4,6]. Kao [5] investigated sputtered Ti/Al/Ti and sputtered and electroplated Ti/Cu/Cu RDLs and found activation energies 0.72-0.96 eV and 1.31-1.41 eV with current exponent (n) between 2 and 2.5 respectively. The result indicated that damage is due to grain boundary diffusion. Kudo [6] studied Enhanced Cu redistribution layer (ENCORE) and compared it to conventional Cu RDL. Its lifetime was found to be higher, which was attributed to the interfacial modification of the Cu traces. In the ENCORE structure, the traces were completely covered with two types of inorganic dielectric where the first dielectric constrains Cu migration to prevent Cu oxidation to improve EM reliability. In a conventional RDL, the weak adhesion between the Cu trace and a single organic dielectric passivation greatly degraded EM reliability.

Results reported by Moreau [4] confirm these observations that the main diffusivity paths are grain boundaries and passivation polymer at the Cu/SiN interface (Figure 1). SiN cannot directly cover all the Cu RDL, thus risking Cu RDL corrosion. A separate study on Wafer Level Integrated Fan-Out Technology (InFO) by Tseng et al. [15] reported an activation energy of 0.9 eV and a current exponent of 2 for Cu RDL, consistent with results by Kao [5]. A recent EM test was conducted on 2µm/2µm L/S Cu RDL embedded in polyimide for Fan-Out Package at 8.8×10^5 A/cm^2 at 180°C in accordance with the JEDEC standard. The EM test was found to accelerate Cu oxidation and void formation at the oxide/RDL interface, indicating a thermal induced failure mechanism [16,17]. EM tests were also conducted on 2-µm-wide RDLs formed with nanotwinned Cu and capped with polyimide. The nt-Cu RDLs showed an improved EM lifetime than regular Cu RDL lines but with serious oxidation in the Cu lines [18].

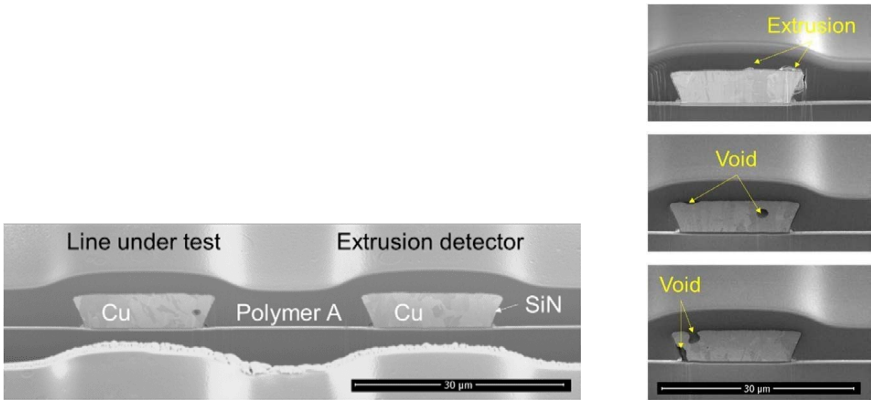


Figure 1: SEM micrograph of a Cu RDL line passivated by a bilayer material (SiN/polymer) after EM test (200 °C, 500 mA). a) Overview, b) types of electromigration-induced defects (voids, extrusion) and localizations (bulk, Cu/SiN interface) [4].

2. EM in TSV for High Density Integration

EM test structures used in these studies are standard NIST-type structures [7, 8]. For TSVs fabricated by the Single Damascene (SD) process, copper depletions usually occur in lines connected to TSVs, depending on current flow. In Dual Damascene Cu lines, voids always grew in the M1 metal lines directly below the TSV, depending on the current flow. EM damage occurred mainly at the interface between the high-density TSV and the BEOL interconnect [8], as shown in Figure 2.

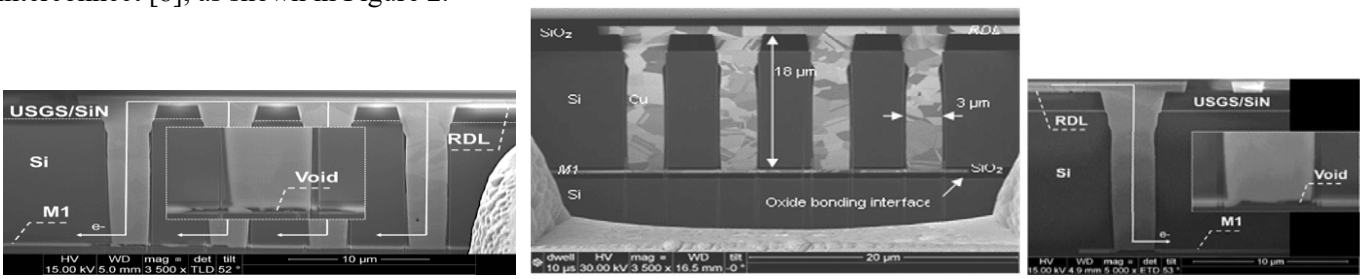


Figure 2: DD TSV and EM damage at the interface between TSV and BEOL interconnect [8].

In these studies, the Black EM parameters extracted from EM tests are in good agreement with typical values obtained for copper interconnects. Postmortem failure analysis has revealed voids in SD lines ended at both sides of TSV, RDL or BEoL interconnects, depending on the direction of electron flow. For DD interconnects, voids are in lines ended at the BEoL side of the TSV. Generally, voids in the TSV bulk have not been detected (low current density). Complicated void-void interaction has been observed in the electrically connected TSV array. 3D FEA simulation was applied to guide the failure analysis [9], as illustrated in Figure 3.

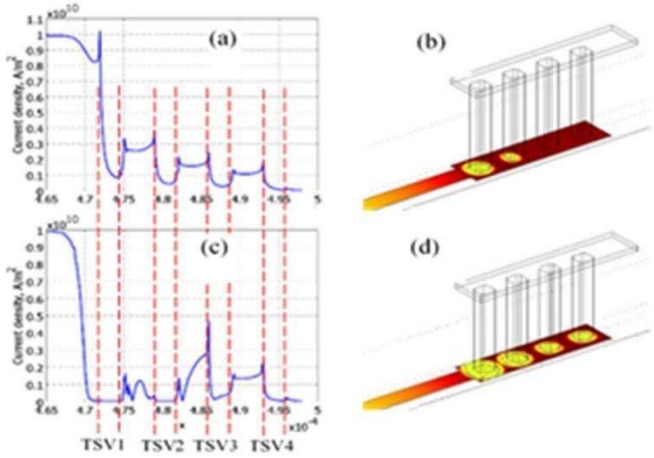


Figure 3: FEA results for up-stream configuration showing (a) initial current distribution, (b) first TSV failure, (c) current distribution after failure of two TSVs, and (d) the circuit failure. Color maps in (b) and (d) show the hydrostatic stress level: yellow is zero, darker red shows higher tensile [9].

3. EM and Modeling for Hybrid Bonding Structures

Structures formed with Hybrid Bonding (HB) technology are limited by low TSV scalability. Electromigration tests were performed to investigate the failure mechanism using the NIST-type test structures together with multi-link daisy chains for analysis of yield related issues. In NIST-type test structures, EM-induced voids are usually found in the SD BEoL lines at the top or bottom wafers, depending on the electron flow direction, Figure 4. Intrinsic bonding voids formed in processing did not move under the electric current.

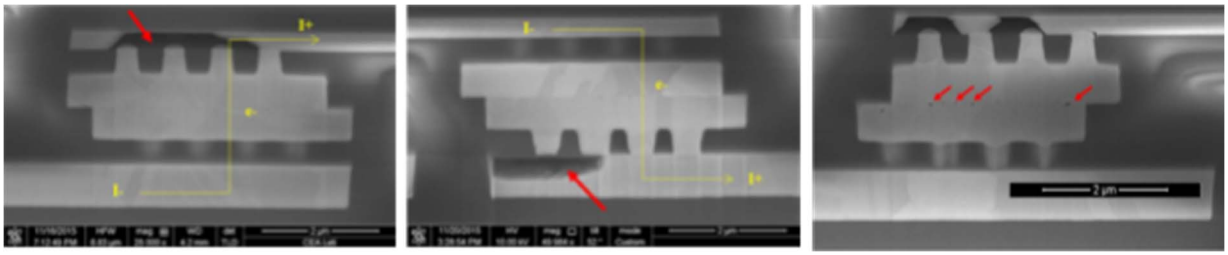


Figure 4: EM-induced voiding in the NIST-like structures [4].

Recent tests performed on Cu/SiCN to Cu/SiCN hybrid bonds also found EM damage dependent on the electron flow direction. For electrons flowing up, failure occurred in the top current feed line, and a lower bound j_{max} at 100 °C of 8.9 mA/cm² is estimated. For electrons flowing down, voids and gap occurred at the bonding interface, then moved up to the top of the top pad along the TaNTa/Cu interface, which suggests damage formation due to EM induced void agglomeration in hybrid bonds [19].

Finite element modeling was used to investigate the EM reliability for a 100-link daisy chain. No EM-induced void was found along the daisy chain, due to the short line length; instead, voids were found to localize only at the cathode side in the feed line at the top layer of the BEoL. Present software- and physics-based models are able to match the experimental results [9].

4. EM in Solders, Cu Pillars and Micro-bumps

EM reliability has been investigated recently over a wide spectrum of far-backend interconnects, including micro-bump, copper pillar, thermocompression flip chip bump, lead-free bump and solder ball, to rank their performance and identify key parameters for reliability [10]. For this class of solder structure, the EM lifetime depends on the amount of Cu consumption due to CuSn intermetallic (IMC) formation, so their EM performance can be classified according to the solder-to-Cu ratio:

- Solder balls, Pb-free bumps and Cu pillars on narrow traces – Solder/Cu > 3
- Cu pillars, thermal compression C4 and micro-bumps – Solder/Cu <3

Low performance was found for solder/Cu ratio >3 where Cu is mostly consumed by IMC formation. EM failures are caused by void formation at IMC interfaces or in the cathode Cu traces. High performance was found for solder/Cu ratio <3 where some Cu remains intact after IMC formation with almost no void formation. A steady-state or near steady-state condition can be reached where no EM voids appear and the resistance is stable. Micro-bumps have the lowest solder/Cu ratio, so more Cu remains after IMC formation and EM lifetime can become immortal [10,11].

The current carrying capability can be classified into two regions depending on the solder/Cu ratio, with high performance for solder/Cu < 3 and low performance for solder/Cu > 3. The difference is more than 10×, as shown in Figure 5.

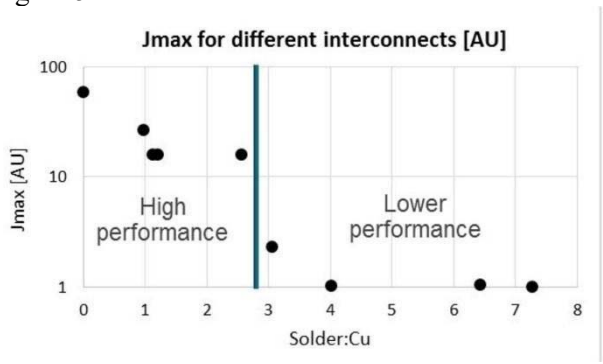
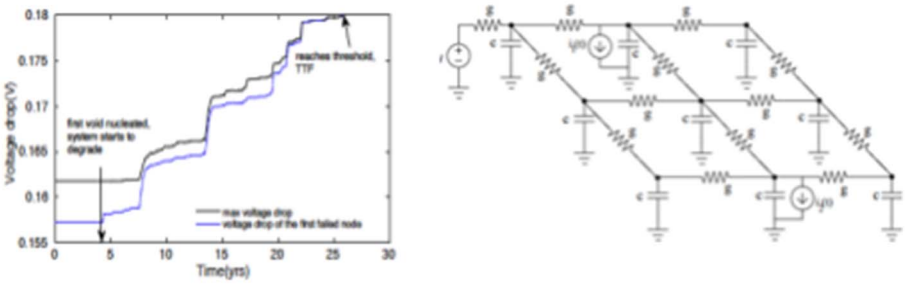


Figure 5. Classification of Jmax behavior for different solder-to-Cu volume ratios for solder interconnect structures [10].

5. EM and Wiring Design for Power Grids

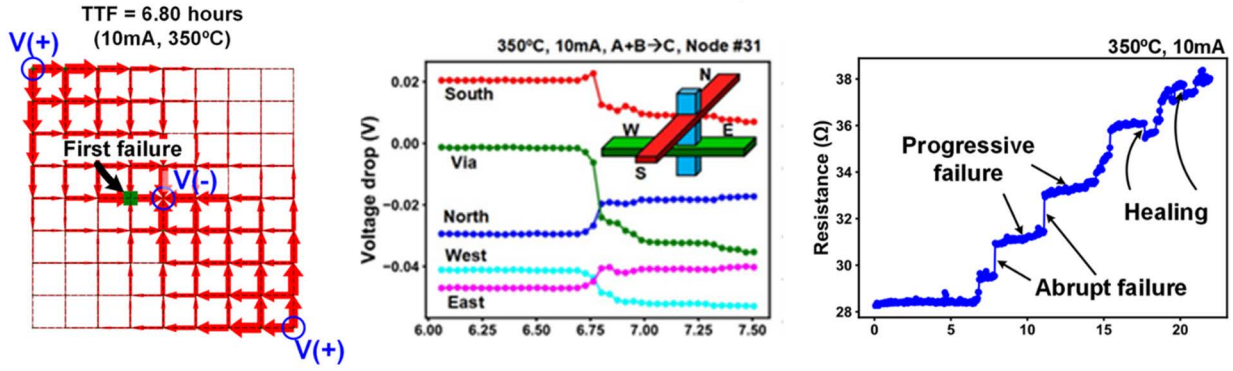
The parallel network configuration for power grids has EM characteristics distinctly different from individual lines, since the standard weakest-link criterion would be too pessimistic for projecting the EM lifetime and current density capability [12]. In addition, the Black equation currently used to calculate the MTTF of individual links ignores the material flow between branches. In power grids, many branches in the mesh structure are connected on the same level with no diffusion barriers in between. This forms an interconnect network where atomic flux can flow freely between the branches, invalidating the Black equation for projecting EM lifetime. This would also make the immortal prediction for individual short branches based on the Blech effect too optimistic and thus misleading for wiring design. A new analysis using a mesh network to account for the grid redundancy has been developed where EM failure occurs only when the grid interconnect cannot deliver the voltage required for the circuit to function properly. This yields a timing error or a reduction of the noise margin, corresponding to a performance loss and a parametric failure, which is a more realistic and practical failure criterion for the power grid systems, as shown in Figure 6a [12,13]. Such analyses have been performed on many industrial-grade power grids to show that the current assessment is too pessimistic, by designing the grid to survive 40 years or more while it has to survive only 10 years. This can be a big problem for power grid design, resulting in overuse of metal area and leaving little room for signal routing with increasing design complexity and design time. In contrast, the newly developed physics-based model can provide a more realistic EM assessment for power grids with user-specified current sources and voltages (Figure 6b). Such an approach can effectively relax the current density design rules with significant improvements in power, time-to-market and design cost.



(a) (b)

Figure 6. (a) Voltage drop of the first failed node and the maximum voltage drop in a power grid as a function of time [13], (b) Power grid schematics with user-provided current sources and voltages [14].

The proposed EM assessment approach was proved recently by experiments [20]. A 65 nm test chip, which was designed to study EM events in integrated circuit power grids, was using M3 and M4 metal layers stressed under constant current and constant voltage modes. Using the voltage tapping technique, the voltage drop map across the entire power grid was directly observed for the first time, Figure 7a, b. Subtle changes on the monitored voltage map uncovered mechanical stress-dependent failure locations as well as self-healing due to redundant current paths, Figure 7c. The EM failure rate and order of failure locations were also analyzed.



(a) (b) (c)

Figure 7. (a) Measured voltage drops and first EM failure location for a fresh power grid. Arrow indicates the magnitude and polarity of voltage drop between adjacent nodes. (b) Example of nearby voltage drop traces after an EM event. The increased voltage across the via suggests a change in the via resistance. (c) Resistance traces exhibiting EM abrupt failure, progressive failure and temporary healing behavior [18].

6. Difficult Challenges and Potential Solutions

1. High heat dissipation in 3D IC chips due to increasing current densities and power requirements associated with the use of very thin dies will cause local hot spots and non-uniform EM-induced failure, making it difficult to predict real MTTF.
   Potential solutions:
   - A priori measurement-based extraction of MTTF as a function of temperature for critical heterogeneous integration components;
   - Precise measurements of intra-stack temperature distribution with designed-in temperature sensors;
   - Multilevel sub-modeling to evaluate warpage, stress concentration and interfacial fracture.
2. It is difficult to project EM reliability statistics with increasing system complexity.
   Potential solutions:
   - Make separate measurements of MTTF of all EM-affected components in the integrated stack, then combine to assess system reliability;
   - Accurate projection of MTTF and statistics, taking into account the redundancy in the design of power grids, standard cell connections to the grid, TSV and bump arrays, and other elements of the power delivery infrastructure;
   - Nodal voltage evolution measured with on-grid voltage sensors to validate and calibrate the novel EM assessment methodology.
3. Difficult to predict system-level EM reliability with distinct failure rates of system components. Challenge in combining with optimization of system-level power distribution through power grids while maintaining system performance and EM reliability.
   Potential solutions:
   - The system-level EM lifetime is subject to power and performance constraints. Employ dynamic voltage and frequency scaling at the system level to achieve an optimum trade-off between EM lifetime and energy/performance. Develop learning-based energy optimization to manage and optimize energy and to meet reliability, power budget and performance requirements.

References

1. C. H. Yu, et al., “High Performance, High Density RDL for Advanced Packaging”, IEEE ECTC 2018.
2. J. Lau, et al., “Reliability of Fan-Out Wafer-Level Packaging with Large Chips and Multiple Re-Distributed Layers”, IEEE ECTC 2018.

3. L. Dumas, et al., “Electromigration and adhesion improvements of thick Cu/BCB architecture in BiCMOS RF technology”. Proc. 5th WSEAS Int. Conf. on Micro. nanoelectronics, optoelectronics (MINO'06), pp. 139-142.
4. S. Moreau, et al., “Passivation Materials for a Reliable Fine Pitch RDL”, IEEE ECTC, 2018.
5. C.-L. Kao, et al., “Investigation of electromigration reliability of redistribution lines in wafer-level chip-scale packages,” Microelectronics Reliability, 54, 11, 2014, pp. 2471-2478.
6. H. Kudo, et al., "Demonstration of High Electrical Reliability of Sub-2 Micron Cu Traces Covered with Inorganic Dielectrics for Advanced Packaging Technologies," IEEE ECTC, 2017, pp. 1849-1854.
7. T. Frank, et al., "Electromigration behavior of 3D-IC TSV interconnects," IEEE ECTC, 2012
8. S. Moreau, et al., “Reliability of Dual Damascene TSV for high density integration: the electromigration issue”, IRPS, 2013.
9. J.-H. Choy, et al. “Finite-difference methodology for full-chip electromigration analysis applied to 3D IC test structure: Simulation vs. experiment”, 2017 Int. Conf. on Simulation of Semiconductor Proc. Dev. (SISPAD), 41-44, 2017.
10. C. Hau-Riege, et al., “Key metrics for the electromigration performance for solder and copper-based package interconnects”, IRPS, 2016.
11. R. Labie, et al., “Outperformance of Cu Pillar Flip Chip Bumps in Electromigration Testing”, IEEE ECTC, 2011.
12. V. Sukharev et al., “Electromigration check: where the design and reliability methodology meet” IEEE TDMR, v. 18, pp. 498-507, 2018.
13. X. Huang, et al. “Physics-based electromigration models and full-chip assessment for power grid networks”, IEEE Trans. Comput.-Aided Design Inegr. Circuits Syst., vol. 35, no. 11, pp. 1848-1861, Nov. 2016.
14. S. Chatterjee, et al. “Power grid electromigration checking using physics-based models,” IEEE Trans. Comput.-Aided Design Inegr. Circuits Syst., 2018. DOI: 10.1109/TCAD.2017.2666723.
15. Chien-fu Tseng, Chung-Shi Liu, Chi-Hsi Wu & Douglas Yu “InFO (Wafer Level Integrated Fan Out) Technology, IEEE EPS 66th ECTC, 2016.
16. Chien-Lung Liang, et al., “Electromigration Failure Study of a Fine-pitch 2?m/2?m L/S Cu Redistribution Line Embedded in Polyimide for Advanced High-density Fan-out Packaging” IEEE EPS 70th ECTC, 2020, p.361.
17. C. Liang et al., "Electromigration Reliability of Advanced High-density Fan-out Packaging with Fine-pitch 2lm/2lm L/S Cu Redistribution Line," in IEEE Transactions on Components, Packaging and Manufacturing Technology, doi: 10.1109/TCPMT.2020.2997824.
18. I. Tseng, et al. "Electromigration in 2 ?m Redistribution Lines and Cu-Cu Bonds with Highly <111>- oriented Nanotwinned Cu," 2020 IEEE ECTC, Orlando, FL, USA, 2020, 479-484, doi: 10.1109/ECTC32862.2020.00083.
19. Joke De Messemaeker et al. “Electromigration Behavior of Cu/SiCN to Cu/SiCN Hybrid Bonds for 3D ntegrated Circuits”, 2018 International Conference on Solid State Devices and Materials, Tokyo, 2018, pp449-450.
20. C. Zhou, et al., “Electromigration Effects in Power Grids Characterized from a 65 nm Test Chip”, IEEE TDMR, v. 20, pp. 74-83, 2020.

Section 12 Reliability Aspects of Heterogeneous Integration

Initial Team: A. Dasgupta (Univ. Maryland), R. Rao (Inphi), S. Sahasrabudhe (Intel), W. D. van Driel (Signify/TU Delft)

1. Introduction, Overview and Scope

To understand the reliability challenges in heterogeneous integration (HI), it is instructive to reiterate the multidimensionality of the HI revolution. One important dimension of HI is the diversity of length scales, since HI is relying not only on wafer-scale ‘more Moore’ advances (including the reduction of semiconductor features sizes to 5 nm range or below, as well as wafer-scale SoC integration), but also on chip-to-chip and package-to-package ‘More-than-Moore’ integration, in both 2.5D and 3D SiP/SoP configurations, ranging up to length scales of several centimeters in many cases. The other critical dimension of HI is the diversity of multi-physics technologies that have to integrate together for different applications. Examples include microelectronic devices, photonic and LED devices, and MEMS/NEMS devices for bio-chemical, acoustic, fluidic, electro-chemical and electro-mechanical applications. Needless to say, such HI requires tremendous diversity of design methods, materials technologies and manufacturing processes. A convenient representation is available in Yole’s Table presented in Figure 1. A more updated version of this table will show that feature sizes are now commercially approaching 5 nm and research labs are exploring technologies below 5 nm.

An example of such a system is presented in Figure 2.

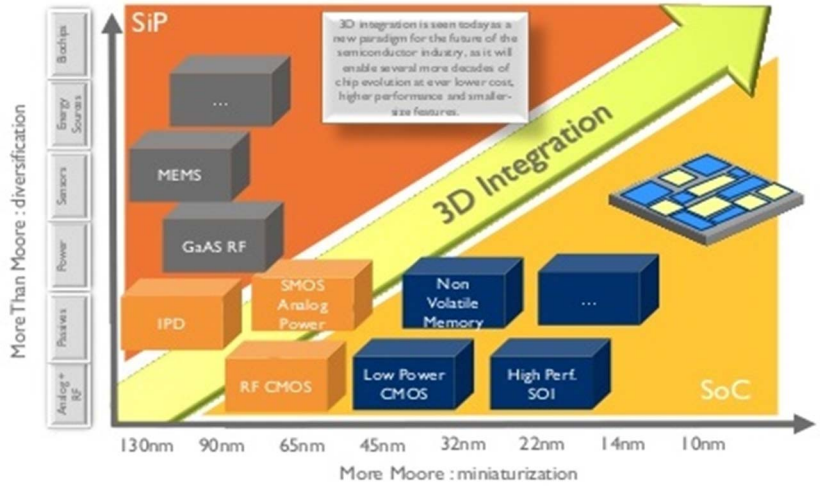


Figure 1. Multi-scale and Multi-Physics perspective of Heterogeneous Integration Systems (Source: Yole, 2015)

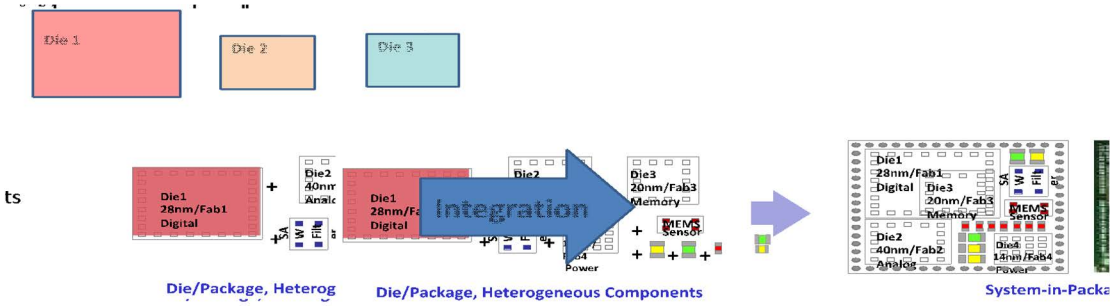


Figure 2. Heterogeneous Integration and System in Package (SiP) (Source: ASE)

From a reliability perspective, this multidimensional diversity presents unprecedented challenges because of:

- multiscale, multiphysics material behavior (ranging from nms to cms);
- novel multiscale manufacturing methods and associated effects on quality and material behavior;
- multiphysics operational and environmental stresses;
- multifunctional performance targets.

All of this diversity will lead to ***multiphysics aging and degradation mechanisms*** in the novel materials and interfaces that will become increasingly prevalent in future HI systems. System resilience and system availability will require novel approaches.

To develop and sustain highly reliable HI systems, product developers will need the knowledge-base, resources and tools for:

- strategies to build-in robust design margins for dominant degradation/failure mechanisms;
- resilient system design strategies;
- process control strategies for managing HI manufacturing processes;
- effective product qualification strategies using a combination of virtual and physical qualification methods;
- real-time in-situ self-cognizant capabilities for systems to monitor and manage their own health throughout the life-cycle for effective sustainment.

The scientific and engineering eco-system (knowledge-base, methodologies, tools and resources) needed to successfully achieve these goals of achieving high reliability and availability in future HI systems include:

- quantitative understanding of multiphysics performance and reliability goals;
- efficient methods for multiphysics stress assessment;
- reliability physics predictive models and tools for multiphysics degradation mechanisms;
- characterization and modeling of multiscale multiphysics materials behavior.

- accelerated stress-testing methods that can be used for virtual and physical qualification of products under the action of multiphysics stress conditions;
- advanced data analytics (e.g. deep machine learning and artificial intelligence methods) that can be built into system firmware to cope with system complexity and dynamic changes throughout the life-cycle;
- methods to handle and manage uncertainty and risk;
- methods to manage complex diverse non-vertical supply chains;
- methods to manage mismatches in technology obsolescence cycles of different subsystems used in complex long-life systems;
- methods to monitor and manage the cost and affordability of all the reliability actions listed above.

It is also important to understand that the knowledge-base and eco-system discussed above is not a static goal. Instead, the goalpost will continue to move as technology evolves and grows. Managing such a dynamic eco-system will not happen by serendipity. Instead, this will require careful pan-industry strategizing and proactive investment strategies in the technology infrastructure. Standardization is a tried-and-true industry tool to manage diversity in the product landscape, but in the diverse world of HI, we must be particularly cautious about developing overly prescriptive approaches to one-size-fits-all industry standards for product development and qualification, as excessive standardization could stifle disruptive innovations. Radically new concepts are needed in standards development strategies, to nurture and encourage application-specific innovations in different segments of the HI ecosystem.

In summary, the HI roadmap for reliability assurance must recognize that the competitive pressures of an ever-expanding global market require companies to be more aggressive in their development, to increase the mix of off-the-shelf technologies and customized technologies, and to develop strategic partnerships. The emergence of new technologies, components, and processes has further taxed the infrastructure of electronic system developers. An increasingly wide array of technologies and resources have to be made available for the development of cost-effective and competitive timely products and systems so that technology risks of HI can be assessed against the backdrop of market risks and companies can develop holistic cradle-to-cradle approaches for managing the life-cycle of HI product families.

It is clear that in such a diverse landscape, a single roadmap will not address all the problems for all the stakeholders. This roadmap will therefore limit its scope for now to the next decade of the microelectronics industry. In the future, this roadmap could be expanded to the photonics and MEMS/NEMS industries.

2. HI System Reliability Principles

Heterogeneous Integration (HI) will launch unprecedented changes, not only in multiscale system complexity, functionality and density (using either single chips or multiple chiplets), but also in multi-physics diversity of technologies (involving combinations of digital, analog, power, RF, sensors, MEMS, photonics, chem/bio-electronics and other devices) within a unified system-in-package (SiP) configuration. Such extreme complexity will inevitably result in increased densities of intrinsic material defects, manufacturing flaws and stochastic variabilities. HI systems of the future will therefore have to combine increasingly resilient and fault-tolerant designs with self-monitoring, self-cognizance and varying degrees of adaptive reconfiguration and self-healing capabilities, to provide high reliability and availability. These systems will have cradle-to-grave reliability management using 'digital twins' which will be based on hybrid methods that will combine bottom-up reliability physics (RP) approaches[1-7] with top-down artificial intelligence (AI) methods. This section lays out the scope, challenges, disruptive opportunities and potential approaches for achieving an integrated approach across the entire product stack-up (chip to system) hierarchy, in HI technologies that are likely to emerge over the next 0-5, 5-10 and 10-15 years.

Reliability describes the ability of products to meet intended performance targets throughout their useful life (typically quantified with probabilistic metrics such as failure distributions, failure rates, MTTF, etc). Managing hardware reliability (the focus of this section) starts with adequate definition of: (i) customer's reliability expectations; (ii) product micro/macro life-cycle environmental and operational duty cycles; and (iii) impact of the life-cycle stresses on wearout and overstress damage mechanisms, based on product technology characteristics. As illustrated in Figure 3, reliability risk is often visualized as a stress-strength interference, where unreliability comes from the probability that the applied 'stress' will exceed the inherent 'strength' of the product. Their interactions can be balanced, as a function of product design, manufacturing variabilities and service expectations, to ensure that the resulting reliability margins are acceptable. In traditional approaches, these interactions were handled at different levels of the supply chain; however, in HI systems, the expertise spectrum must be integrated within a single reliability team with knowledge and skills across all the levels of integration.

The process of quantifying and managing the time-dependent 'stress' and 'strength' interference requires science-based multi-physics, multiscale co-design approaches that leverage: (i) the rich disciplines of multi-physics simulations to identify the electrical, thermal, mechanical and chemical 'stress' distributions at potential degradation sites; and (ii) fundamental RP methods to quantify the corresponding 'strength' distributions at these same sites. AI methods in the era of big-data will provide unprecedented advances in both steps (i) and (ii), by using sophisticated machine learning methods that exploit data collection, data analytics and deep learning technologies. RP will provide a 'bottom-up' approach to enable robust design margins based on assessment of dominant degradation/failure mechanisms at critical sites, while AI will provide a complementary 'top-down' perspective of system-level risk, based on the unprecedented level of real-time field reliability data that will become available via IoT (Internet of Things) infrastructure. Together, they will enable the development of 'digital twins' that are expected to become one of the central backbones of future reliability assurance methods. Figure 4 shows the traditional top-down view of system-level reliability, shown as a 'bathtub' curve (plotted in terms of scaled probability distribution functions (pdfs) instead of hazard rates), with three classical phases of infant mortality: decreasing failure rate (with Weibull scale parameter $\beta<1$), mid-life 'random' mortality (constant failure rate with $\beta=1$) and end-of-life wearout mortality (increasing failure rate with $\beta>1$). The green sub-distributions emphasize the corresponding 'bottom-up' RP view that this system-level failure information actually results from many degradation mechanisms that compete at multiple critical failure sites. End-of-life failures depend on the intrinsic robustness of the design. Premature failures (during the infancy and mid-life portions of the bathtub curve) depend on the distribution of weak sub-populations due to manufacturing and material variabilities/defects. In complex, multi-physics, multi-scale HI systems, developers will be able to leverage both RP (bottom-up) and AI (top-down) approaches via the use of 'digital twins' to estimate these failure rates, ensure system robustness and resilience, reduce time to market and minimize cost of ownership.

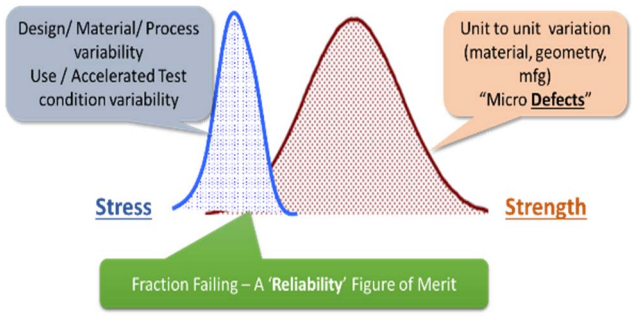


Figure 3: 'Stress' vs. 'Strength' interference

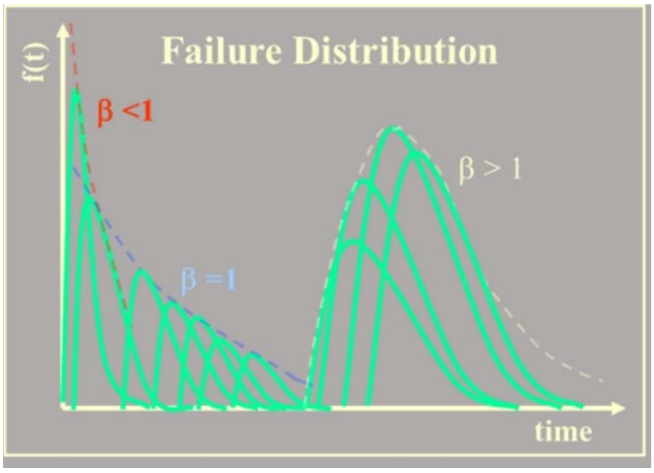


Figure 4: Bathtub Curve (replotted as scaled pdfs), showing multitude of underlying competing failure modes

Figure 5 provides a sample listing of the dominant multi-physics degradation mechanisms in the bottom-up RP perspective of microelectronics hardware systems. 'Overstress' mechanisms are triggered under the action of sudden catastrophic stress events while 'wearout' mechanisms cause gradual damage accumulation throughout the life cycle because of routine operational and environmental stress exposures.

Each listed mechanism represents a rich body of expert knowledge,

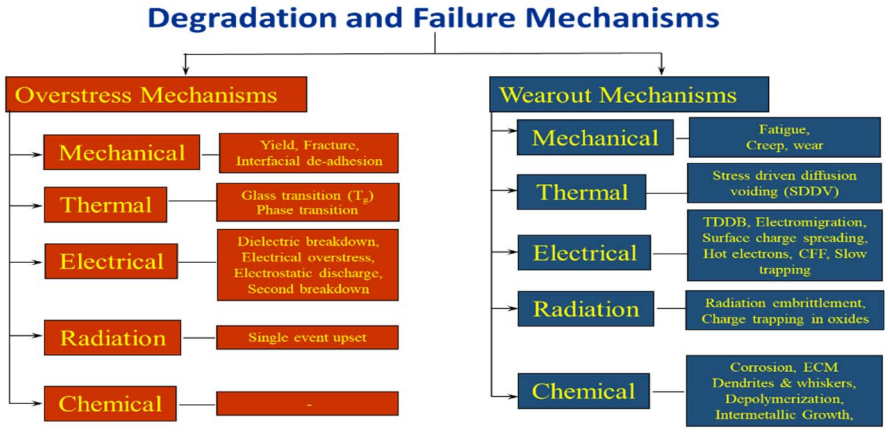


Figure 5. Examples of dominant multi-physics degradation/failure mechanisms in electronic systems, under overstress and wearout stress exposures (TDDB = Time-dependent dielectric breakdown; CFF = conductive filament formation; ECM = electrochemical migration)

quantitative models for assessing design margins and acceleration factors, model constants for current materials, and test methods for quantifying the model constants for new materials. These models must be integrated seamlessly within multi-physics co-design simulation tools so that reliability assessment can truly become a concurrent consideration (along with functional considerations) in a fully integrated ecosystem for product development and sustainment.

In contrast, the top-down concept of data-driven approaches will use AI algorithmic approaches (based on data analytics and machine learning) for product development and prognostic health management (PHM) of complex systems, to ensure high availability. This requires: (i) collection of system data (performance data and environmental stress data); (ii) smoothing and de-noising of the data using filtering methods; (iii) anomaly detection, using supervised and unsupervised machine learning algorithms; (iv) pattern identification with diagnostic algorithms to identify the root-cause source of the anomaly; (v) pattern extrapolation with prognostic algorithms to assess the remaining useful life (RUL); and (vi) actionable responses to RUL estimates (e.g. design support decisions, self-healing actions and feedback for improving the data acquisition-analysis cycle). Together, the RP and AI methods will enable the development of 'digital twins' that will facilitate life-cycle reliability management via fusion PHM, as schematically illustrated in Figure 6.

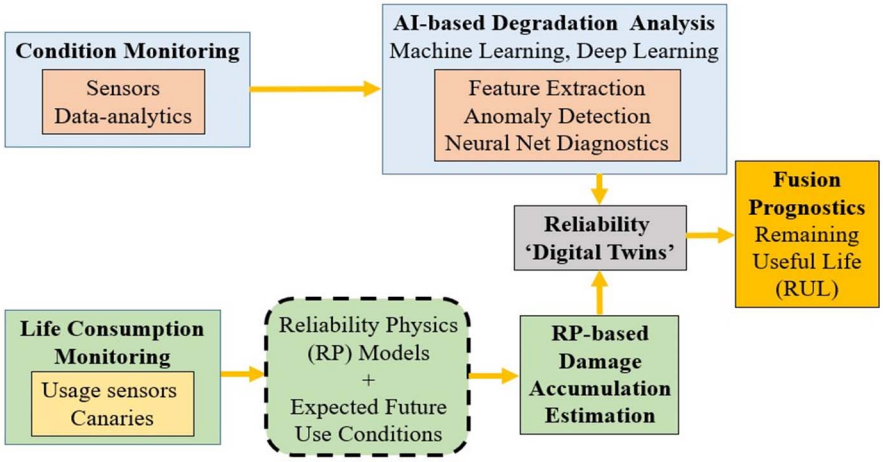


Figure 6. Flowchart for fusion PHM facilitated by use of 'digital twins,' using combination of top-down data-driven machine learning and AI algorithms, as well as bottom-up RP models.

Typical reliability tasks for quantifying and managing the stress-strength distributions and their interference are grouped, for convenience, under seven headers:

i. Identification of customers' **reliability targets** for different market segments and different technology segments
ii. Development of life-cycle user models that include expected environmental and operational **life-cycle conditions** and understanding of system configurations
iii. **Design for reliability (DfR)** tasks using hybrid RP and AI methods, materials-centric approaches, co-design simulation methods, resilient, fault-tolerant design and development of digital twins
iv. **Manufacturing for reliability (MfR)** using knowledge of the effect of processing conditions and variability on material behavior; understanding of process quality, defects and yields; use of appropriate process metrology; AI-based process control; stress screening approaches facilitated by the use of 'digital twins'
v. **Qualification for reliability (QfR)** including knowledge-based testing (KBT), accelerated stress testing for engineering verification testing (EVT), design verification testing (DVT) and process verification testing (PVT), aided by 'digital twins'
vi. **Sustaining for reliability (SfR),** based on personalized in-situ fusion PHM using 'digital twins' to enable condition-based maintenance (CBM) and dynamic adaptive healing/reconfiguration
vii. Integration and managing of reliability best practices across the **supply chain**.

3. Reliability Activities within the Context of the Overall HIR Effort

Managing reliability of complex HI systems will require synergistic efforts throughout the product development, deployment and sustainment process, requiring close interactions with all the HIR TWGs. The overall scope, tasks

and cross-TWG interactions for managing hardware reliability are schematically mapped into a 2D matrix, as shown in Table 1. The horizontal axis shows the seven activity groups discussed above, where new difficulties and challenges are expected and new solutions are needed. The vertical axis shows the three segments of stakeholders involved in the HI roadmap (Technology, HI and Product Market Segments). In the interest of brevity, these 3 segments have not been broken out into further granularity in this document. Detailed information can be found elsewhere in the literature8. The various cells of this 2D matrix are color-coded to show the interactions and cross-fertilization of the Reliability road-map, not only with other Topic Area Teams within the Single-Chip/Multi-Chip Packaging TWG (labeled in red); but also with other TWGs in other segments of the HI Roadmap (HIR) Team (labeled in black). The reliability team will be selected to contain liaison members to each of the TWGs and committees listed in Table 1.

Table 1: Hardware reliability topic matrix

<table>
<tr><th></th><th colspan="2">Reliability Targets
Life Cycle Conditions</th><th>Design for Reliability</th><th>Manufacturing for Reliability</th><th>Qualificaton for Reliability</th><th>Suustaining for Reliability</th><th>Supply Chain</th></tr>
<tr><td>SiP Technologies</td><td colspan="2">TWGs: SCM/MCM , Photonics, MEMS, Power, RF/Analog</td><td rowspan="3">TWGs: Electromigration; Materials; Co-Design and Simulation Topic Teams: Electrical System Performance; Thermal Management; Mechanical Structure and Mechanics</td><td>TWGs: SCM/MCM; Photonics; MEMS/Sensors; Power; RF/Analog; Test</td><td rowspan="3">Test TWG</td><td rowspan="3">Security TWG</td><td rowspan="3">Supply Chain TWG</td></tr>
<tr><td>Package Integration</td><td colspan="2">TWGs: WLP, 2.5D/3D, Interconnects, SIP Topic Teams: WST; Substrate</td><td>TWGs: WLP; 2.5D/3D; Interconnects; SIP, Test; Topic Teams: WST; Substrate; Board Assembly</td></tr>
<tr><td>Applications</td><td colspan="2">TWGs: Mobile, IoT, MHW, Automotive, HPC, Aerospace</td><td>TWGs: Mobile; IoT; MHW; Automotive; HPC; Aerospace; Test</td></tr>
</table>

In the initial (2019) version of this HIR Document, the scope of the Reliability Topic was limited to the Single-Chip/Multi-Chip Package Integration Segment (Second row of Table 1). In future versions, the Reliability Topic will be expanded to include additional subject matter relevant to the remaining two Segments (i.e. the Technology Segment, and the Product Market Segment). The current section will serve as a template and can be leveraged when extending this section to include those other segments. The relevant topic areas in single-chip and multi-chip package integration are: Wafer level packaging (WLP) using both fan-out and fan-in (FO/FI) processes; 2.5D/3D package integration process; wafer singulation and thinning processes; chip-package interactions; interconnection processes; substrate and interposer assembly processes; board assembly processes; sub-system integration and interconnection processes for SOC/SIP/SOP formats. Detailed reliability roadmaps for each of these topic areas are omitted here for brevity and are discussed in detail elsewhere in the literature[8].

The hardware reliability topic areas within the Package Integration Segment can be broadly grouped into three sub-segments: IC Reliability, Substrate/Board Reliability, and Interconnect/Assembly Reliability. Detailed tasks for these sub-segments are presented elsewhere in the literature8. Chiplet integration is key to HI and reliability concerns and includes: interconnect technologies such as TSV (Through-Silicon Via), RDL (Re-Distribution Layer), Cu/Cu bonds, µBumps and regular C4 bumps. Anticipated new failure modes due to chip-package interactions (CPI) in chiplet integration are: (a) Multi-physics failures under high current density, temperature, temperature gradient and thermal mechanical stress, electro-migration and stress migration induced voiding, interconnect cracking and interface fracturing; (b) stress effect on Cu/ELK (extra low k) die – Cu/ELK cracking and circuit performance drift. Methods to quantify and mitigate reliability risks will require a DOE (Design of Experiment) matrix to perform reliability studies of these new failure modes. The approach includes: (a) Simulation techniques to identify the critical factors for each major failure mode; (b) Design and fabrication of effective test structures to study the relevant failure modes; (c) Application of effective stress to accelerate the failure modes with the aid of RP models; (d) Statistical distribution plots based on failures in tests; (e) Acceleration factors to assess risk of in-service failures.

Based on such studies, reliability engineers can: (a) Extract reliability design rules for major failure modes; (b) Develop EDA (electronic design automation) flow to implement reliability design rules in the design flow; (c) Establish knowledge based qualification tests to qualify final packaged IC product; and (d) Identify difficult challenges and disruptive opportunities.

3. Difficult Challenges and Disruptive Opportunities

Reliability tasks must be closely synergistic with the HI technology roadmaps proposed by other TWGs, shown in Table 1. While it is very difficult to predict technology evolution over the next 15 years, it is possible to make some informed speculations[1-4] that new technologies may see a diverse mix of: (i) advanced nodes (below 5nm) vs conventional nodes (above 5 nm); (ii) conventional binary digital logic devices vs quantum processors and quantum computing devices; (iii) conventional microelectronics vs molecular-electronics (including transistors made from nanotubes and few-atoms); (iv) silicon vs wide bandgap (WBG) semiconductor devices; (v) microelectronic vs photonic devices; (vi) interconnects based on conventional vs advanced nanomaterials (e.g. 2D materials like graphene or stanene; and 1D nanomaterials like nanotubes and nanorods made from carbon and other conductive materials); (vii) complex materials and structures fabricated by additive vs conventional (subtractive) processes; and (viii) SOC vs SIP multi-device integration. Such revolutionary technology inflexions will be essential to get past the fundamental physical and economic limitations of Moore's Law. Significant changes will be required in materials, processes, equipment and metrology, e.g. the switch to cobalt as a conductor material.

Chip-Package Interactions (CPI) and package-interposer-substrate-module-interconnection interactions in single-chip and multi-chip integration for HI will be a major challenge and are listed below (details are presented elsewhere in the literature[8]): (a) global thermal mechanical stresses from Chip-Package thermal expansion mismatches; and from local stresses due to FinFET self-heating effects; (b) Stress-induced and temperature-induced transistor performance shift; (c) CPBI: microbump failure, microvia failure, RDL failure, TSV/TSG failures; (d) BEOL failure; (e) Electrical consequences of CPI; and (f) Management of CPI risk. The overall CPI challenges come from three aspects:

- The 'intrinsic strength' is continuously decreasing due to the scaling of Si technology. The adoption of ELK is one of the major reasons. Both the ELK material strength and the Cu/ELK interface strength are lower in more advanced nodes;
- On the other hand, the 'intrinsic stress' is continuously increasing due to such factors as increase in die size, reduction in die thickness, adoption of Pb-free interconnection bumps, and use of coreless substrates;
- Both the fab and assembly process variations are increasing in advanced Si nodes and packages. That causes wider distribution of both 'stress' and 'strength,' thus increasing their interference (and compromising reliability) in the more advanced nodes and packages.

Table 2: Difficult Challenges and Approaches for Reliable HI Technologies

	Reliability Targets	Life Cycle Conditions	Design for Reliability	Manufacturing for Reliability	Qualificaton for Reliability	Suustaining for Reliability	Supply Chain
SiP Technologies	**1-5 Years**: Multi-physics fusion approaches for co-design and reliability assurance of HI systems • Bottom-up Reliability Physics based approaches, tools, infrastructure • Top-down big data and Artificial Intelligence based approaches, tools, infrastructure						
Package Integration	**5-10 Years**: Fusion approaches for co-design (based on 'digital twins') and life-cycle PHM of next-gen robust HI systems • Fault-tolerant systems • Resilient systems						
Applications	**10-15 Years**: Fusion approaches for co-design and life-cycle management of future intelligent HI systems • Self-cognizant systems • Self-healing systems						

Table 2 provides a tabular discussion of the overall difficult challenges expected in making the proposed new leading-edge HI hardware technologies reliable, dependable and affordable. Detailed versions are presented elsewhere in the literature[8]. The biggest reliability challenges will be from the new yet-unknown degradation

modes and stochastic process/material variabilities that have already (and will continue to) become significant at small length scales and to prevent a 'zero- defects' philosophy. It will be hard to find highly integrated supply chains that have the know-how and ability to manage such complex reliability challenges.

Science-based approaches to reliable HI technologies require a thorough understanding of the failure modes and root-cause degradation mechanisms that are expected to dominate in current and future technologies. This team has initiated an effort to catalog the known and expected failure mechanisms in HI systems and an overview of the results are summarized in Table 3 (attached at the end of this document).

4. Approach and Roadmap for Addressing Difficult Challenges

To solve the reliability challenges that are likely to emerge in technologies discussed in the HI Technology roadmaps (and briefly overviewed in Section 3 above), we will need investments in new RP knowledge-base, AI methodologies, machine learning algorithms, and data management infrastructure (via IoT), to provide designers the right set of models and tools to allow life-cycle reliability management via use of 'digital twins.'

To deal with the technological complexities, dimensional down-scaling, new materials and processes, short technology development cycles and fast production ramps, the industry needs to invent and adopt lean, effective design methodologies, manufacturing ecosystems, qualification approaches, and reliability assessment and monitoring methodologies. Such a systematic proactive approach will offer unique 'cradle-to-grave' opportunities to reduce time to market for new product introduction (NPI) and to minimize the cost of ownership over the product's life-cycle and across multi-generational technology families. Innovations in reliability ecosystems are needed in the following areas:

Design for reliability (DfR): RP and AI knowledge-bases are needed for new degradation physics in ultra-complex multi-chip SiP architectures that are based on advanced semiconductor systems and new emerging packaging material systems. Co-design methods using 'digital twins' are needed to consider multi-physics, multi-scale CPI and reliability up-front in an integrated seamless concurrent manner.

Manufacturing for reliability (MfR): Process variabilities and materials variabilities will have to be characterized to quantify their effects on hardware reliability, using a combination of empirical studies, fundamental RP models and AI approaches. Co-existence of 3D additive and subtractive process technologies will require innovative solutions for process metrology and process control strategies, in order to minimize defect densities and maximize yield. 'Zero defects' will not be a realizable goal at these length scales and at this level of complexity, so solutions will need resilient product designs facilitated by massive redundancies and the use of virtual manufacturing simulations guided by 'digital twins.'

Qualification for Reliability (QfR): RP and AI tools will be needed to develop acceleration models for customizing accelerated stress testing, for EVT/DVT/PVT, based on the specifics of individual designs, life cycle usage conditions and reliability goals. Such knowledge-based customized testing must be adopted across the supply chain. Innovations will be needed for built-in testing strategies for ultra-complex HI architectures. Physical testing may become increasingly non-economical and must be supplemented with RP/AI-aided virtual testing of 'digital twins.' The IoT infrastructure must be harnessed to harvest fertile reliability data in real-time from fielded products, in order to supplement EVT/DVT/PVT findings.

Sustainment for Reliability (SfR): Fusion of RP and AI methods will be essential for personalized real-time PHM of fielded products in the post 'zero-defect' era. Integrated PHM canaries must become a regular feature of self-cognizant, intelligent, bio-mimetic hardware that can survive and function and 'age with grace' instead of failing unexpectedly in service. Innovations must include neuro-morphic adaptive reconfigurability and self-healing capabilities.

Supply-chain integration: New supply-chain models and management practices will be essential to transition to complex new multi-physics HI ecosystems. New IP business models and novel availability-based contracting models will have to be developed.

5. Software Reliability

Traditional reliability approaches are focused on prevention of hardware failures, but software reliability will have to be an integral part of future HI systems with enhanced connectivity and adaptive control. Software reliability methods are fundamentally different from hardware reliability methods. Software reliability or robustness is the probability of failure-free software operation for a specified period of time and environment. Unlike hardware failures, software systems do not degrade over time unless modified. Software failures are caused not by faulty components, wear-out or physical environment stresses such as temperature and vibration; instead by latent software defects that were introduced into the software as it was being developed, but were not detected and removed before

the software was released to customers. The best approach for achieving higher software reliability is to reduce the likelihood of latent defects in the released software. Software reliability growth models (SRGM) are based on mathematical functions that describe fault-detection and removal phenomenon in software [9]. These models, in combination with Bayesian statistics, need further attention within the hardware-oriented reliability community in the coming years. The 10-year horizon includes the items listed below to be given central visibility/priority in the HI roadmap:

- Develop software reliability growth models able to predict the remaining number of software defects (bugs).
- Create mainstream industry eco-systems for approaches for validated bug-finding rates.
- Establish techniques that can combine SW reliability metrics with HW reliability metrics.

5. Summary

This document is a preliminary high-level description of Reliability Tasks for making single- and multi-chip HI systems highly reliable, available, and affordable. In the interest of brevity, this document provides only an overview while details are provided elsewhere in the literature[8]. Potential difficult challenges with solution approaches, and necessary infrastructure are discussed here. This section has laid out the importance of an integrated approach towards reliable HI systems, based on 'digital twins' that result from strategic integration of RP with powerful AI algorithms. Such hybrid approaches will have to leverage the unprecedented level of real-time field reliability data that is becoming available via IoT infrastructure. The business case for such an integrated approach rests in the tremendous opportunities for reducing time to market and 'cradle-to-cradle' cost of ownership across multiple generations of NPI. The present version of this document is necessarily generic and lacking in specific details since reliability practices are dependent on the specifics of the technology roadmap. This TWG will continue to work with other TWGs in the HIR Roadmap team, to continue to add granularity and specificity to this section in future versions. Furthermore, in future versions, this document will expand its scope to include other aspects of system reliability (including software security and dependability of human-machine interactions) for the entire HI ecosystem.

References

1. L. DiCarlo, et. al., "Demonstration of two-qubit algorithms with a superconducting quantum processor," Nature, volume 460, pages 240–244 (09 July 2009).
2. Will Knight, "Quantum Inside: Intel Manufactures an Exotic New Chip," MIT Technology Review, October 10, 2017
3. Dong Xiang, Xiaolong Wang, Chuancheng Jia, Takhee Lee, and Xuefeng Guo, "Molecular-Scale Electronics: From Concept to Function," Chem. Rev., 2016, 116 (7), pp 4318–4440, DOI: 10.1021/acs.chemrev.5b00680
4. Ayelet Vilan, Dinesh Aswal, and David Cahen, "Large-Area, Ensemble Molecular Electronics: Motivation and Challenges," Chem. Rev., 2017, 117 (5), pp 4248–4286, DOI: 10.1021/acs.chemrev.6b00595
5. Dasgupta, A., Contributor, Integrated Circuit, Hybrid, and Multichip Module Package Design Guidelines, Editor M. Pecht, Pub. Wiley Interscience, 1994.
6. Dasgupta, A., Contributor, Plastic Encapsulated Microelectronics, Ed. M. Pecht, L. Nguyen, E. Hakim, pub. Wiley, 1994.
7. Dasgupta, A., "Hardware Reliability," Chapter 5, pp. 95-133, in Product Reliability, Maintainability, and Supportability Handbook, Editor M. Pecht, CRC Press, 1995.
8. Dasgupta, A., Rao, R., Sahasrabudhe, S. W. VanDriel, "Detailed Perspective of Reliability Roadmap for HI Systems," to be published.
9. H. Pham, "System Software Reliability," in Springer series in Reliability Engineering, London, Springer, March, 2006, ISBN-10: 1852339500

Table 3. Summary of Failure Modes and Mechanisms expected to dominate in current and future HI systems

Multi Scale and Multi Physics Reliability Failure Modes and Design Simulations, *by Richard Rao*												
Scale of Integration	Multi Scale and Multi Physics	Electrical Stress		Moisture		Thermal Mechanical Stress		Mechanical Stress		Simulation/Modeling and Co-Design Flows		
		Failure Modes	Failure Mechanism & Reliability Models	Failure Modes	Failure Mechanism & Reliability Models	Failure Modes	Failure Mechanism and Reliability Models	Failure Modes	Failure Mechanism & Reliability Models	Failure Modes	EDA Flows	PDK/ADK
Transistor	FinFET/GAA IGBT/WBG/SOI device failures	1. FinFET/GAA NBTI/PBTI 2. Hot Carrier Injection 3. Gate Dielectric Breakdown 4. ESD	1. BTI Models with recovery effect 2. HCI Model 3. TDDB Weibull model 4. Oxide and junction breakdown	No known failures	None	1. Self Heating Effect induced localized stress in FinFET channel and in GAA 2. uBump/C4 bump/TSV and package/system level stress transmission to FinFET. 3. Channel stress alter FinFET/GAA characteristics	1. FinFET/GAA SHE Model a. Steady Model b. Transient Model c. Model with heatsink effect 2. e-CPI Model 3. Piezo-electrical models	No known failures	None	1. Transistor aging induced circuit degradation; 2.How the thermal and stress effects be considered in circuit aging simulation? 3. How does the process variation interact with aging for 16nm/7nm? How should this interaction been simulated?	1. Cadence Relxpert 2. Mentor Graphics 3.	1. Device SPICE Model 2. Device aging model 3. SHE model 4. TDDB model
Inter-connects	MEOL/BEOL Metal/Via /ELK	1. Electromigration 2. Inter Layer Dielectric ELK Breakdown 3. MEOL Oxide Breakdown 4. EOS	1. Pad and underline metal corrosion 2. Cu/ELK delamination	1. Electrochemical corrosion 2. Interface degradation due to moisture absorption	1. Stress Migration 2. Joule/SHE effects on the stress in Cu/ELK; 3. uBump/TSV/RDL and package/system level stress transmission to MEOL/BEOL. 4. Cu metal line fatigue	1. Creep induced voiding 2. CTE mismatch 3. CTE mismatch 4. SHE induced localized thermal cycling	No known failures	None	CPI induced Cu/ELK cracking - how will the internal stress due to JHE/SHE and external stresses from bumps/TSV/RDL/Packaging affect the failure of Cu/ELK over time?	1. Ansys/Mentor	
	FBEOL RDL/Dielectric	1. RDL Electromigration		1. 2D. cracking	ubumps/TSV/Package/Board effects on RDL stress				CPI or CBPI induced RDL cracking/delamination - how will the stress from bumps/TSV/Package/board level stresse affect RDL failure?	EM- How does the temp and stress affect RDL EM?		
	Au/Cu Wirebonding	1. Electromigration		1. IMC Corrosion	1. Bond wire fatigue		1. Cu/ELK cracking	Bonding force induced				
	uBump/C4 Bump/UBM	1. Electromigration			1. Bump joint cracking 2. Bump joint fatigue	1.Thicker IMC from uBumps			Multi Physics Bump FM - How does local current, temp, temp gradient and stress affect bump EM?	Bump fatigue - How does local temp, temp variation and stress affect fatigue life?		
	TSV/Interposer	1. Electromigration 2. Barrier Dielectric breakdown			1. Cu pumping/TSV pop up	1. Si and Cu CTE mismatch causing Cu extrusion 2. Non reversible plasticity at high temp			1.TSV FM - How does local current, temp, temp gradient and stress affect TSV FM? 2. TSV Pop Out and effects on TSV/Si interface delam	Barrier breakdown - How does voltage/current, temp and stress affect TSV barrier BD?		1. Package material thermal/mechanical properties 2. Die metal stack and thermal/mech properties 3. uBump/C4 bump/TSV thermal mechanical properties 4. Cu/ELK Fracture criteria 5. Void initiation and propagation criteria 6. Interconnect fatigue/creep model 7. Package interface fracture criteria 8. Moisture diffusion and vapor pressure model 9. IMC thermal/mech/electrical properties 10. Photonics optical properties
Packaging/ System	Passivation	1. Passivation cracking	1. EOS induced cracking	1. Passivation cracking 2. Passivation delamination 3. Delamination between passivation and underfill/Molding compound	1. Passivation cracking	1. Stress from packaging 2. Stress concentration in SiN						
	Underfill			1. Underfill to die/substrate delamination 2. UF Expansion	1. Moisture degrade bonding interface 2. UF Moisture absorption	1. Bump joint cracking						
	High Density Substrate	1. Metal trace electromigration		1. Metal trace corrosion		1. Metal trace/via cracking				Cu trace/Via cracking - How does voltage/current, temp and stress affect Cu trace/via cracking?	Cu trace/Via cracking - How does voltage/current, temp and stress affect Cu trace/via cracking?	
	Wafer Level Package					1. Warpage						
	Fanout Package					1. Warpage						
	2.x/2.5D Interposer Package					1. Warpage						
	3D Package					1. FinFET Ion shift due to a. TSV/Si CTE mismatch b. uBump stress c. Underfill/EMC shinkage 2. TSV effects on BTI/HCI 3. BEOL cracking 4. Cu pillar joints failure						
	Laser/Receiver/ Modulator											
	SiPhotonics Chip											
	Co-packaged IC and Photonics					1. Optics misalignment						
	Wafer Level Integration of IC and Photonics											
	Known Good Die	ESD				1. Die edge cracking 2. Under bump ELK cracking						
Module/ System	Printed Circuit Board Assembly					1. Solder joint cracking 2. PCB failure		1. Solder joint cracking 2. Pad damage		Board level Solder Joint Reliability	ANSYS Mechanical	1. PCB thermal/mech properties 2. Solder joint fatigue/creep model 3. Solder joint dynmnic/impact properties

Section 13 Summary, Difficult Challenges and Contributing Team

1. Summary and Difficult Challenges

This chapter starts with an Executive Summary followed by 12 sections on key technology building blocks from Knowledge Base and Data to Manufacturing and Physical Infrastructure.

☐ Knowledge Base and Data
- Electrical Analysis and System Requirements
- Thermal Management
- Mechanical Analysis
- Electromigration
- Reliability

☐ Manufacturing and Physical Infrastructure
- Wafer Singulation and Thinning
- Wire bond
- Flip Chip
- Substrate
- Board Assembly
- Additive Manufacturing

The Knowledge Base and Data, and Manufacturing and Physical Infrastructure, form the base foundation for Assembly, Packaging and Integration technologies for all market applications from High Performance Computing – Data Centers, Mobile, Automotive, IoT and 5G, Wearables, and Aerospace and Defense.

Electronics are deeply embedded into the fabric of our society, changing the way we live, work and play while bringing new efficiencies to global lifestyle, industry, and business. We are entering the era of the digital economy and myriad connectivity. Our industry has reinvented itself through multiple disruptive changes in market, products, and technology.

At this triple inflexion point of rise of tech disruption, with Moore's Law plateauing and the explosive expansion of digital data growth, and developments in IoT and artificial intelligence, continued progress requires a different phase of electronics innovation.

In the later part of Gordon Moore's celebrated 1965 paper [1] he adopted a system focus: "It may prove to be more economical to build large systems out of smaller functions, which are separately packaged and interconnected. The availability of large functions, combined with functional design and construction, should allow the manufacturer of large systems to design and construct a considerable variety of equipment both rapidly and economically."

Our purpose in Heterogeneous Integration is to build large systems out of smaller functions – System in Package – which are separately designed, packaged, interconnected and manufactured. The twelve sections in this chapter articulate the basic tool sets – infrastructure and knowledge – for heterogeneous integration for all the market segments.

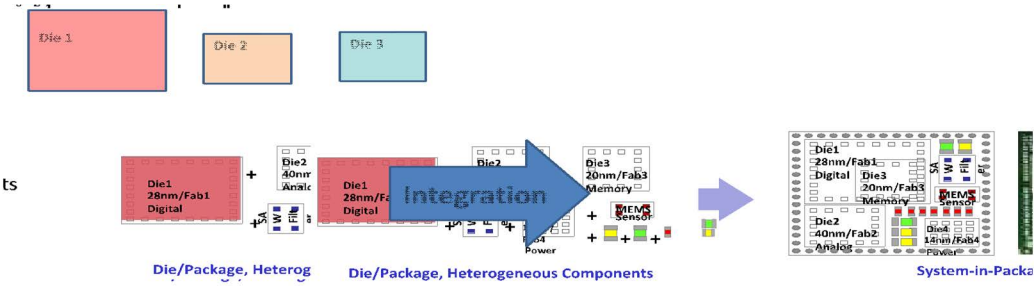


Figure 13.1: Heterogeneous Integration & System in Package (SiP) Source ASE

This single- and multi-chip chapter has been designed to understand the current state of the art of the key technology building blocks from knowledge base and data to manufacturing and physical infrastructure, and to ask the questions:

- What are the challenges looking into the future?

- What are the potential solutions?

In this chapter we have described the current state of the art, and roadblocks in the path going forward, to stimulate pre-competitive research and innovations 15 years ahead.

For this 2020 version of the Heterogeneous Integration Roadmap (HIR 2.0), we shall collaborate more closely and dig deeper with other TWGs into each other's difficult challenges and potential solutions for a stronger and more effective message to the Roadmap stakeholders and users.

2. Difficult Challenges:

While each section has articulated their difficult challenges and potential solutions, we have summarized the major challenges, issues and their potential solutions below:

Challenge: Integration of multiple die or devices from different sources into a single SiP.

Issues: Wafer yield, design, and sourcing for bare die from different nodes and sources, test for KGD, interface protocol, test for known good SiP

Potential Solutions: The examples shown together with chiplet initiatives, and the DARPA Chips program, have shown that the business and technical issues are being resolved.

Challenge: Warpage and Stress

Issues: Differential thermal expansion of different materials, particularly between silicon and many packaging materials including laminate substrate. Warpage and stress control for advanced nodes (3 nm) and automotive applications will be essential

Potential Solutions: Low CTE materials development, 3D warpage imaging metrology, co-design for stress simulation through product assembly and harsh environment.

Challenge: Signal integrity - package and board

Issue: laminate materials properties and copper conductor surface roughness

Potential Solutions: new organic materials and interface process development, full scale simulation.

Challenge: High Volume and High Mix manufacturing and assembly

Issue: Consumer drive for "custom" products

Potential Solution: integrate additive manufacturing into mainstream manufacturing infrastructure

Challenge: Length of Product Design and Development Cycle

Issue: High complexity of design, development, and qualification across market segments; automotive, HPC, aerospace and defense, mobile infrastructure, wearables and health

Potential Solution: Full utilization of AI & Sensor Driven data across production line, field trial and ramp-up

3. Acknowledgments and Contributing team

William Chen, Annette Teng, Lei Shan, Benson Chan, Ivy Qin, Mark Gerber, Kyu-Oh Lee, Jim Wilcox, Bahgat Sammakia, Abhijit Dasgupta, Richard Rao, Shabhada Sahasrabudhe, Paul Ho, Valeriy Sukharev, Stéphane Moreau, Kris Erickson, Eugene Chow, Richard Misawa, Hugo Pristauz, Stefan Behler, Willem van Driel, Dae Young Jung, Seungbae Park, Urmi Ray, Ravi Mahajan, Brandon Prior, Bob Chylak, Srikanth Rangarajan, Scott Schiffres


William Chen (ASE)


Lei Shan (Tekollect)


Benson Chan (Binghamton U)


SB Park (Binghamton U)


Annette Teng (Promex)


Ivy Qin (K&S)


Mark Gerber (ASE)


Kyu-oh Lee (Intel)


Jim Wilcox (UIC)


Kris Erickson (Facebook)


Paul Ho (U Texas)


Stephane Moreau (CEA)


Valeriy Sukharev (Mentor)


Abhijit Dasgupta (U Maryland)


Richard Rao (Marvel)


Shubhada Sahasrabudhe (Intel)


Bahgat Sammakia (Binghamton U)


Srikanth Rangarajan

Edited by Paul Wesling