HIR

HETEROGENEOUS INTEGRATION ROADMAP

2019 Edition

Chapter 8: Single Chip and Multi-Chip Integration

http://eps.ieee.org/hir

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

50-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 "The future of computing" in the business magazine *The Economist* as "Moore's Law Saturation" (Figure 1). The semiconductor industry is implementing EUV and FinFET technology at the 7 nm node. The 5 nm half-node and 3 nm node are in sight. The message in the March 2016 article is as relevant as ever today.
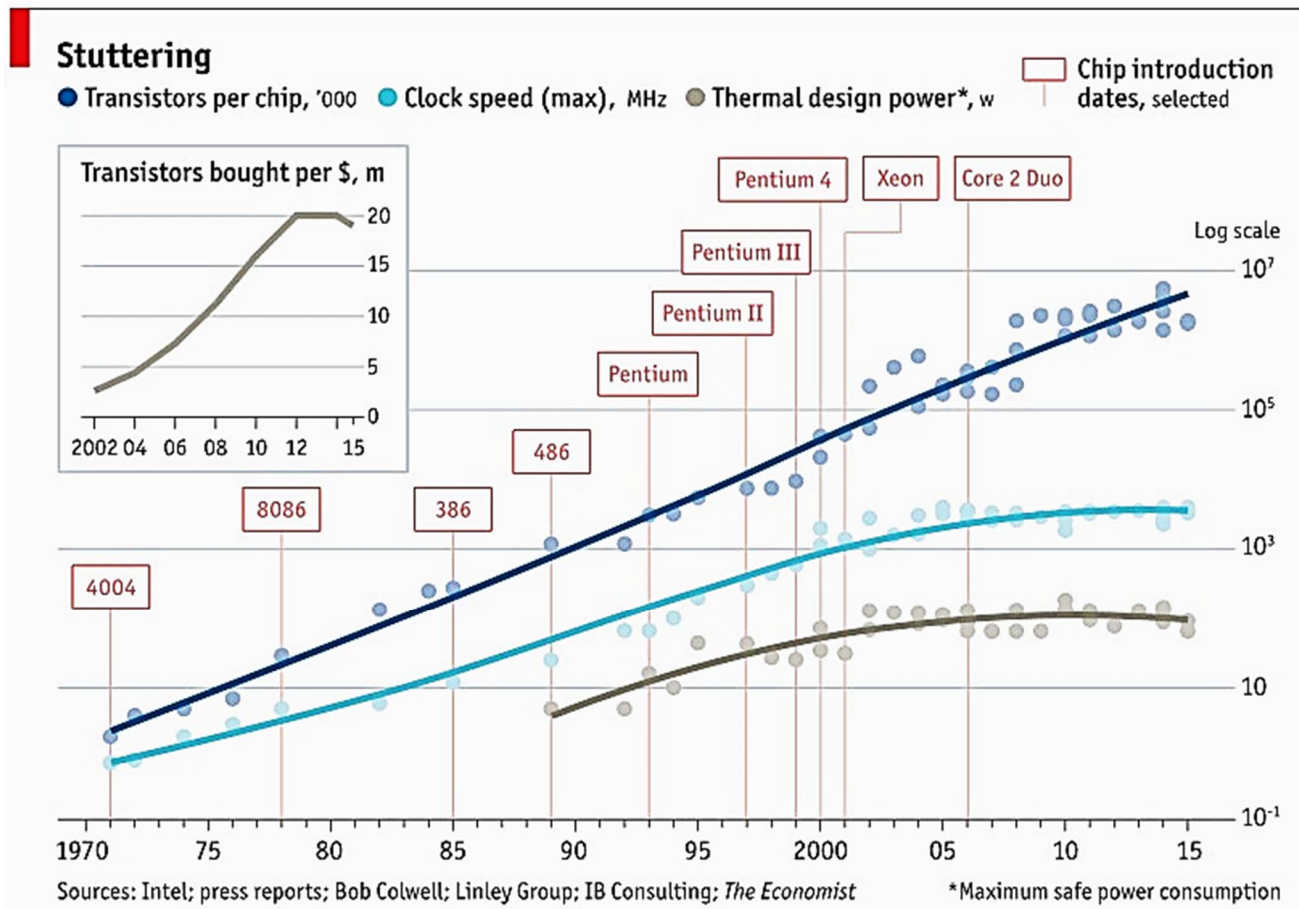
Figure 1. Moore's Law Saturation – Performance & Economics. (Source: The Economist March 12, 2016)

We are entering the era of the digital economy and massive connectivity, with data migration to the cloud, smart devices everywhere, Internet of Things to Internet of Everything, the introduction of 5G, and the emergence of autonomous vehicles. The business landscape is seeing great changes with the rise of technology 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 listed the top 10 publicly traded companies by market capitalization in 2006 and again in 2019. While there was only one technology company in 2006 in that list, in 2019 the top 5 of the 10 are all technology companies, signifying this transition to the digital age.
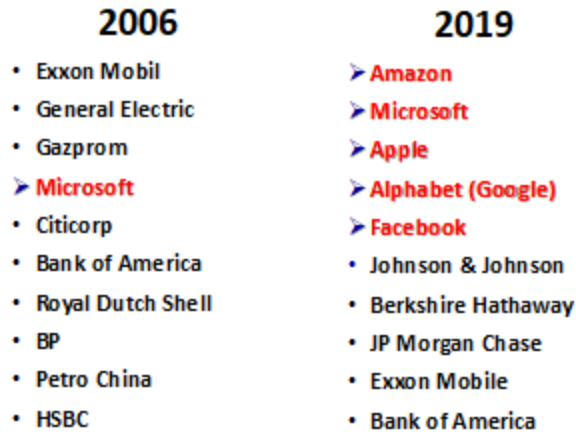
Figure 2. Rise of the tech companies (Source: www corporateinformation.com, Feb. 2019)

At this triple inflection point of plateauing of CMOS's scaling advantage, the transition to big data analytics and artificial intelligence in the digital economy, 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 our industry has succeeded tremendously well indeed, in investment, technology and science of scaling, "cramming" billions of transistors into integrated circuits from wafers to chips and chips to products, fostering a global electronics industry for the benefit of society.

In the same 1965 paper, Dr. Moore turn 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 – System in Package (SiP) – 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 assembly (SiP) that, in the aggregate, provides enhanced functionality and improved operating characteristics.

> *In this definition, components should be taken to mean any unit 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 for this enormous advance.

The Heterogeneous Integration Roadmap addresses six specific market segments: High Performance Computing and Data Center, IoT and 5G, Smart Mobile, Automotive, Wearable & Health, and Aerospace & Defense, as illustrated below in Figure 3.

Figure 3. Six major applications markets

This Single and Multichip Integration Chapter covers the basic knowledge-base tools and physical manufacturing infrastructure tools across all of these market segments. IC devices start with wafers from foundries, thinned and singulated into "chips". While wirebond remains the workhorse of the industry, there is very strong growth in dollar value for Flip Chip (BGA-CSP) and WLCSP, and specifically high rapid dollar-value growth for advanced packages such as 2.5D, FO WLP/PLP, as shown in Figure 4.
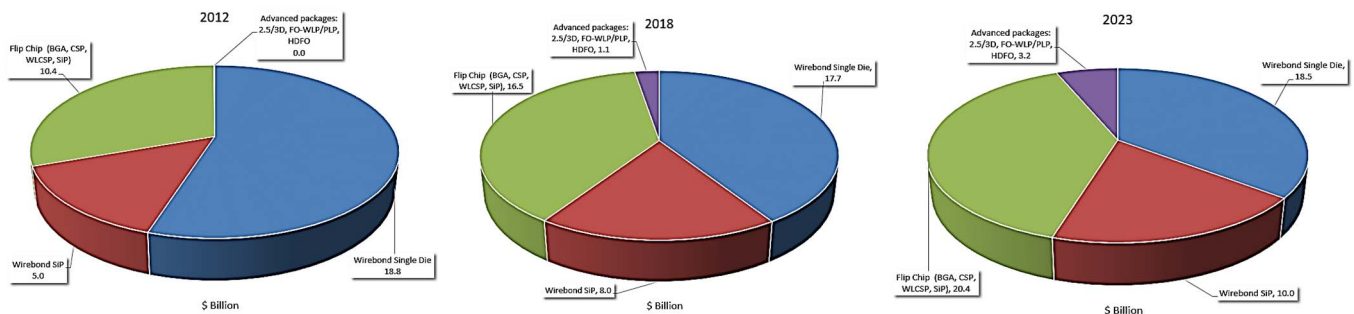
Figure 4.   Chip Package interconnect technology trends (US$ billions) (Source: Prismark Partners 06-2019)

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 they form the knowledge and manufacturing infrastructure for 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 which are separately designed, packaged, interconnected, qualified, manufactured and sold.

Tool Box for Heterogeneous Integration (Source: ASE)

Figure.5. Tool box for heterogeneous integration technologies (Source: ASE)

A basic tool box for SiP and heterogeneous integration is illustrated in Figure 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 – well-established companies and new startups building SiPs through heterogeneous integration for 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 & knowledge – for heterogeneous integration for all the market segments.
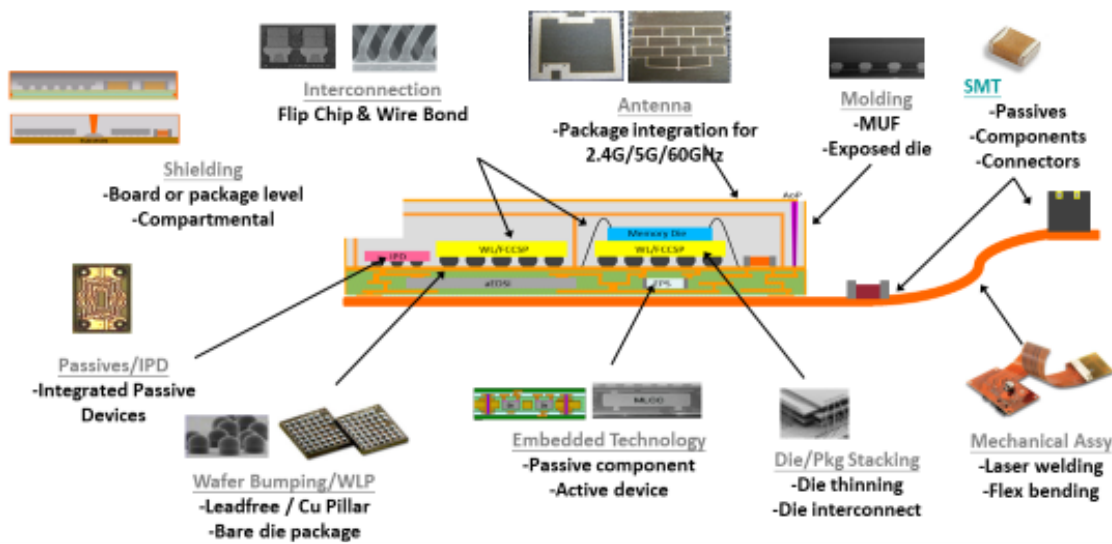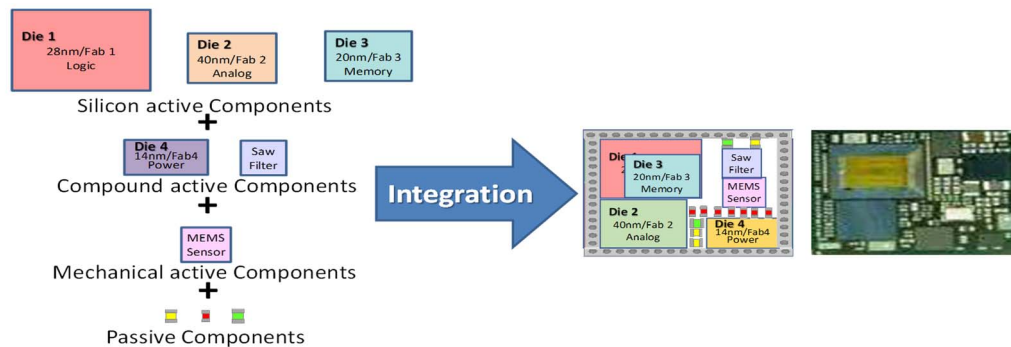

Figure 6. Heterogeneous Integration & System in Package (SiP) (Source: ASE)

The Mobile Smart Phone industry has adopted heterogeneous integration through SiP very effectively, developing new functions and expanding capabilities in their products while maintaining the phone's form factor in this highly competitive global consumer market. This is a very good example of Dr Moore's suggestion – an industry which has been an early adopter of heterogeneous integration technology in the use of System in Package (SiP) for their advantages in miniaturization, modularity and co-design, from product debut to subsequent generations of product roll out to the market place. The application processors, housed in PoP packages with the memory component stacked in close proximity on the top, is almost always of the most advanced node (now at 7 nm). They are the premier examples of System-in Package being able to incorporate the most advanced-node processor integrated with memory die in close proximity for this demanding consumer market.

The list accompanying the figure:

1. Apple A11/DRAM (InFO PoP)
2. Samsung Flash (Stacked die)
3. Qualcomm MDM9655 BBP/Memory (stacked die)
4. Qualcomm WTR5975 TRx (WLP)
5. Skyworks GSM PA (SiP)
6. Skyworks 3G/4GPAD (SiP)
7. Broadcom 3G/4GPAD (SiP)
8. Skyworks Diversity Rx Module (SiP)
9. Skyworks Diversity Rx Module (SiP)
10. Qualcomm Envelope Tracker (SiP)
11. USI WLAN/BT Module (SiP)
12. Qualcomm PMD9655 Power Manager (WLP)
13. Power Manager (WLP)
14. Audio Codec (WLP)
15. Audio Amplifier (WLP)
16. Audio Amplifier (WLP)
17. Audio Amplifier (WLP)
18. Broadcom Wireless Charger (WLP)
19. TI Battery Charger (WLP)
20. NXP Display Controller (WLP)
21. NXP NFC Controller (WLP)
22. Bosch MEMS Accel/Gyro (SiP)

Figure 7: Apple iPhone X main board (source: Prismark Partners)

Shown in Figure 7 is the main board for the iPhone X. The main A11 processor die is housed in a PoP package with flip-chip, assembled on an advanced substrate, together with a wirebonded memory component on the top. There are 8 other SiPs and 11 WLC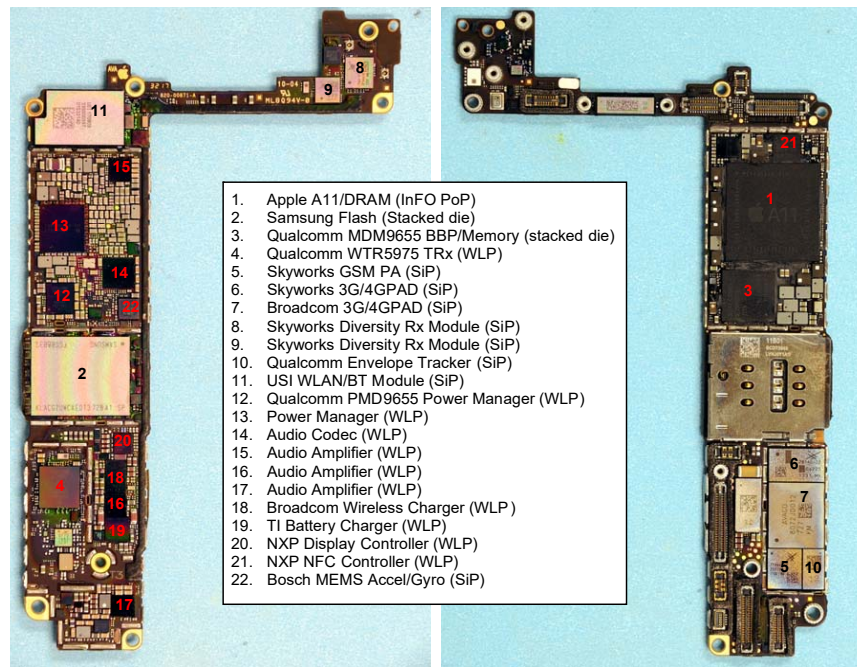SP-packaged devices tightly assembled with many passives on a rigid-flex board. To paraphrase Dr. Moore, availability of large functions in the form of SiPs, combined with functional design and construction, should allow the manufacturer of large systems (such as the smartphone) to design and construct smartphone models from one 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 integration of the processor and high-bandwidth memory (HBM) stack in a 2.5D silicon interposer platform. The package provided massively parallel high-bandwidth connectivity to the HBM, a significant power saving, and greater than 50% shrinkage in X-Y form factor.
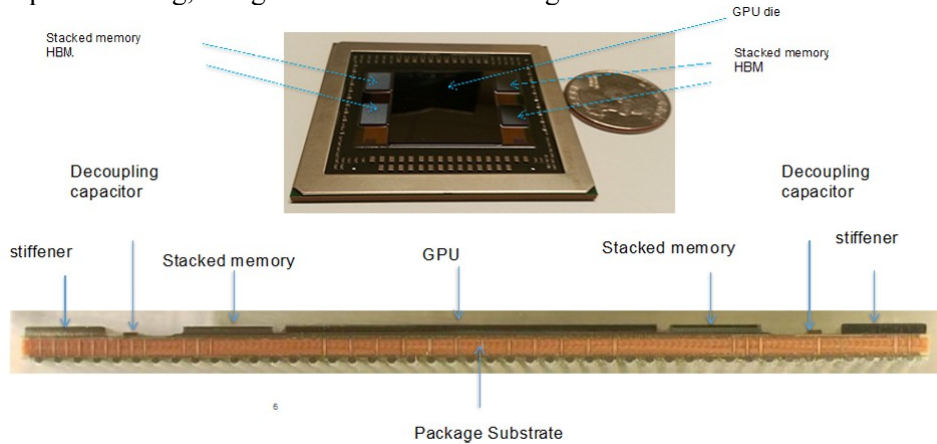
Figure 8: AMD Fiji GPU-HDM on Si Interposer 2.5 D Package (Source: ASE)

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 an ecosystem of discrete modular IP blocks, to be assembled into a system using existing and emerging integration technologies.

At the advanced nodes the die yield falls exponentially with die size. At the same time, die cost per unit area is escalating [2]. Splitting a large monolithic SoC into smaller tightly coupled die, first demonstrated by Xilinx on a silicon interposer, is now being seriously considered and executed. Shown below (Figure 9) 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, using 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 using heterogeneous integration to optimize unit area die cost.
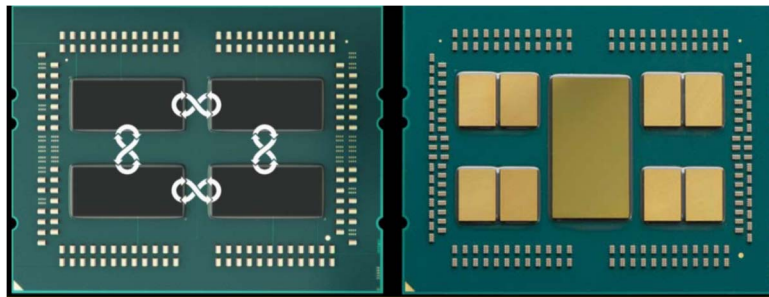
Figure 9: Examples of System Integration - EYPC Server Processors Source AMD

Components of different nodes or from different companies may be heterogeneously integrated together in one SiP, such as implemented in the Intel Kaby Lake G card incorporating Intel CPU with AMD GPU linked to 4GB of HBM2. Silicon bridge on organic substrate (EMIB), developed by Intel (Figure 10), was used to link multiple die together in close proximity in the package [3].
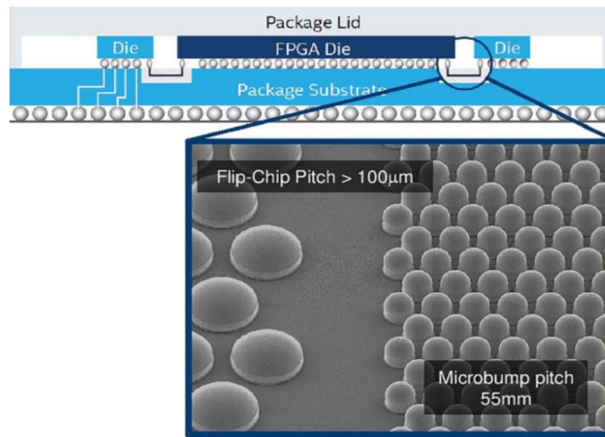
Figure 10: Silicon bridge on substrate technology (EMIB) from Intel for Heterogeneous Integration (Source: Intel)
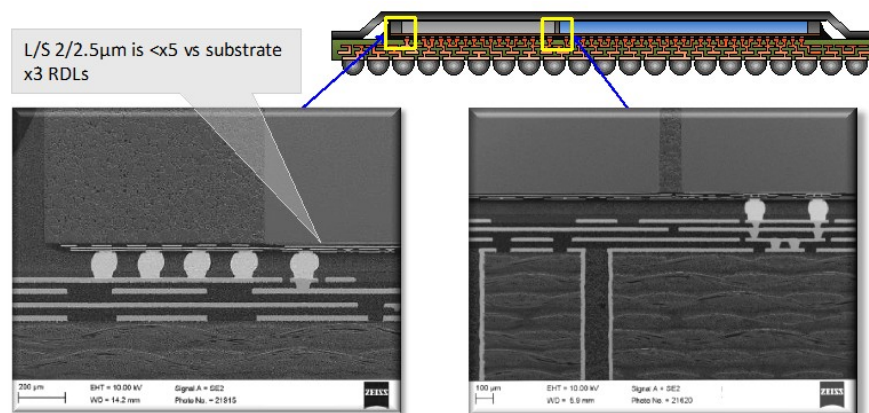
Figure 11: Wafer Level Fan-Out integrating 14 nm & 22 nm die in multichip package (Source: ASE)

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 (Figure 11).

The previous examples demonstrate the growing momentum for SiP in high-performance computing taking a new look at system architecture and expanding innovations using our packaging tool box. 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 is an important trend going forward [4-10].

The scope of this Single Chip & Multichip chapter is designed to present 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 heterogeneous integration?
- What are the challenges ahead, looking into the future?
- What are the potential solutions?

We shall address these questions, and include in this roadmap 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 – System in Package (SiP) – which are separately designed, packaged, interconnected and manufactured. This chapter starts with this section on Scope followed by 11 sections on key technology building blocks, from 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
  - Wirebond
  - Flip Chip
  - Substrate
  - Board Assembly
  - Additive Manufacturing

These two broad categories – **Knowledge Base & Data**, and **Manufacturing & Physical Infrastructure** – form the base foundation for advanced packaging and integration technologies, spanning SiP (chapter 21), 2D, 2.5D & 3D (chapter 22), and wafer-level packaging (chapter 23). This chapter works closely with the chapters of other Technical Working Groups (TWGs) in this Heterogeneous Integration Roadmap. Working across the TWGs from diverse market drivers, we will develop specific focus to guide future directions. 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.

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 the 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 Computewre 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

2.1 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 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. Heterogeneous system integration also drives the increase in package dimensions and hence maximum pin count as well as package profile. Various interconnects will be detailed in the following sections.

2.2 On-package interconnections

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.

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:

Year of Production	2018	2019	2020	2021	2022	2025	2028	2031	2034
Chip size (mm^2)									
Memory (DDR/HBM)	100/40	60/90	60/90	60/90	60/130	60/130	60/130	60/130	60/130
Smart Phone / Laptop	85/140	80/130	80/130	80/130	80/130	80/130	80/130	80/130	80/130
High-performance (note1)	750	500	500	500	500	500	500	500	500
Maximum Average Power Density (W/mm^2)									
Memory (DDR/HBM) (note2)	0.06/0.15	0.06/0.12	0.06/0.12	0.06/0.12	0.05/0.11	0.04/0.1	0.04/0.09	0.04/0.08	0.04/0.07
Smart Phone / Laptop	0.4	0.45	0.45	0.45	0.5	0.55	0.6	0.65	0.7
High-performance	0.85	0.9	0.9	0.9	0.95	1	1.05	1.1	1.15
Core Voltage (Minimum Volts)									
Memory (DDR/HBM)	1.2/1.35	1.1/1.2	1.1/1.2	1.1/1.2	1/1.1	0.9/1	0.8/0.9	0.75/0.8	0.75/0.75
Smart Phone / Laptop	1	0.85	0.85	0.85	0.8	0.8	0.75	0.75	0.75
High-performance	1	0.85	0.85	0.85	0.8	0.8	0.75	0.75	0.75
Package Pin count Maximum									
Memory (DDR/HBM)	288/2400	288/3200	288/3200	288/3200	288/3200	326/4100	326/4100	350/4700	350/4700
Smart Phone / Laptop	960/4094	1155/5260	1155/5260	1155/5260	1212/7000	1275/7000	1396/9600	1444/9600	1588/11000
High performance (note3)	6400	6400	6400	6400	7800	7800	7800	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	0.19/0.75	0.19/0.75	0.19/0.75	0.19/0.75	0.19/0.75	0.19/0.75
High-performance	1.5	1.5	1.5	1.5	1.5	2	2	2	2
Performance: On-Chip									
Memory (DDR/HBM), MHz	400	800	800	800	800	1200	1600	2000	2400
Smart Phone / Laptop, GHz	2.5/4	2.8/6.4	2.8/6.4	2.8/6.4	3.2/8.0	4/9.6	5.2/11.2	6.4/12.8	8.0/16
High-performance, GHz	4	6.4	6.4	6.4	8	9.6	11.2	12.8	16
Interconnect: Chip-to-Chip (note4)									
Memory (DDR/HBM), Gb/s	3.2/1.0	6.4/2.0	6.4/2.0	6.4/2.0	8.0/4.0	9.6/8.0	12.8/12.8	16/16	25/25
Smart Phone / Laptop, Gb/s	25	50	50	50	100	100	200	200	200
High-performance, Gb/s	28	56	56	112	112	224	224	224	224
Interconnect: Pkg-to-Board									
Memory (DDR/HBM), Gb/s	3.2	6.4	6.4	6.4	8	9.6	12.8	16	25
Smart Phone / Laptop, Gb/s	25	50	50	50	100	100	200	200	200
High-performance, Gb/s	28	56	56	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)	75	70	70	70	70	70	70	70	70
Operating Temperature Extreme: Ambient (℃)									
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	(white)
Manufacturable solutions are known	(yellow)
Manufacturable solutions are NOT known	(red)

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

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 erupt unprecedentedly and result 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.
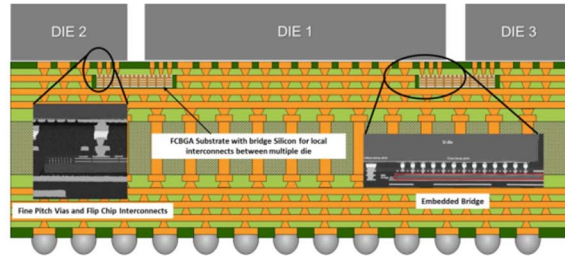
Figure 1: Embedded Multi-die Interconnect Bridge (EMIB) [1]

- **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 to 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 for designs 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.
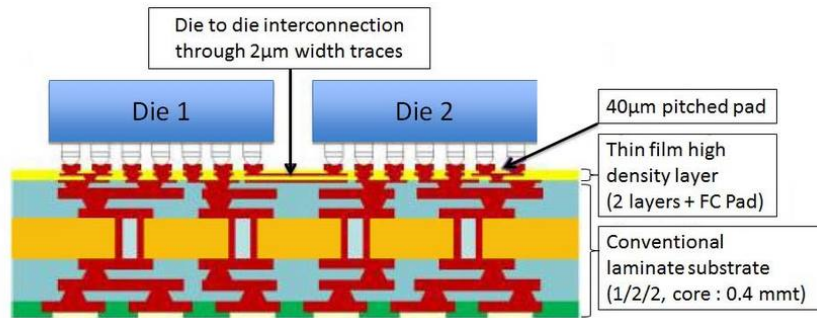
Figure 2. Emerging high-density organic substrate [2]

- **3D integration on the horizon**: At the end of 2018, Intel announced 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.
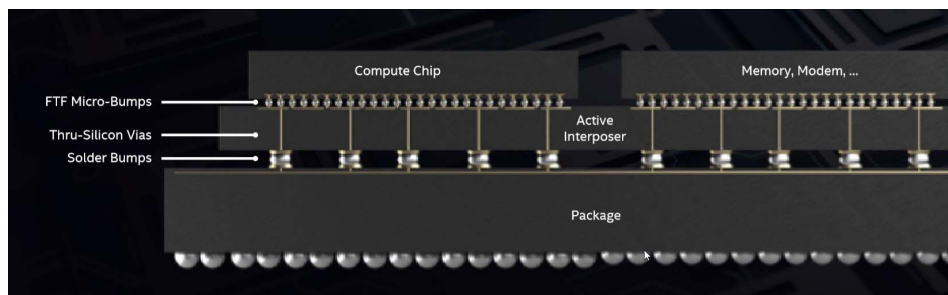
Figure 3. FOVEROS 3D chip stacking by Intel [3]

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.
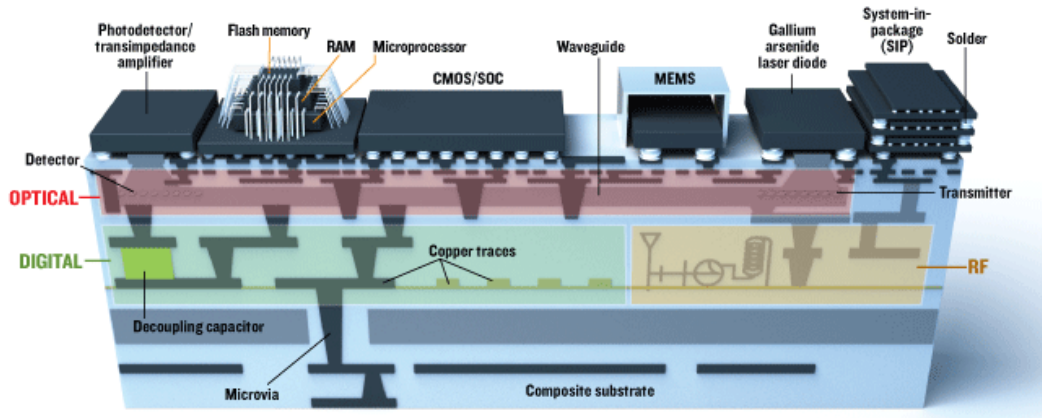
Figure 4: 3D heterogeneous integration

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 have 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.

2.3 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

- **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 was hardly settling down in 2017, the industry had already started searching 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.
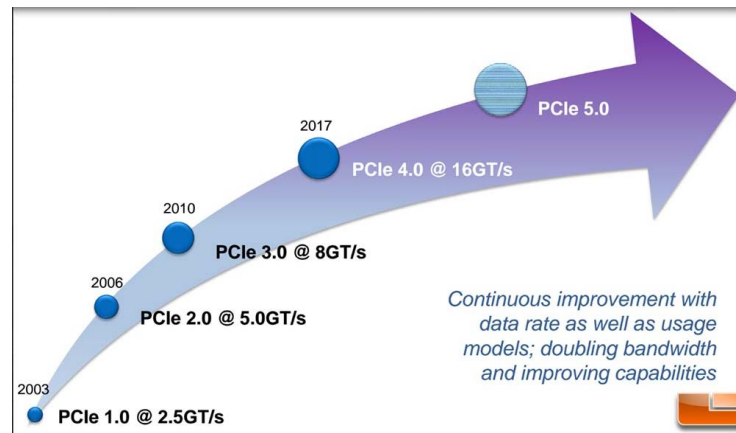
Figure 5: PCIe standard evolution [5]

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. The table below shows SERDES I/O speed, distance, and channel topologies. Off-package 56Gbps data-rate is expected by 2020 with PAM4 signaling.

Table 1: SERDES I/O speed, distance, and channel topologies (compiled from various sources)

Parameter	USR	XSR	VSR	MR	LR
Data Rate (Gbps)	18 -58	36 -58	36 -58	36 -58	36 -58
BER	1E-15	1E-15	1E-15	1E-15	1E-15
Distance	10 mm (~0.4")	50 mm (~2")	150 mm (~6")	500 mm (~20")	886 mm (~34.9")
Interconnect	MCM	PCB+ 0 connector	PCB + 1 connector	PCB + 1 connector	PCB + 2 connectors
Insertion Loss (dB) (at f_N)	2	4 (PAM4) 8 (NRZ)	10 (PAM4) 23 (NRZ)	20 (PAM4) 36 (NRZ)	30 (PAM4)
Modulation	NRZ	NRZ or PAM4	PAM or NRZ	PAM4	PAM4
FEC	N	N	Y/N	Y	Y

- **Electrical/optical "flyover" cabling:**

To mitigate the impairments of via and solder joint transitions in package and PCB, direct "flyover" cabling as shown in the figure below 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, 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.
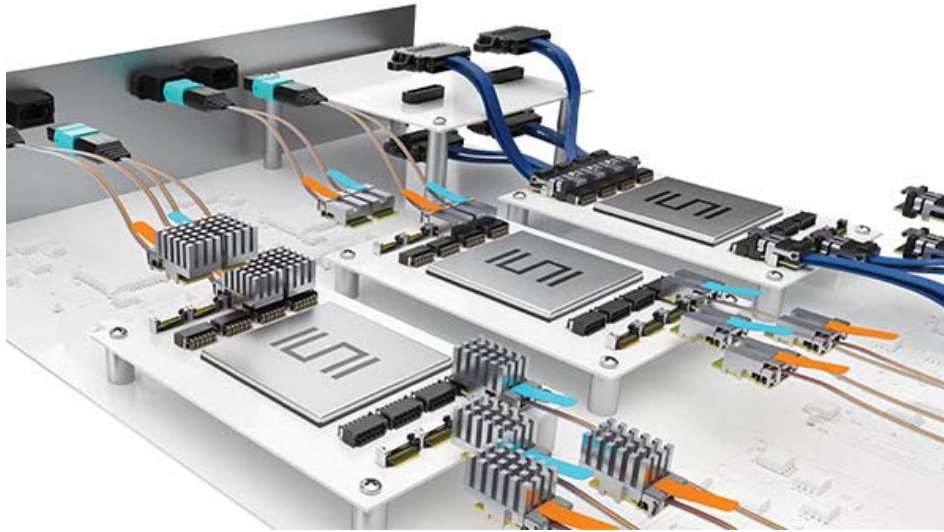
Figure 6: Electrical/optical "flyover" cabling [6]

2.4 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, $\tan\delta$=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 cross-talk. For via transitions, via-stub removal by using blind via or backdrilling is critical, and a smaller via diameter may be needed for via impedance control and cross-talk reduction. Further, signal conditioning and equalization will be widely adopted to compensate f or excessive loss, ISSI, and cross-talk. For data rates beyond 50Gbps, PAM4 signaling will prevail, for much lower Nyquist frequency.

2.5 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

2.6 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, Febuary 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] Pirzada, U., "Hot Chips 2017: PCI Express 4.0 Standard Coming In 2017 But Will Be Short-lived – PCIe 5.0 Landing in 2019", in Wccftech, Aug 29, 2017, available at: https://wccftech.com/pci-express-4-0-standard-coming-in-2017-but-willbe-short-lived-pcie-5-0-landing-in-2019/
[6] Samtec Micro FireFly Flyover System (samtec.com)

Section 3: Thermal Management

Introduction

Since there is a separate Thermal Management Technical Working Group (see Chapter 20), we shall make brief reference to the challenges and issues in this Single & Multichip Integration Chapter and refer the reader to that Chapter for the output of their work. The Thermal Management TWG focused on three key areas, namely the die level, the package integration/SIP/module level, and the system level (which is limited to the board level by definition, for the purpose of this TWG). The Thermal Management TWG has outlined the current trends and projections for cooling requirements, the currently available technical solutions, including solutions that require significant engineering advances but are considered tactical, and finally the major challenges and advanced strategic solutions that may require new discoveries and potentially significant changes in design, manufacturing infrastructure and deployment in the field.

Challenges for Single & Multichip Integration

Looking out into the future, there are a number of unique thermal management challenges that significantly differ from single chip packages in the past. One of the concerns is the difference in chip height above the substrate. For example, typically the HBM stacks and other components may be higher than the processor chip, necessitating a way to bridge the gaps between the cold plate or heatsink and the devices. Another unique issue in Heterogeneously Integrated Packages arises due to the multiple devices and packaged components that may be present within a package, from multiple manufacturers. This may cause the need to use more than one thermal interface material in the same package. Packages that have devices that are highly sensitive to thermal gradients, both spatially as well as temporally, such as optical devices, may require thermal isolation from the other devices in the package. Another possible unique consideration is that different devices may have different allowable operating conditions including maximum junction temperatures, maximum thermal gradients and different thermomechanical tolerances. All of this may significantly complicate heatsink design, TIM selection, and assembly processes, and limitations may arise. This also complicates the design of the interposer from a thermal perspective. The TWG examines the different available interposer solutions including Si, EMIB and glass from a thermal perspective.

Thermal solutions and challenges

Air cooling: Existing thermal solutions include the use of forced air cooling with advanced heat sink designs, plus advanced TIM solutions including thermal paste, thermal adhesives, graphite sheets, and solder. Vapor chambers and heat pipes enable improved spreading and can be optimized for specific applications. The use of advanced designs may allow for heat flux levels in the range of $100W/cm^2$ under ideal circumstances. Larger devices and advanced heat spreading strategies may increase that limit.

Liquid and 2-phase cooling: Indirect water cooling solutions may provide cooling solutions for devices in the range of $450W/cm^2$ or higher, under optimal conditions. Direct liquid cooling can achieve significantly higher heat flux levels by eliminating the resistances due to TIM and spreaders. However, these solutions require significant changes in the overall designs of the package, and possibly the server and the rack.

Section 4: Mechanical Requirements

Mechanical stresses stem from differential thermal expansion of different material within a package built-in during fabrication. Warpage is a manifestation of physical deformation and built-in stresses. Warpage engineering is thus an essential part of mechanical stress management.

Warpage Engineering and Stress Management

Warpage engineering has long been an integral part for 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 gets 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.
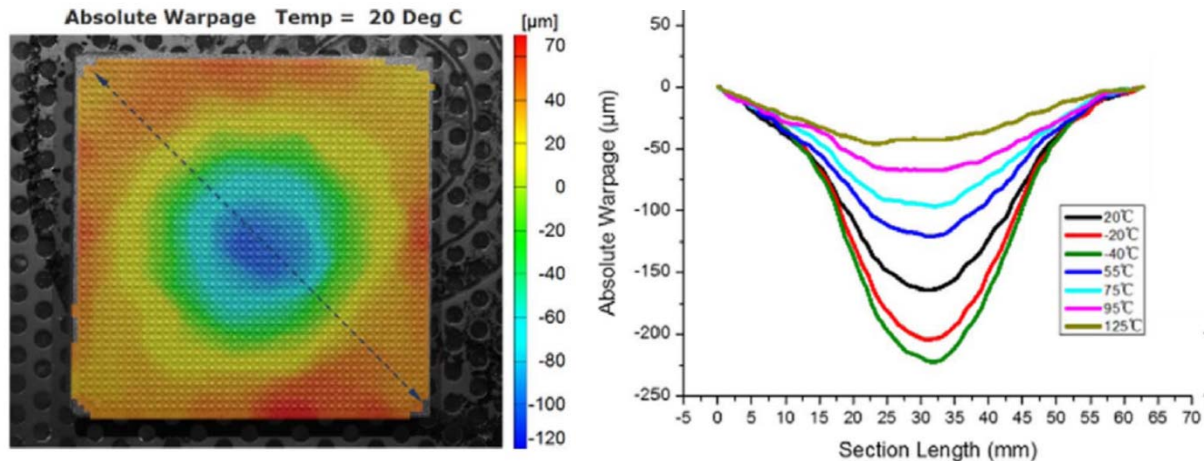
Figure 1. Thermal deformation (warpage) of a high-performance flip chip PBGA package (SB Park, Benson Chan)

For System-in-Packages (SiP) through Heterogeneous Integration – the characteristic 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 co-planarity 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 & simulation tools? What are the metrology tools?

Finite Element Analysis tools has 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 deformartion package

Warpage Metrology tools available today:
- Shadow Moire
- Projection Moire
- 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

Figure 2 is an example of a flip chip package which lies flat entering the reflow oven. As it exits the reflow oven, the part starts 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 in their warpage behavior and stress management. Figure 3 shows the assembly process for a 2.5D device; the package will undergo reflow 3 times with associated warpage phenomenon.
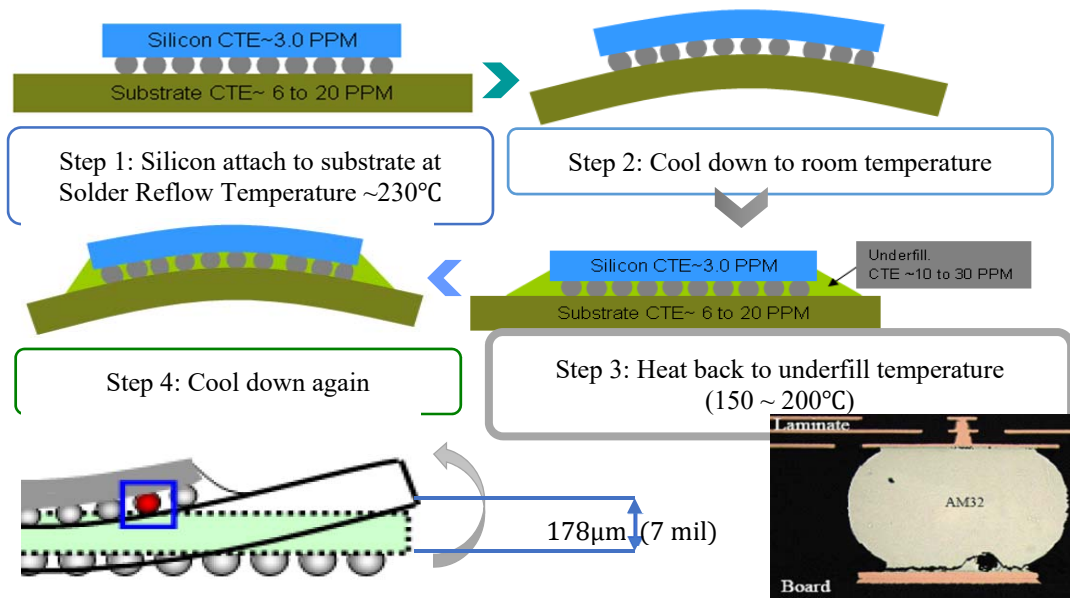
Figure 2. Warpage resulting from assembly processing (SB Park, Benson Chan)
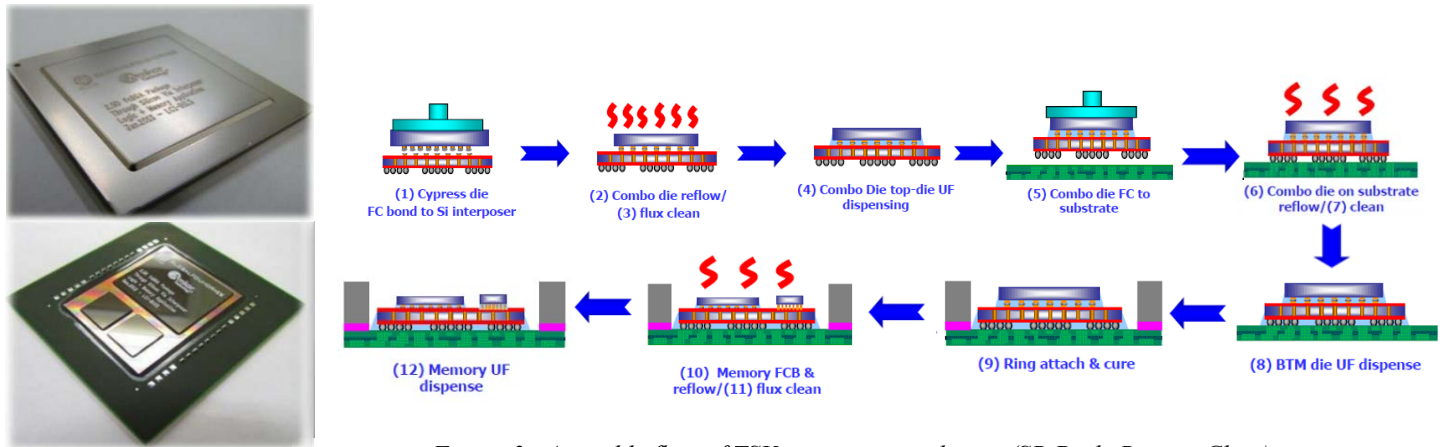
Figure 3. Assembly flow of TSV interposer package   (SB Park, Benson Chan)

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 (Figure 4) is created to show the movement of the points that were captured.
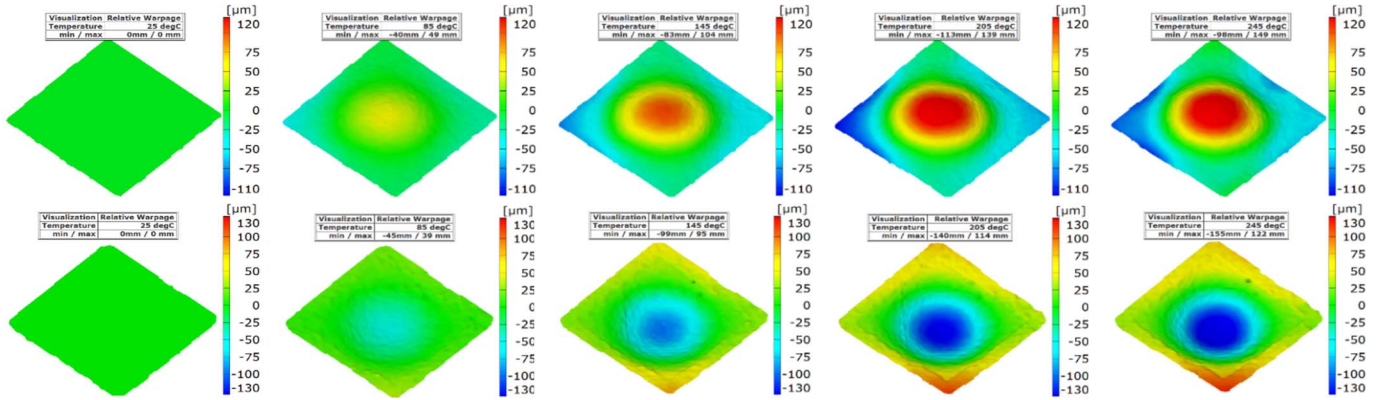
Figure 4. Warpage progression through thermal cycling (SB Park, Benson Chan)

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).
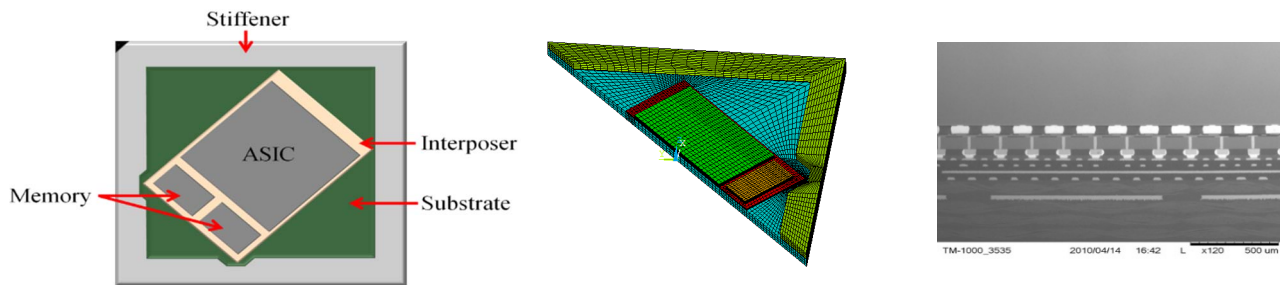
Figure 5. FEA model of 2.5D package and micro bump cross section (SB Park, Benson Chan)

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 below 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.
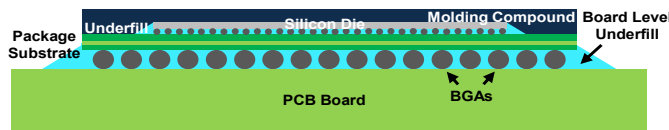
Figure 6. Structure of a FCCSP Package (SB Park, Benson Chan)
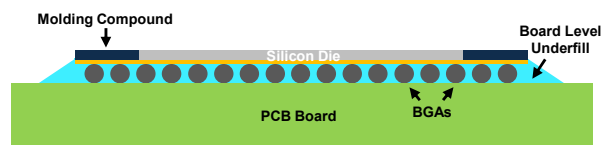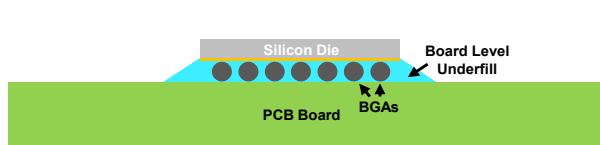
Figure 7. Structure of a WLP Package (SB Park, Benson Chan)   Figure 8. Structure of a FOWLP Package (SB Park, Benson Chan)

Table 2 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 80mm x 80mm to 90mm x 90mm 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 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 2: Warpage Allowance across two market segments

Segment	Year of Production / Pitch (mm)	2018	2019	2020	2023	2026	2029	2030
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
HPC	1.0	-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
HPC	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
HPC	0.8	-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
HPC	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
HPC	0.65	-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
Mobile	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
Mobile	0.5	-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
Mobile	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
Mobile	0.4	-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
Mobile	0.3	-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
Mobile	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
Mobile	0.25			-0.055, +0.055	-0.055, +0.055	-0.05, +0.05	-0.045,+0.045	-0.045,+0.045
Mobile	0.2			-0.055, +0.055	-0.055, +0.055	-0.05, +0.05	-0.045,+0.045	-0.045,+0.045
Mobile	0.2				-0.045,+0.045	-0.045,+0.045	-0.045,+0.045	-0.045,+0.045
Mobile	0.15				-0.045,+0.045	-0.045,+0.045	-0.045,+0.045	-0.045,+0.045
Mobile	0.15				-0.025,+0.025	-0.025,+0.025	-0.025,+0.025	-0.025,+0.025
Mobile	0.1				-0.025,+0.025	-0.025,+0.025	-0.025,+0.025	-0.025,+0.025
Mobile	0.1					-0.020,+0.020	-0.020,+0.020	-0.020,+0.020

Manufacturable solutions exist, and are being optimized

Manufacturable solutions are known

Interim solutions are known

Manufacturable solutions are NOT known

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.

Future Challenges & Opportunities

With artificial intelligence and development of advanced DIC-like 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: Wafer Thinning and Singulation

1. Introduction

All silicon wafers start out at the foundry at between 0.7 to 0.8mm in thickness. To fit inside advanced low-profile single and multichip packages, wafers are thinned by processes where the wafers are ground on the backside using abrasive rotary grinding and polishing wheels. Thinning is a well-established process which can be highly controlled to produce ultrathin wafers down to below 10μm and even below 5μm. Smaller dies like microLEDs can achieve 2 to 5um in thickness; however, "thinness" is limited by area of the die. Currently volume production has been established at 30μm for wirebond dies and 50μm for flipchip dies.

2. Scope of thinning and singulation

Ultrathin dies at 30μm and less have been driven by the homogenous die-stacking memory companies so they can integrate more gigabytes. A 0.6mm profile package can accommodate a 10-die stack if the thickness of the die is 25μm, as shown in Figure 1. For 5μm die with 5μm die-attach film (DAF), a total stack of 50 is possible within a 0.6mm package.
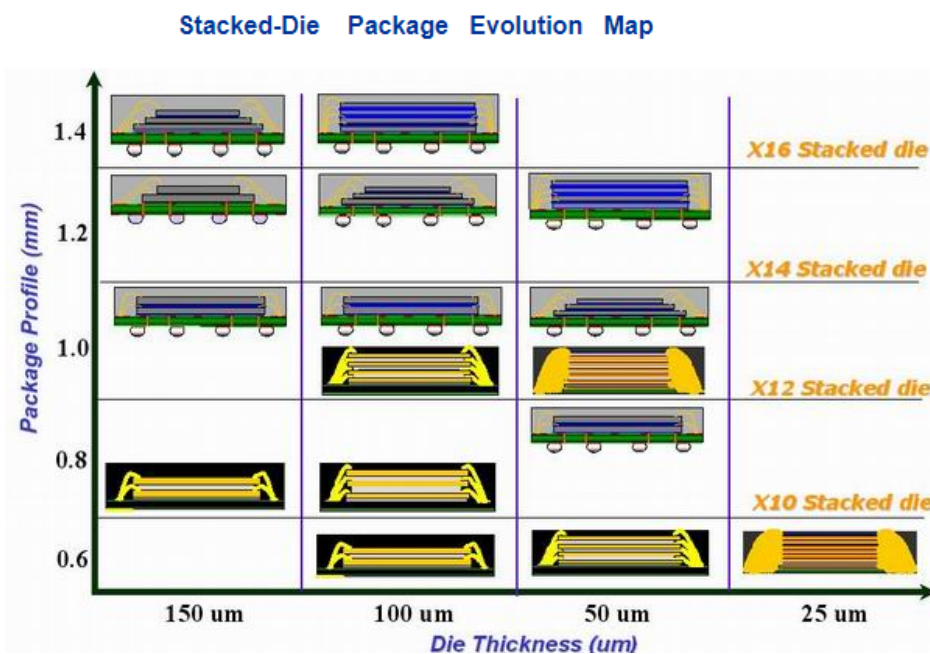
Figure 1. Die thickness vs die count of stacked packages. (Source: King Pak)

Another factor driving the current packaging trend towards ultrathin dies is the wearables market. Going below 25μm thickness allows non-planar flexibility, where dies can bend inside wearable devices.

3. Difficult Challenges: Singulation of Ultrathin die that will not crack

The challenge has been to keep the ultrathin silicon die strong without cracking and without affecting device performance. Die curling and cracking must be resolved through extensive process controls and equipment refinement as we push towards ultrathin dies. Reduction of process temperatures and good clamping to temporary carriers are key. Plasma etching using the Bosch technique is capable of producing silicon dies with no flaws on the 4 sidewalls, and the dies are therefore stress-free to flex. The Bosch process has provided payback for diode and RFID wafers where there are thousands of streets to cut. A batch process using plasma to singulate can lower time and cost for processing wafers, whereas conventional one-cut-at-a-time processes may take hours for a 300mm wafer with >1 million dies. An alternative method which can possibly lower cost is using a wet technique to form deep vertical trenches of 100 μm in depth for chemical dicing, using hydrogen peroxide and hydroflouric acid.

Figure 2. Bendable 15μm Si (from University of Glasgow)

3.1 Reduction in street and guard ring geometries

With plasma dicing, there is less of a need for a robust seal ring at the perimeter of every die. It is possible for the seal ring to be eliminated, thus freeing up more silicon for increased yield and cost reduction. Reducing street width (currently at 100μm for most wafers) to below 50μm will increase greatly the die per wafer and productivity. Street width is dependent on wafer thickness during dicing. In fact, ultrathin wafer street width can be designed to be the same as the wafer thickness during dicing: i.e. street width equals die thickness. Hence, it is important for design engineers to work closely with package and manufacturing engineers to make decisions resulting in the most efficient and effective use of the silicon real estate.

Geometries for dies and films are trending thinner and narrower alongside trends for interconnect bump technology. New processes for ultrathin wafers allow street width between dies on a wafer and seal ring width around each die to be significantly reduced. There will likely be a convergence of dimensions to <5 μm in the years ahead for thinness of dies, seal ring border width, street width, thinness of adherent films and various interconnect joint diameters.

3.2 Support and Pick Challenges of ultrathin dies

With thinner dies, infrastructure to support these dies includes support tapes and temporary wafers with temporary adhesives. The development, implementation and cost of this infrastructure for handling ultrathin die is quite high. New ways to simplify the handling of ultrathin wafers with fewer steps and avoidance of temporary support wafers would be ideal. One such process that American Semiconductor, Inc. has been offering is FleX Silicon-on-polymer, sandwiching a 12μm thin wafer in between polyimide.
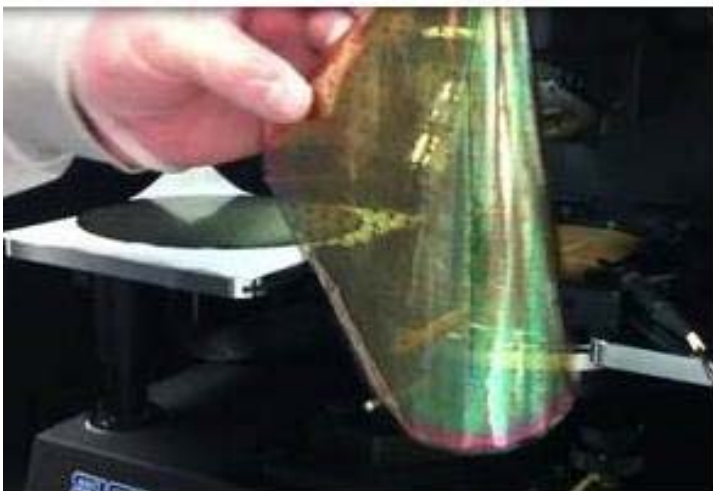
Figure 3: 200mm diameter wafer supported by polyimide (from American Semiconductor, Inc.)

One of the riskiest assembly process steps is when the ultrathin die is removed from its support media for assembly to single or multichip modules. The traditional one-size-fits-all tooling is been replaced by highly customized tooling

for peeling die with decreasing thicknesses. Vacuum is installed in pick and place machines now to facilitate the peeling (pulling) of the tape from the die, such as the system used on Datacon's tool. [1]
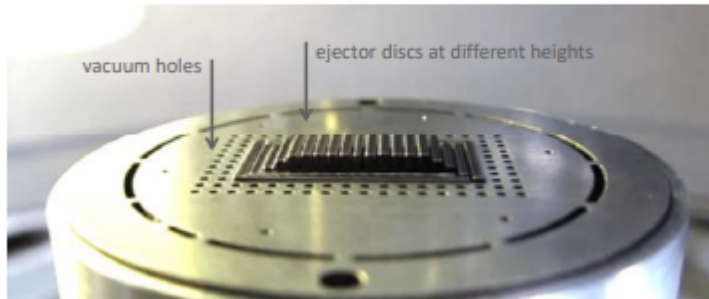
Figure 4. Peeling fixture for flexible dies (from Datacon)

This area could use some new ideas that allow for a less mechanically based transfer. A potential technology is Microassembly Printing from research done by Eugene Chow[2] of Xerox, PARC. The Microassembler Printer uses electrophoretic and dielectrophoretic forces of the die to sort, orient and align small dies (chiplets) into an array for batch transfer and packaging. For now, the microassembly printing works well on thin and small dies such as diodes. Thin and fragile dies are not damaged, as there is no contact with the pick-up or ejector tools. Figure 5 summarizes the batch process of picking and placing.

	Throughput	Accuracy	Heterogeneous	Programmable
Xerography	☑	☒	☑	☑
Pick & Place	☒	☒	☑	☑
Atomic Force Microscope	☒	☑	☑	☑
EUV Lithography	☑	☑	☒	☒
Chemical Self Assembly	☑	☑	☒	☒
Microassembly Printer Goals	☑	☑	☑	☑

Figure 5 [2]

4. Combining DAF Dicing with Die Singulation

For wire-bonded single and multichip, die attach films (DAF) are essential, unlike for bumped chips. The DAF is both the adherent material for the die as well as the support film. Wafers with DAF are most easily singulated using the saw method; however, with the divergence away from sawing, wafer film that is resistant to plasma and other etchants needs to become available. DAF films with 5 to 25µm thickness are commercially available.[3] They are also available in silver-filled sheets for heat conduction. Their thermal conductive properties are somewhat limited for now and need to be improved for conductive DAF to be more widely accepted. Breakthroughs are needed to produce DAF with higher thermal conductivity, such as the use of alternative fillers.
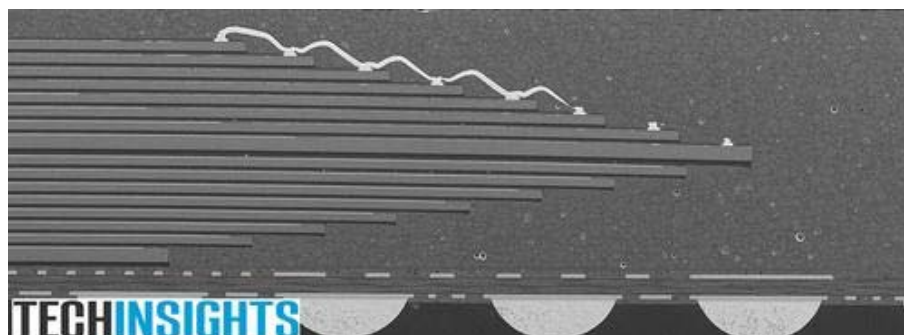
Figure 6: Samsung K9UGB8S7M 48L V-NAND Flash memory showing dies interlayered with DAF.

References:

1. Stefan Behler, "How to Peel Ultra-Thin Dies from Wafer Tape", EPS Santa Clara Valley Technical Talk, Dec 7, 2017
    (www.cpmt.org/scv)
2. Eugene Chow et al, "Chiplet Micro-Assembly Printer", IEEE 69th ECTC, May 28-31, 2019.
3. Yasuko Ferris, Nitto, Inc. (private communication)

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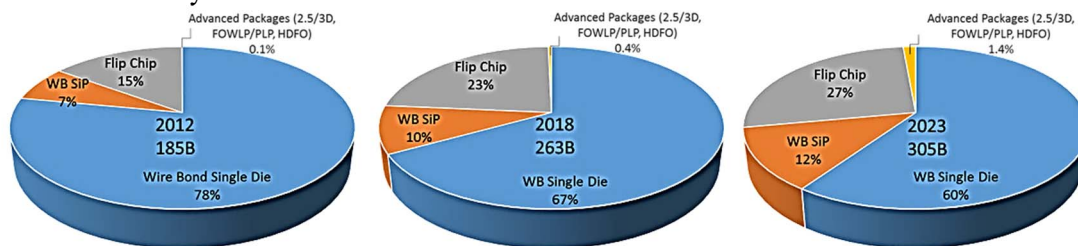
Figure 1: Semiconductor package unit growth by interconnect type. IC Package Shipment excluding Discrete, LED and Opto. Source: Prismark May 2019.

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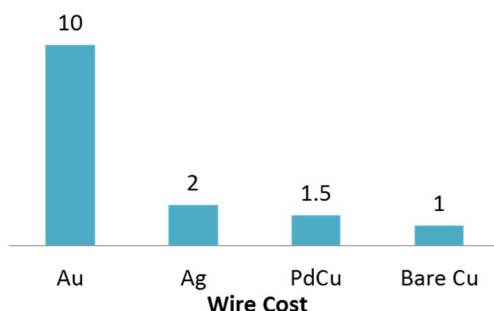
Figure 2: Relative cost for different type of wires, based on 25um diameter wire (Source: Heraeus and Tanaka). 94% Ag alloy is used in the comparison.

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, the 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].
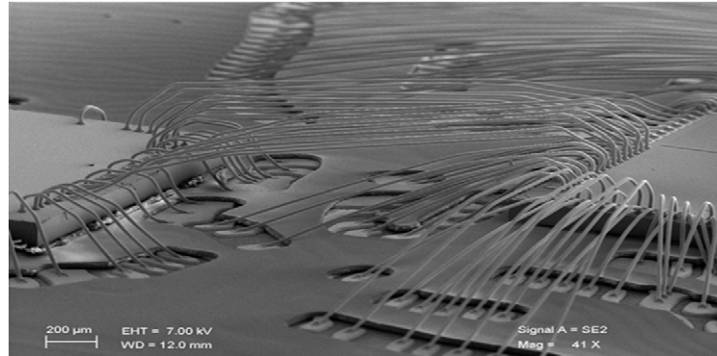
Figure 3. Advanced multiple tier looping and die-to-die bonding loops (K&S)
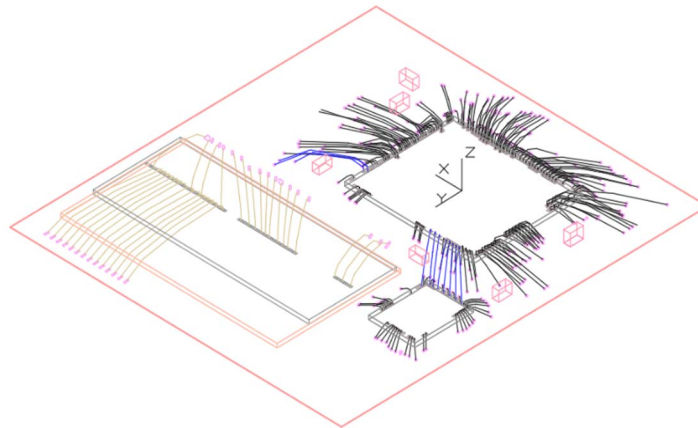
Figure 4. 3D loop design and clearance check software (K&S)

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].

Multiple-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]. A close-up of a SSB bump is shown in Figure 7. The industry is driving toward finer SSB pitch capability.
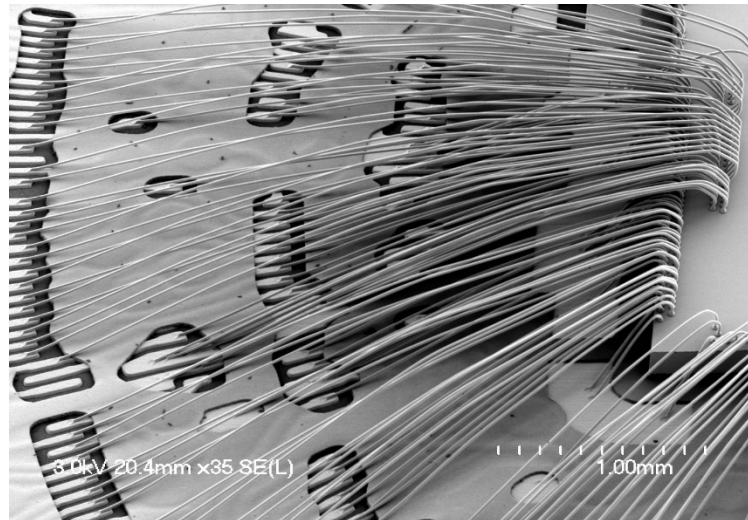
Figure 5. High density multi-tier package (K&S)

*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)*
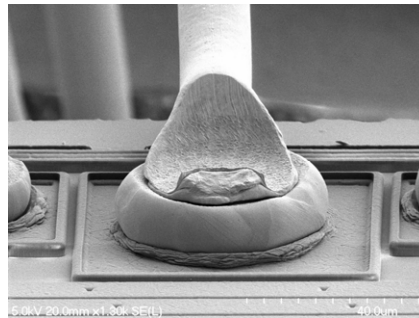
Figure 7. Stand-Off-Stitch Bond (SSB) process is more challenging than forward bond reduced fine pitch capability (K&S)

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]. The two commonly used stacking methods are vertically stacking devices in a single package, and vertically stacking multiple packages. An example of the first method is a memory device that stacked 4, 8 and 16 dies as shown in Figure 8. Higher number of stacked die such as 32 and 64 stacks are in the R&D phase and low volume production.
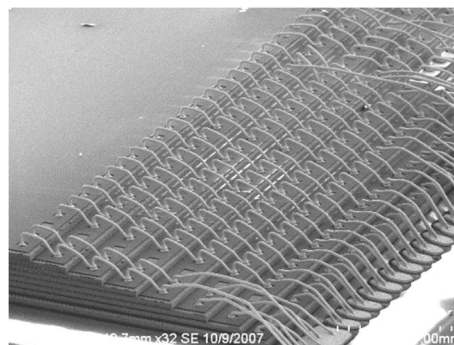
Figure 8. Example of a stacked-memory device (K&S)

In stack die packages, an 'overhang' configuration is very common. In overhang configurations, one or more dies may be unsupported (Figure 9). 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.
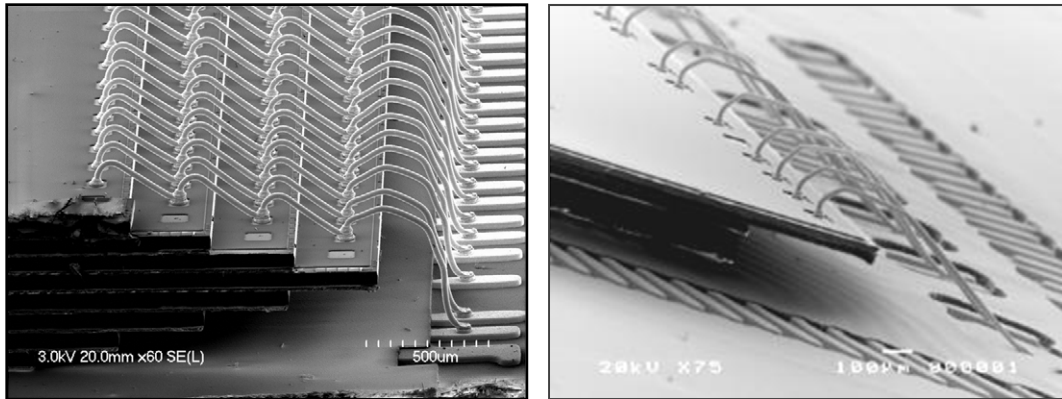
Figure 9. Examples of bonding on overhang die (K&S)

In order to achieve overall package height requirements for multi-stack memory packages, low loop heights of 100 um or lower 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.
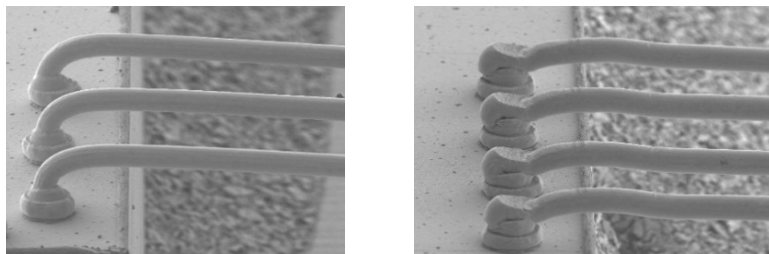
Normal Loop   Compressed loop
Figure 10. Examples of low loop bonding with Ag wire using a normal loop and a compressed loop (K&S)

Many breakthroughs in packaging solutions were enabled recently through Fan-out Wafer Level Packaging (FOWLP). FOWLP has several advantages such as a smaller form factor and thinner packages [13]. One such assembly process flow was demonstrated by the Institute of Microelectronics (IME) in a Fan-out Wafer Level PoP package achieving ~2400 I/O counts [14]. A thick photoresist lithography process for copper pillar Through-mold-interconnection (TMI) formation is normally used. This tall Cu pillar process is challenging and costly for high density and fine interconnect requirements. A simplified and 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 11).
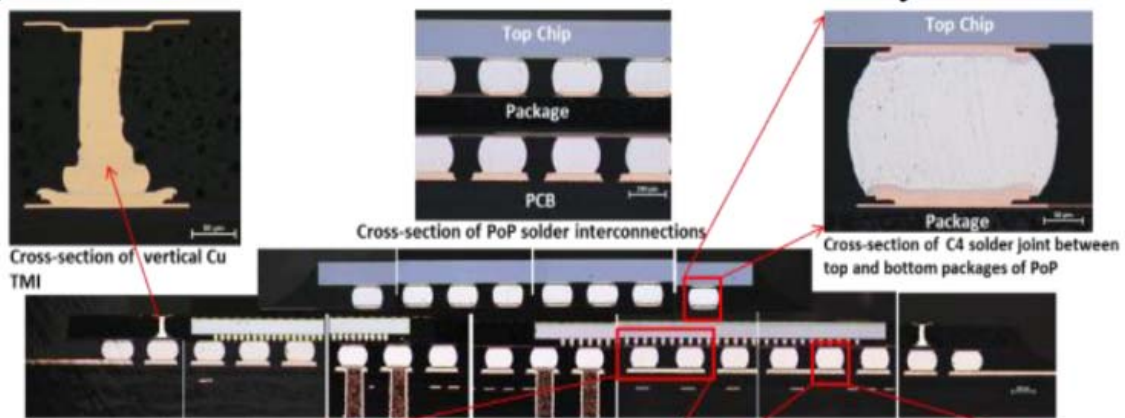
Figure 11. Cross section of assembled FOWLP package with wire bonded Vertical Interconnects [15] (Source: IME).

Future Challenges and Roadmap towards 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 [15]. 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 [16]. 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.

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 Shringking 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] T. Braun, M. Töpper, K.-F. Becker, M. Wilke, M. Huhn, U. Maass, I. Ndip, R. Aschenbrenner, K.-D. Lang, "Opportunities of Fan-out Wafer Level Packaging (FOWLP) for RF applications", IEEE 16th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF), 2016

[14] 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 Vagas, NV, Jun 2016, pp. 1522-29

[15] 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.

[16] I. Qin et al., "Smart Wire Bonding Processes for Smart Factories", Electronic Packaging Technology Conference, December, 2018

Section 7: Flip Chip and Interconnects

1. Introduction

Flip Chip continues to grow across multiple package platforms as higher package pin counts and performance are needed (See Figure 1). Today's applications are driving unique ways of organizing advanced semiconductor system solutions. System on a Chip (SOC), which was once the preferred design layout option, has moved over toward best of class silicon for best of class cost/performance. The alternative to SOC is to look for ways to heterogeneously integrate key portions of the design into a close side-by-side configuration that can leverage optimized silicon nodes for portions of the design. As part of heterogeneous integration, the interconnect of the die also become critical because the inductance of the lines, connecting each subsystem, can significantly impact performance. Interconnect pitch, size, metallurgy and even substrate technology can play a critical role in ensuring that a design can meet necessary high-performance standards. This section will concentrate on the key challenges and considerations for flip chip when looking how best to leverage heterogeneous integration.

Bn Units	2016	2017	2022F	2017/ 2016	CAAGR 2017-2022
Flip Chip QFN/MIS	0.6	0.8	3.3	33%	33%
Flip Chip SiP	3.0	3.4	7.0	13%	16%
Flip Chip CSP	3.4	4.4	6.0	29%	6.4%
Flip Chip CSP for DRAM	3.0	3.3	8.0	10%	19%
Flip Chip BGA/PGA/LGA	0.9	1.0	1.3	11%	5.4%
WLCSP	28	31	43	10%	7%
FO-WLP	0.8	0.7	2.9	-13%	33%
COF/COG	8.0	9.0	11	13%	4.1%
Subtotal Flip Chip	**47.4**	**53**	**82**	**12%**	**9.1%**
Subtotal Wire Bond	172.9	192	208	11%	1.6%
IC TOTAL	*220.3*	*245*	*290*	*11%*	*3.4%*

Figure 1: Flip Chip IC Package Unit Growth (Prismark 2018)

2. Scope for Flip Chip

The Flip Chip interconnect technology extends across multiple bump types and package platforms. Solder bump, Copper Pillar and Gold stud have been the primary flip chip package interconnects but evolving interconnect options such as direct copper to copper interfaces using copper to copper bumps or copper nano-paste are being developed to help address pitch, electro-migration and reliability issues. The interconnect decision will depend on the bump pitch, power, speed/frequency requirements and die size. Figure 2 is a chart that shows the main types of flip chip interconnects.

Bonding Method	C4 FC (Contolled Collapse Chip Connect)	C2 FC (Chip Connect)	TC/LR (Local Reflow) FC	TC FC
Schematic Diagram				
Major Bump Pitch Range at Application	> 130 um	140 um ~ 60 um	80 um ~ 20 um	< 30 um
Bonding Method	Conventional Reflow	Reflow with Cu pillar	Thermal Compression with Cu pillar	Thermal Compression
Bump Metallurgy	Solder (SnAg or SnAgCu)	Cu + Solder (SnAg or Sn)	Cu + Solder (SnAg or Sn) Cap	Cu
Bump Collapse	Yes	No	No	No
Underfill Method	- Capillary - No flow	- Capillary - No flow - Wafer Level	- No flow - Wafer Level	- No flow - Wafer Level

Figure 2: Common Flip Chip Interconnect Options (Provided by Atotech)

Cu Nano paste, and similar technologies, are not mainstream but are in development and are to be watched in the coming years as companies look to improve reliability of the interconnect by eliminating solder.
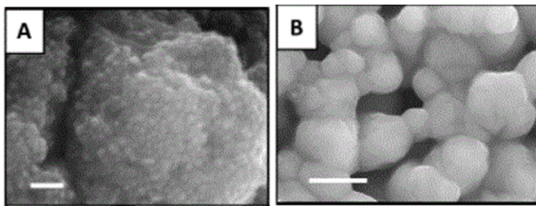
Figure 3: SEM micrographs of copper nanoparticles isolated from the synthesis (A) and material after fusing at 210°C for4 minutes (B). The scale bars represent 100 nm and 500 nm for panels A and B respectively (Lockheed Martin)
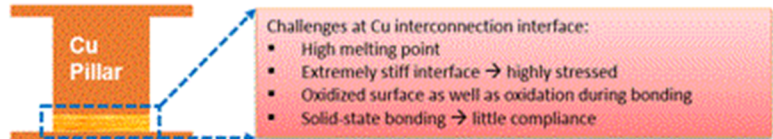
Figure 4: Direct Cu to Cu Challenges (Georgia Tech)

3. Difficult Challenges

Challenges fort flip chip depend on the application which drives the package platform. For smaller body mobile type applications, x, y and z height are critical and because of the size requirements, the bump pitch has been pushed harder than for larger body applications. In addition, the printed wiring boards that are used for mobile applications are getting more aggressive in their line/space/via requirements and most are now using IC substrate technology to enable their designs.
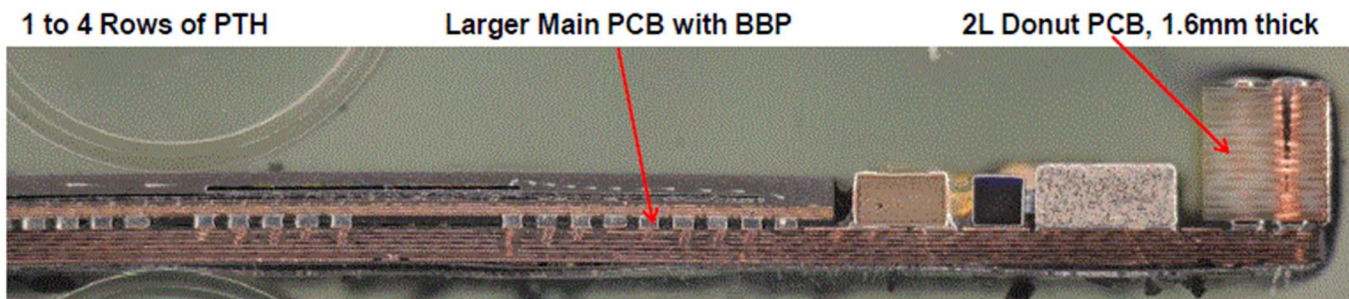
Photos source: Prismark/Binghamton University

Figure 5: Cross Section of iPhone X showing PCB and subsystem.

Flip Chip pad pitches for the smaller CSP type packages (<23mm) can leverage design rules much lower than the traditional solder flip chip technology. Most of these use copper pillar and may have larger center array pitches for power and ground but the periphery pitches can range from 150um down to 40/80um staggered or 50um inline. Underfill challenges exist with these packages mainly centered around the keep out zone areas around the die. The underfill fillet will typically bleed out and can impact other device pads. For this FCCSP space, both oval shaped and round shaped copper pillars are used and sometimes even mixed. Oval shapes can be beneficial when you need more current carrying area in the bump, more solder volume for stress reduction at the die level and to help with routing when escape routing is challenging. As pad pitches continue to reduce, future challenges will include addressing heat uniformity across the die during die bond, die warpage for mass reflow, and ultra low K dielectric stress sensitivities. In addition, the substrate technology will also play a key role and will also need to ensure ultra-flat surfaces during assembly to ensure a uniform solder joint. Die bond alignment at finer pitches, such as 10um, will also be a key challenge focus area. For flip chip die, there is a cost tradeoff with high accuracy on most bond equipment and the tighter the pitch, the slower the die bond alignment process.

In this smaller body FCCSP space, key considerations of the interconnect system can also influence high speed/freq applications. For sub 7Ghz applications like 5G, there are multiple package platforms that will need to be optimized to minimize electrical/thermal losses and improve performance. This challenge becomes even more difficult when you move into the mmwave space >24Ghz. For 5G mobile, complex antenna designs and substrate layer stacks play into Flip Chip modules while also having to consider shielding requirements for close proximity interference. Considerations such as high speed frequency skin effect can also drive the needs for good etching control of the copper lines in a substrate. In other applications, such as 77Ghz radar for automotive, come to the market, more smoother lines and optimized interconnect systems will need to be developed. Today there is a good amount of trade-off analysis done between the fan out platform (RDL Process Based) and flip chip packages (usually organic coreless) when considering these higher speed applications for smaller body sizes. A few of the key drivers for the analysis between these two technologies is the thin dielectric layers with lower loss through vias, smooth line surface and predictable cross section features and the trade-offs between a solder-based interconnect and direct cu to cu.

For applications related to artificial intelligence (AI), most industry leaders think of 2.5D as the main platform for this solution due to the high bandwidth memory (HBM) integration and complex ASICs but more and more companies are integrating AI capabilities into their chips, with some key mobile industry players making announcements this year. For the more complex 2.5D large body package, the die to die pitches and density of interconnects have been key enablers and have driven some new capabilities. For CSP, some of these lessons learned may also be leveraged but on a smaller scale. For deep learning applications, the use of wide-IO HBM type memories could drive much higher interconnect densities in a smaller body package. Today, 2.5D HBM memory stacks have IO counts around 5000 due to the multi-die TSV stack configuration. The nature of deep learning will drive the need for high density memory if the system does not rely on a server/internet-based data storage system. This could drive much higher densities for smaller body/CSP packages which could help accelerate flip chip pitch and substrate line/space/via roadmaps.
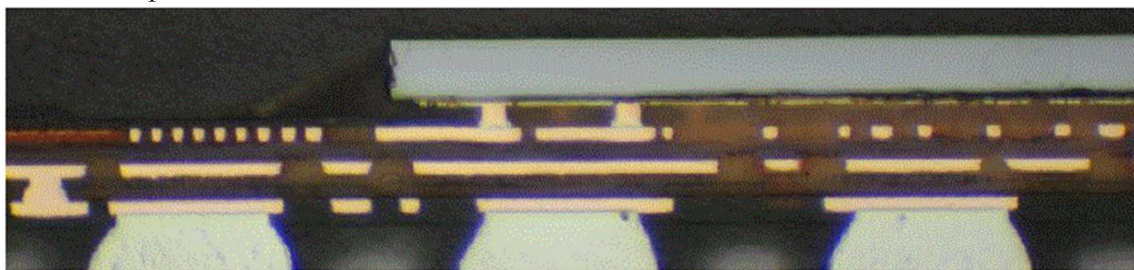
Figure 6: 3 Layer Coreless Flip Chip CSP with Copper Pillar

For larger body Flip Chip devices, a number of new challenges are emerging as heterogeneous integration being considered for new devices. Heterogeneous and Homogeneous integration are both adding new design rules to the mature Flip Chip MCM package platform. Underfill design rules, specifically die to die spacing is critical when trying to minimize inductance between 2 or more chips. Companies looking to break out SERDES blocks are being driving more advanced design rules to allow very close spacing (<70um) between these chiplets and the larger ASIC. In addition to the die to die spacing, keep out zones for discrete components are also driving new component spacing rules. In addition to the placement consideration, how to handle thermal challenges is a growing. Functional blocks

that have been broken off the main die or separate die/components may have different power characteristics and the heat dissipation challenges can be a major design factor. For many of the current 2.5D solutions, a design that allows for all die in the system to be planar with the surface has been a requirement. Whether this is achieved by molding the system and then planarizing or designing the die/component heights to be the same, a planar surface for heat dissipation through the top heat spreader is critical while managing the mechanical stresses associated with new die system.

Substrate technology is key enabler to the more advanced Flip Chip package solutions. Integrating jumper chips with finer lines/space and looking at new ways to solve denser routing requirements is accelerating and will need a lot of industry support. As the density continues to increase, substrate technologies will move closer toward the fab based process where higher end equipment and similar clean room environments will be needed for yield control. Sputter based metallization is already a key enabler for coreless and advanced substrates and other benefits have extended many substrate roadmaps. When interchanging different material stacks, metallization's and even integrating jumper silicon/glass, modeling of subsystems will become increasingly important as part of the larger system.

Because there are so many complex tradeoffs between electrical, mechanical and thermal in these HI systems, early modeling analysis for co-design is critical. Component location and orientation in addition to material selections can be a challenges task. Figure 7 shows some of these considerations for a complex 2.5D Flip Chip package.
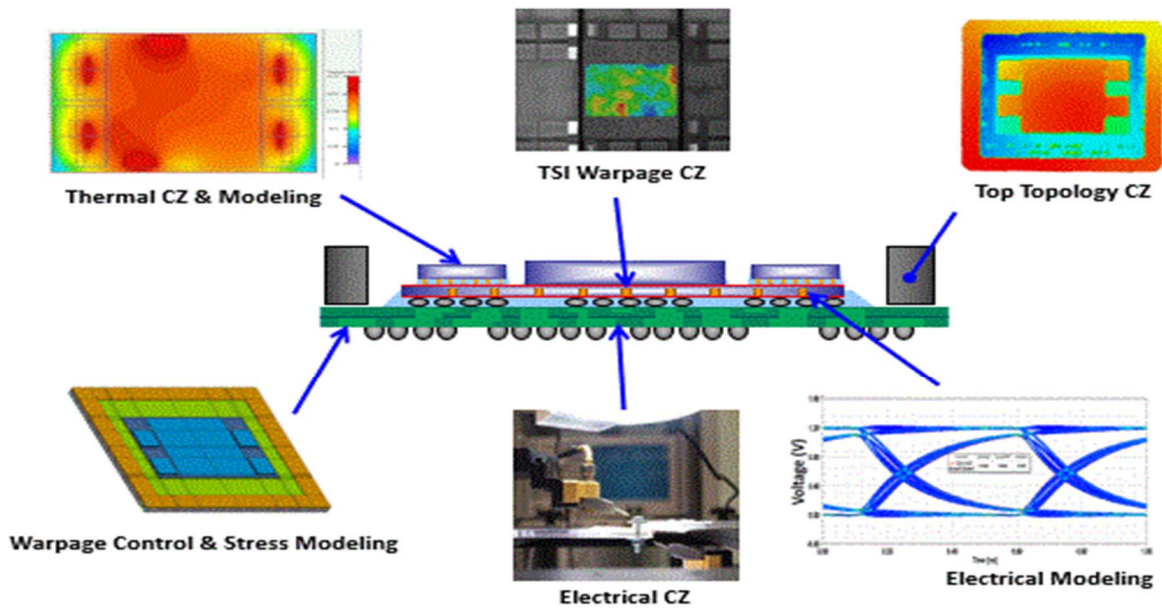
Figure 7: Modeling and Characterization key considerations for High End Flip Chip (ASE Group)

4. Conclusions

Flip Chip continues to be a vital package platform for the semiconductor industry and as the cost and yield challenges continue to grow with advanced silicon node technologies, heterogeneous integration packaging solutions will rely heavily on advanced flip chip to enable effective solutions. All the key components of flip chip will be challenged in the coming years, including bump type, pitch, substrate, underfill and thermal capabilities which will drive new development activity to enable a scalable long-term roadmap.

5. References

[1] A. A. Zinn, R. M. Stoltenberg, J. Beddow, and J. Chang, "NanoCopper Based Solder-free Electronic Assembly Material" IPC Proceedings, pp.4-5

[2] Frank Kuechenmeister, Dirk Breuer, Holm Geisler, Jens Paul, Chirag Shah, Kashi Vishwanath Machani, Sven Kosgalwies, Rahul Agarwal, Shan Gao, "Chip-Package Interaction: Challenges and Solutions to Mechanical Stability of Back End of Line at 28nm Node and Beyond for Advanced Flip Chip Application"14thEPTC 2012

[3] Dirk Breuer, C. S. Premachandran, Matthias U. Lehr, Frank Kuechenmeister, Holm Geisler, Michael Hecker, Jens Paul, Chirag Shah, Kashi Vishwanath Machani, Sven Kosgalwies, Rahul Agarwal, Shan Gao, Lothar Lehmann, "Chip Package Interaction"; IEEE International Reliability Physics SymposiumIRPS2014; CPI Tutorial

[4] M. Madsen, K. Takei, R. Kapadia, H. Fang, H. Ko, T. Takahashi, et al., "Nanoscale semiconductor "X" on substrate "Y"– processes, devices, and applications," Advanced Materials, vol. 23, pp. 3115-3127, 2011.

[5] B. K. Appelt, H. Chung, C. Chienfan, and R. Wang, "Fine pitch flip chip chip scale packaging," Microelectron. Packag. Conf. (EMPC), 2011 18th Eur., pp. 1–4, 2011.

[6] S. W. Yoon, D. W. Yang, J. H. Koo, M. Padmanathan, and F. Carson, "3D TSV Processes and its Assembly / Packaging Technology," Packag. (Boston, Mass.), 2009.

[7] B. Schmaltz, "Packaging Materials for 2.5/3D Technology," 15th Int. Symp. Exhib. Adv. Packag. Mater. (APM 2013), vol. 2, no. Paper 8, pp. 93–101, 2013.

Table 1: Flip chip interconnects

Year of Production	2018	2019	2020	2021	2022	2025	2028	2031	2034
Flip Chip Pitch									
Flip Chip- Large Body Solder >12mm Sq Die	135	130	130	130	130	130	130	130	130
Flip Chip- Small Body Solder <12mm Sq Die	135	130	130	130	130	130	130	130	130
Flip Chip - Cu Pillar Small Body <12mm Sq Die (Bond on Trace)	45	30	30	30	30	20	20	20	20
Flip Chip- Cu Pillar Large Body >12mm Sq Die	110	100	100	100	90	90	90	90	90
Flip Chip Solder - COW	50	50	50	50	50	50	50	50	50
Flip Chip Cu Pillar -COW (Chiplets)	40	35	35	35	30	25	25	25	25
Flip Chip Cu Nano Particles	30	30	30	30	30	30	30	30	30
Wafer to Wafer Cu to Cu Interconnect	5	5	5	2	2	2	1	1	1
Embedded Die In Substrate Interconnect Pitch	120	120	120	120	90	90	70	70	70

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

Notes for Table 1:

1. For very fine pitch chip to package bonding, alternative technologies such as TSV and bumpless interconnect will be used as alternatives to technologies in this table.
2. Finer pitch is technically possible for most categories but does not meet cost constraints.
3. Pitch for emerging connection technologies is not included in this table. See chapter text.
4. Flip chip array, low end & consumer and mobile products are too small to meet the requirements of low cost FCBGA substrate.
5. Area array pitch should be ca. 200um to use low cost FCBGA substrate.
6. Chip on film is a new package type that is used as a replacement for TAB in some cases.

Section 8: Substrates

Scope

In this substrate section we shall focus on substrates used as key components in four market application areas: (a) High Performance Computing, (b) Mobile & 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 and 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, especially for the emerging chiplet applications. This demands significant innovation for interconnect architecture, substrate materials and process development.

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

Table1. Substrate interconnect scale roadmap (micrometers)

Since high-performance-computing die are reaching and exceeding full reticle size, disagregating into multiple die with a mix of advanced and later nodes and perhaps different foundries may proving to be an effective economical and realistic option. Bump pitch can scale more aggressively for smaller dies, since bump coplanarity and shrinkage scaling is much less than for a full-reticle die.

Leading-edge FC-BGA build-up substrates at 14/14μm L/S are in production. Advanced laminate for EMIB is reported at 9/12μm L/S.[1]. Further line and space scaling will require a very smooth Cu surface, with roughness R_q below 100nm. Potential solutions would require thinner seed metals below 0.5μm, 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 R_q, 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 40μm using a 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), photoimagable 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 70μm bump pitch, below which electroplating technology is expected to be a potential solution. One of the challenges in fine-pitch bumping is to enable mixed bump pitch with different pad sizes and uniform bump height.

With 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 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 (see chapter 17 on Test).

As power density increases for high-performance computing applications and the vertical interconnect feature size scales, the demand for current-carrying capability through 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.

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.

	2019	2023	2027
Bandwidth – Package to board (Gbps)	45	55	65
BU dielectrics loss (D_f)	0.007	0.004	0.002
BU dielectric roughness (R_q)	300~400 nm	150~200 nm	100~150 nm
Cu roughness (R_q)	350~400 nm	200~250 nm	50~100 nm

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

Embedding passive components inside the package is an efficient concept for improving power delivery due to the shorter interconnect length between passive components and the die, compared to discrete components attached on the land side or die side. Along with the development of high-efficiency passive components, it is essential to develop an innovative way to embed the larger number of passive components (capacitors and inductors) inside the package.

Mobile and Wearables

The Smart Phone has become the major consumer electronics product across the globe. According to International Monetary Fund (IMF) [2], 1 billion units were sold in 2017, one for every fifth person on earth. The main trends from smart phone consumers are the continuous demand for more usability, longer battery life, and affordability. At the same time, the smart phone 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 smart phone product generation to the next, the product design and release cycle would be contained within each of these functional SiPs. Smart phones today may have twenty or more System in Packages. 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 in a single PoP package. PoP approaches commonly have the baseband or application processor assembled in FC-CSP format, with 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 node, where the wafer yield needs to be closely watched as the foundry product ramps up.

For FC-CSP in the PoP application, 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 to substrate manufacturers. The major challenges are (a) it needs to increase I/O density by decreasing the line and space below 15/15μm L/S (even 10/10μm) and with a blind via diameter below 60 μm. The second challenge is package warpage control for robust package-to-package assembly.

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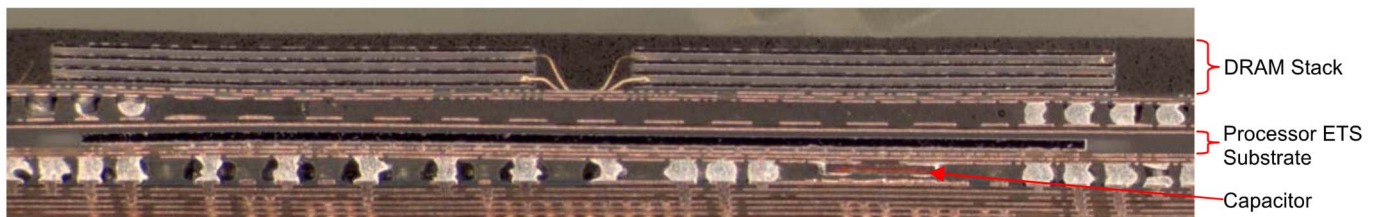

Photo source: Prismark/Binghamton University

Figure 1 PoP cross-section for Qualcomm 855 processor in Samsung Galaxy 10 Smart Phone (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 mother board to die. For in-depth review of the wafer-level and panel-level fan-out roadmap, please refer to WLP – Fan-in and Fan Out, Chapter 23.

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 2. Cross-section of Apple iPhone A12 processor (source Prismark Partners)

APPLE A12 – RDL

- Three-layer RDL fan-out
  - Minimum 8μm L/S
- Additional metal layer plated before initial dicing
- Total of five metal layers
  - 5 – 7μm thick
- Five dielectric layers
  - 5 – 6μm thick
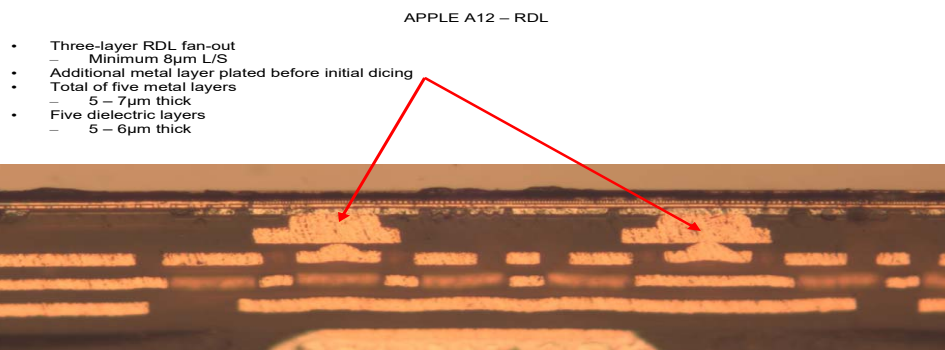
Figure 3. Expanded Fanout RDL cross-section (source Prismark Partners)

TSMC's Fanout (InFO) process has been implemented in the Apple iPhone for its application processors since its adoption for the A10 processor in 2016, the A11 in 2017, and the A12 in 2018, and it was also used for the application processor in the Apple S4 watch in 2018. Samsung has implemented its Panel Level Fanout technology[5] in its Galaxy phone and smart watch since 2018. Readers are encouraged to read the WLP – Fan-in and Fan-out roadmap (Chapter 23) for an in-depth discussion.

Network Connectivity

With the exponential growth of data traffic in the global digital economy and 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, with increased bitrate requirements likely towards 10 Gbps, which can only be enabled through 5G mmWave technologies. This will be described in depth in the RF/Analog/5G Roadmap (Chapter 12).

Network systems applications require large FC-BGA packages for the coming 5G mobile broadband and its higher-bandwidth data communication. This will require higher I/O counts and lower signal transfer loss. High I/O counts require substrates with fine line and space widths and fine vertical interconnect pitches (uVia and plated through hole). A low-loss signal transfer rate requires low loss dielectric materials and smoother conductor surfaces.

These substrate requirements are similar to those in the High Performance Computing discussion in the previous section. Each network application will have an operating environment and operating life requirements different from the High Performance Computing needs.

	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 (μm)	7	3	1.5
Build up roughness, Rq (nm)	500~600	400~300	100~200

Table3. 5G/mmWave bandwidth and substrate materials requirements

Autonomous vehicle

With 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 Test Qualification for Integrated Circuits" provided requirements for Grade 0, Grade 1, Grade 2, and Grade 3 based upon their function. AEC-Q104 (published 9-14-2017) "Failure Mechanism based stress test for Multichip Modules" addresses increasingly complex heterogeneous integration packages, such as SiPs, and MCMs, including 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 the Autonomous Vehicle require a high level of computing power, in-vehicle networking and sensor integration, requiring multi-die large form factor package integration using advanced packaging. The challenge is how these advanced packages can meet the stringent requirements of these automotive applications. 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 (CTE_1 & CTE_2) and interlayer adhesion need to be carefully examined to meet Grade 1 and Grade 0 requirements.

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 spaces

Difficult Challenges

In this substrate section we have reviewed leading edge organic substrates for FC-BGA & FC-CSP used in four major market application areas. We have show that wafer and panel fanout are potential alternatives, displacing conventional laminate substrate technologies in future leading-edge applications. Difficult challenges need to be addressed to retain a 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.

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.

Year of Production	Package Body (mm)				Package I/O pitch (mm)			
	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 (245C).

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 foot print 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 close 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 acheieve in proximity to large-body BGA SiP footprints.

High Accuracy Device Placement

Die stacking technologies and other semiconductor integrations 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 polyimer 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.   Repeatably 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 can 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). Those approaches could be considered as 2.5D printed electronics and will not be the focus of this section. Also not covered are developments 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), Inkjet Printing, 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 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.[1–4]

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.

However, currently AE technologies are largely under-developed for SiP applications, making them more suitable for low-complexity consumer electronics. In this section, 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.

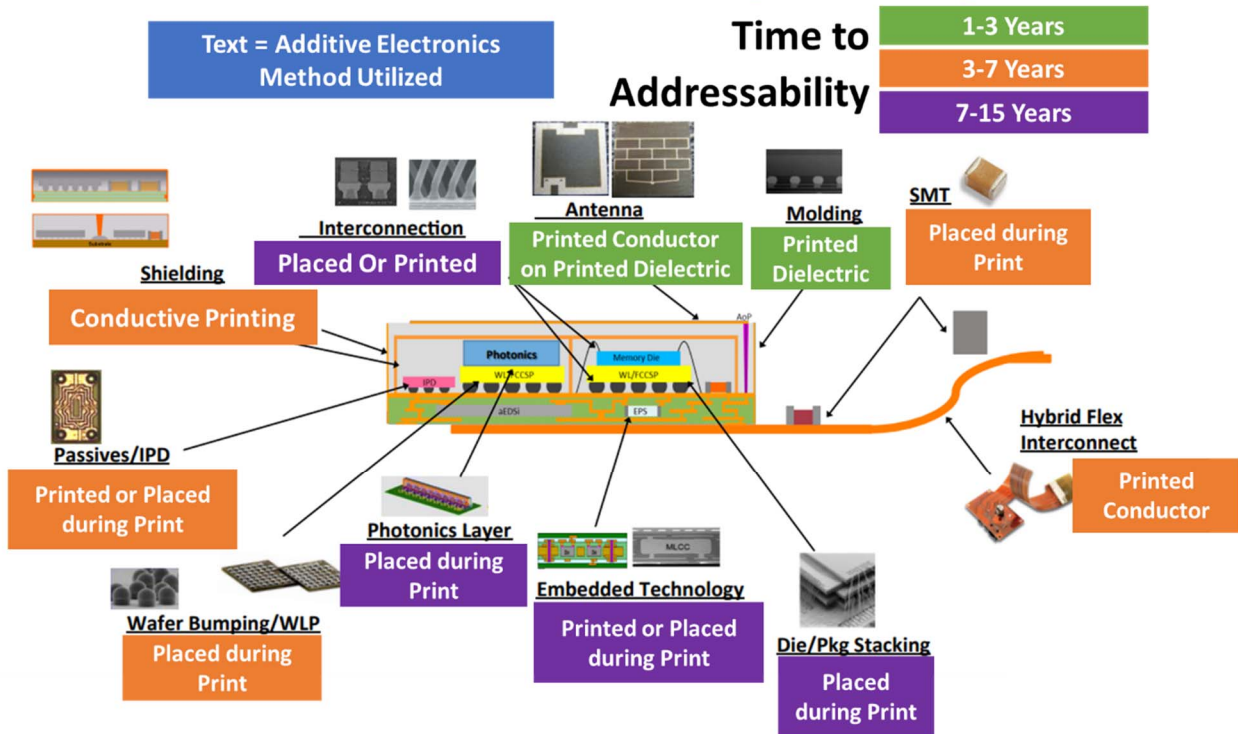## Additive Electronics Possibilities for Hetergenous Integration

Figure 1: Possible uses for AE within SiP applications for heterogenous integration with estimated time horizons for development. See section 1, figure 5. [5]

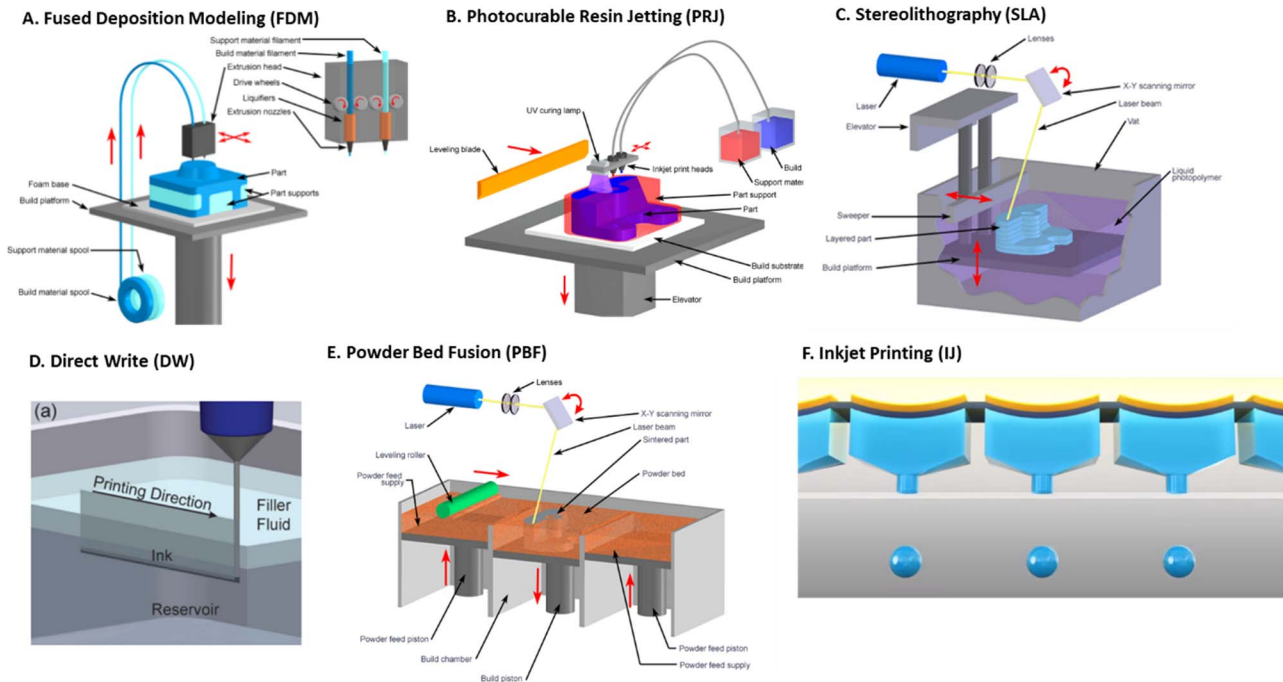## Basic Additive Manufacturing Technologies

Figure 2: Basic AM technologies, which form the basis for AE [9,10]

B. Current State of AE Technologies

A number of AM technologies have generally been applied towards AE, with a few reviews being available.[3,6–8] 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.

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 a graphene-filled polymer) for the selective deposition of conductive features.[11] Alternatively, FDM can be combined with DW using highly conductive pastes (usually silver-based) being extruded during the FDM print (Figure 3A). The combined approached can be complimented with P&P and CNC, with components added during or after the print.[3] FDM benefits include good 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.
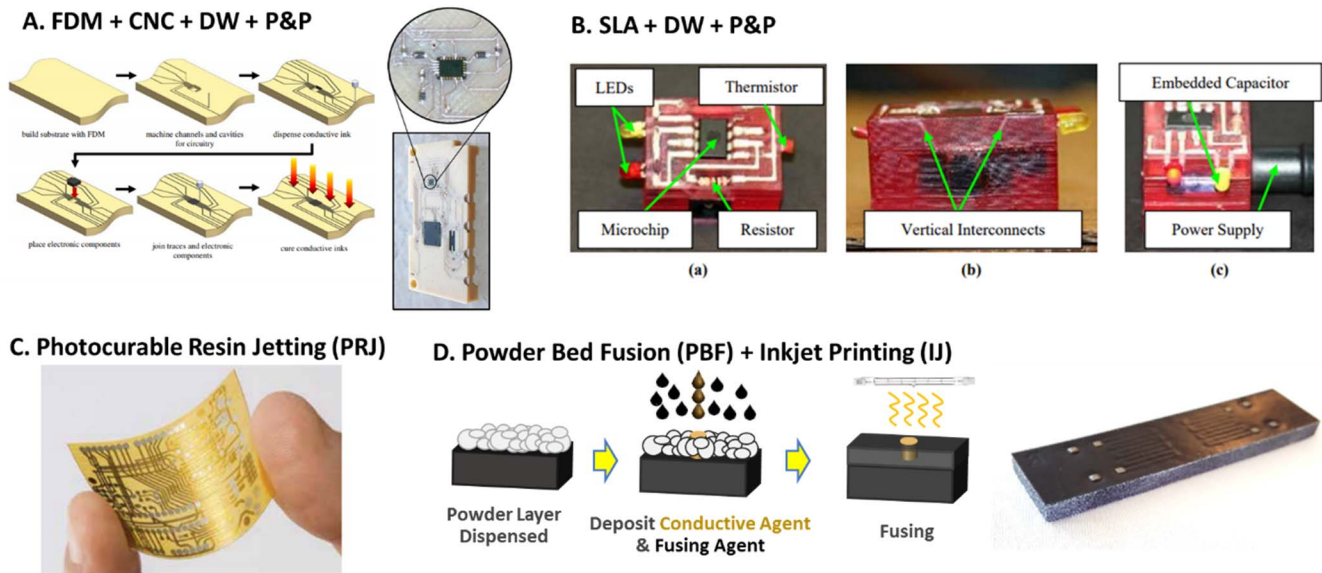

Figure 3: AE technologies, which often combine one or more AM approaches and conventional manufacturing approaches. A,[3] B,[2] C,[12] D[18]

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 photoresin with silver is jetted along with a dielectric support material (Figure 3C).[13] 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.[1,9,14–17] Benefits include high conductivity traces (up to 11.8^{-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.[4] Nanodimension has a PRJ AE commercial printer, specialized for the fabrication of multi-stack PCBs.[12] HP has presented a pre-commercial, research-level PBF + inkjet technology capable of

producing highly conductive traces within a dielectric polymer material.[18] Another common AE commercial tool is Optomec's Aerosol Jet technology, which is widely used for printing conductive features, such as antennas, onto 3D parts,[19] but it is not optimized for high build rates of dielectric material.

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.

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 ℃	FDM (PPSF)	>220 ℃	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.

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

For printing attributes, the line width, space width and conductivity are of primary importance. Resolutions down to 20 μm and conductivities of up to 11.8^{-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 the AE apparatus additional tools such as P&P, PP/R, wire-bond, or others so 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 incorporating 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 present the most significant yet necessary developmental challenge for AE in order to offer a robust SiP solution.

1. Macdonald E, Salas R, Espalin D, et al. IEEE Access. 2014;2:234-242.
2. Joe Lopes A, MacDonald E, Wicker RB. Rapid Prototyp J. 2012;18(2):129-143.
3. Espalin D, Muse DW, MacDonald E, Wicker RB. Int J Adv Manuf Technol. 2014;72(5-8):963-978.
4. https://www.nscrypt.com/solutions/printed-circuit-structures/.
5. https://eps.ieee.org/technology/heterogeneous-integration-roadmap.html.
6. Rim YS, Bae S-H, Chen H, De Marco N, Yang Y. Adv Mater. 2016;28(22):4415-4440.
7. Wang X, Jiang M, Zhou Z, Gou J, Compos Part B. 2017;110:442-458.
8. https://www2.deloitte.com/content/dam/insights/us/articles/3837_3D-opportunity-for-electronics/DUP_3D-Opportunity-Electronics.pdf.
9. Muth JT, Vogt DM, Truby RL, et al. Adv. Mater. 2014; 26, 6307-6312.
10. http://www.custompartnet.com.
11. http://www.graphene3dlab.com/s/home.asp.
12. https://www.nano-di.com/.
13. Layani M, Cooperstein I, Magdassi S. J. Mater. Chem. C, 2013; 1, 3244-3249.
14. Mannoor MS, Jiang Z, James T, et al. Nano Lett. 2013; 13, 2634-2639.
15. Ladd C, So J-H, Muth J, Dickey MD. Adv Mater. 2013;25(36):4953-4953.
16. Guo S-Z, Qiu K, Meng F, Park SH, McAlpine MC. Adv Mater. 2017;29(27):1701218.
17. Hoerber J, Glasschroeder J, Pfeffer M, Schilp J, Zaeh M, Franke J. Procedia CIRP. 2014;17:806-811.
18. https://www.engineering.com/3DPrinting/3DPrintingArticles/ArticleID/12298/The-Future-of-HPs-Multi-Jet-Fusion-3D-Printing.aspx.
19. https://www.optomec.com/printed-electronics/aerosol-jet-technology/.

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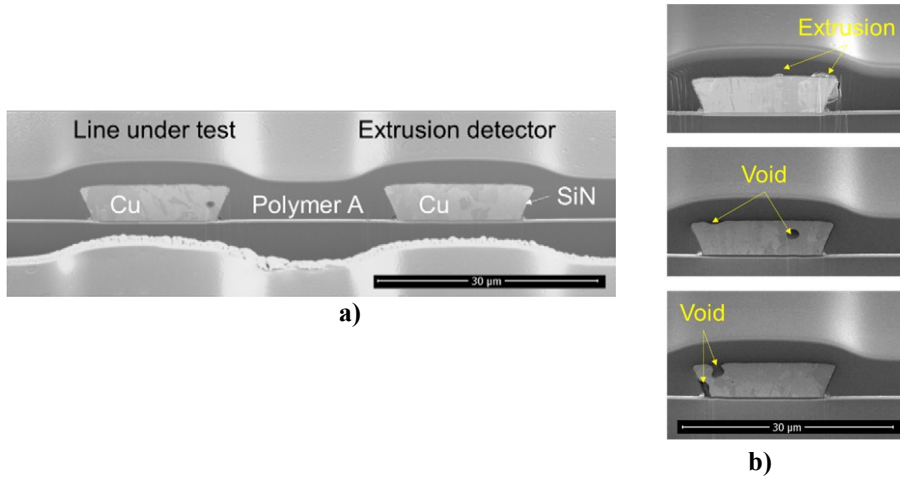
Figure 1: SEM micrograph of a Cu RDL line passivated by a bilayer material (SiN/polymer) after EM test (200 °C, 500 mA). a) Overview, b) types of electromigration-induced defects (voids, extrusion) and localizations (bulk, Cu/SiN interface) [4].

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.
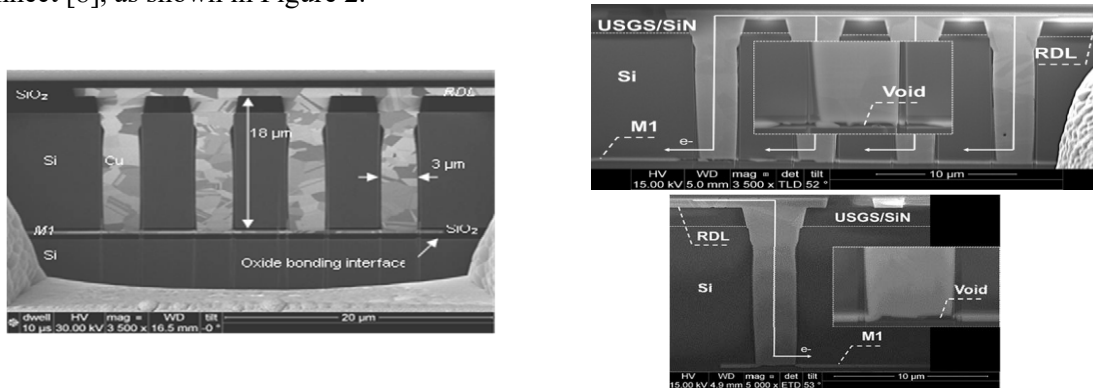

Figure 2: DD TSV and EM damage at the interface between TSV and BEOL interconnect [8].

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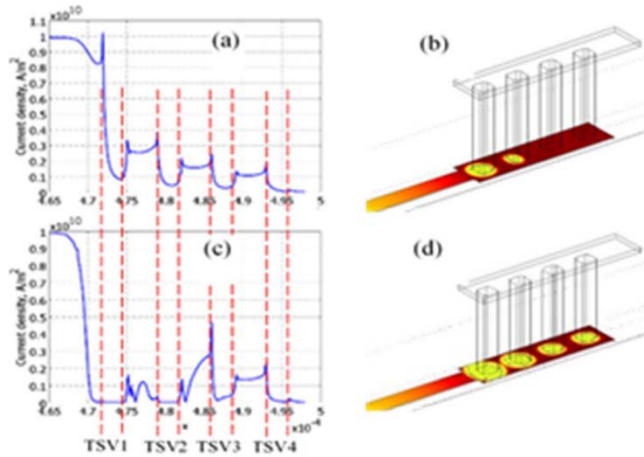
Figure 3: FEA results for up-stream configuration showing (a) initial current distribution, (b) first TSV failure, (c) current distribution after failure of two TSVs, and (d) the circuit failure. Color maps in (b) and (d) show the hydrostatic stress level: yellow is zero, darker red shows higher tensile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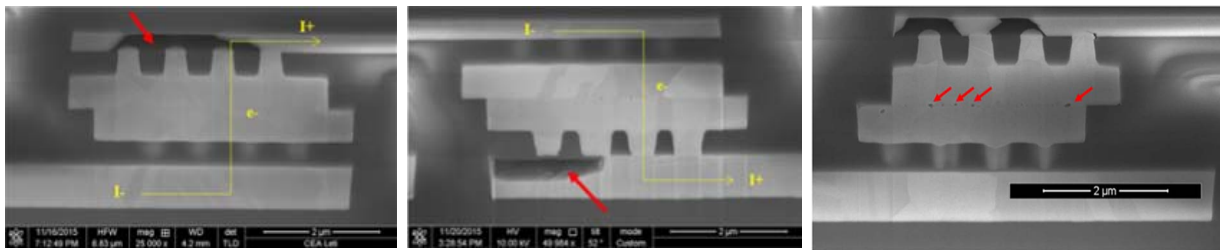

Figure 4: EM-induced voiding in the NIST-like structures [4].

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, thermal compression 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.
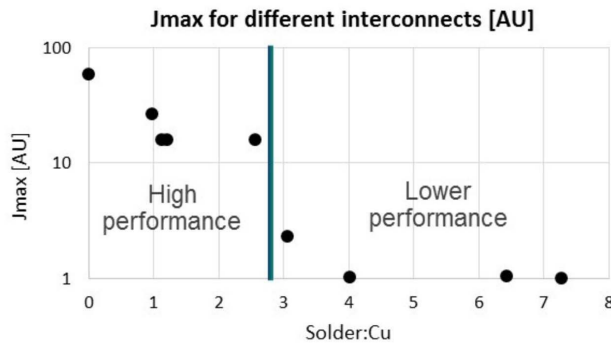

Figure 5. Classification of Jmax behavior for different solder-to-Cu volume ratios for solder interconnect structures [10].

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, as shown in Figure 6b. Such an approach can effectively relax the current density design rules with significant improvements in power, time-to-market and design cost.
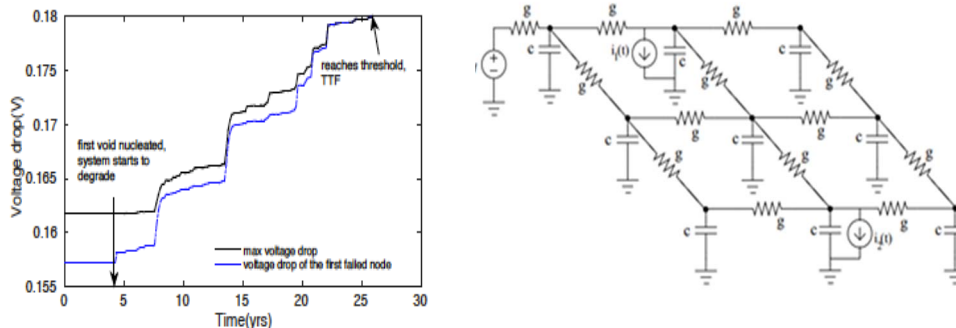
Figure 6. (a) Voltage drop of the first failed node and the maximum voltage drop in a power grid as a function of time [13], (b) Power grid schematics with user-provided current sources and voltages [14].

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 combined 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". In Proceedings of the 5th WSEAS international conference on Microelectronics, 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 International Conference on Simulation of Semiconductor Processes and Devices (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.

Section 12: Reliability

1. Introduction: HI System Reliability

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 (focus of this chapter) 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 1, 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.
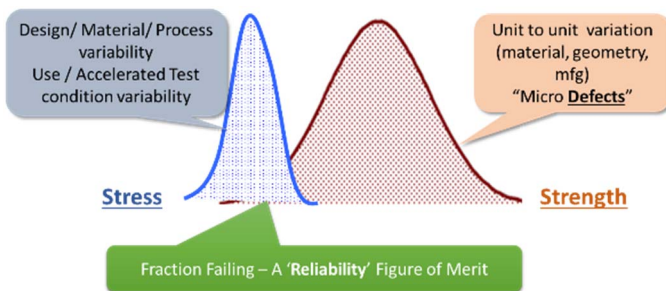

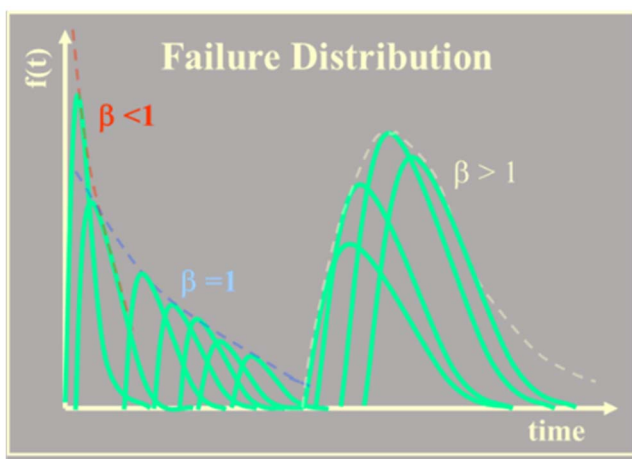*Figure 1: 'Stress' vs. 'Strength' interference*

Figure 2: Bathtub Curve (replotted as scaled pdfs), showing multitude of underlying competing failure

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 2 shows the traditional top-down view of system-level reliability, shown as a 'bathtub' curve (plotted in terms of scaled probability distribution funcions (pdfs) instead of hazard rates), with three

classical phases of infant mortality: (decreasing failure rate (with Weibull scale parameter β<1), mid-life 'random' mortality (constant failure rate with β=1) and end-of-life wearout mortality (increasing failure rate with β>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.
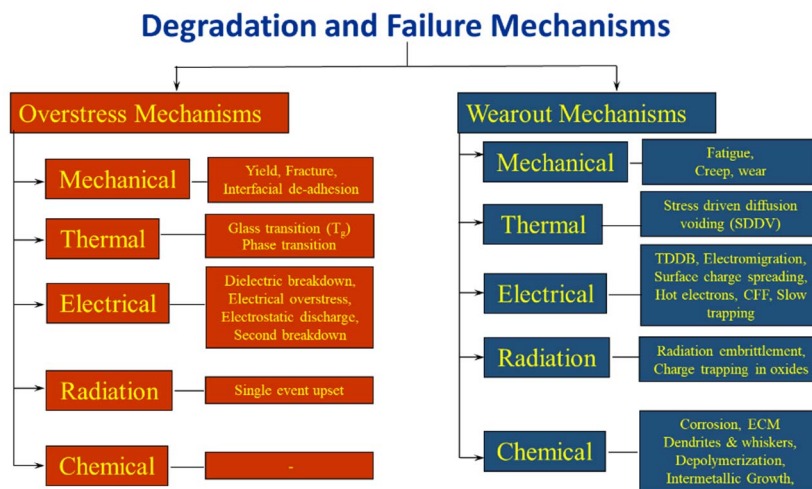

Figure 3a. Examples of dominant multi-physics degradation/failure mechanisms in electronic systems, under overstress and wearout stress exposures
(TDDB = Time-dependent dielectric breakdown; CFF = conductive filament formation; ECM = electrochemical migration)
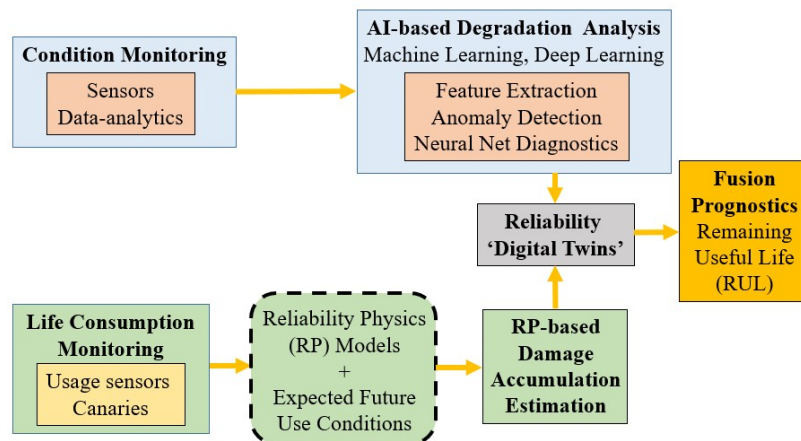
Figure 3a below provides a sample listing of the dominant multi-physics degradation mechanisms in the bottom-up RP perspective of microelectronics hardware systems. 'Overstress' mechanisms are triggered under the action of sudden catastrophic stress events while 'wearout' mechanisms cause gradual damage accumulation throughout the life cycle because of routine operational and environmental stress exposures.

Each listed mechanism represents a rich body of expert knowledge, 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)

Figure 3b. 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.

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 3b.

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**.

2. Scope

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 literature[8]. 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).

In the current version of this HIR Document, the scope of the Reliability Topic will be limited to the Single-Chip/Multi-Chip Package Integration Segment (Second row of Table 1). In future versions, the Reliability Topic may 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 literature[8]. 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.

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

Table 1: Hardware reliability topic matrix

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 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.

Tables 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.

	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: Difficult Challenges &Approaches for Reliable HI Technologies

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-orientated 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 remaining number of software defects (bugs).
- Create main-stream 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-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

Section 13: Summary and Difficult Challenges

This chapter starts with an Executive Summary followed by 11 sections on key technology building blocks, from Knowledge Base & Data to Manufacturing & Physical Infrastructure.

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

- Manufacturing & Physical Infrastructure
  - Wafer Singulation & Thinning
  - Wirebond
  - Flip Chip
  - Substrate
  - Board Assembly
  - Additive Manufacturing

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

Electronics are deeply embedded into the fabric of our society, changing the way we live, work and play while bringing new efficiencies to our 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 inflection point of the rise of tech company disruption, the plateauing of Moore's Law, the explosive expansion and growth of digital data, and adoption of IoT and artificial intellegence, continued progress requires a different phase of electronics innovations.

In the later part of Gordon Moore's celebrated 1965 paper [1] he took 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."

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 & knowledge – for heterogeneous integration for all the market segments.
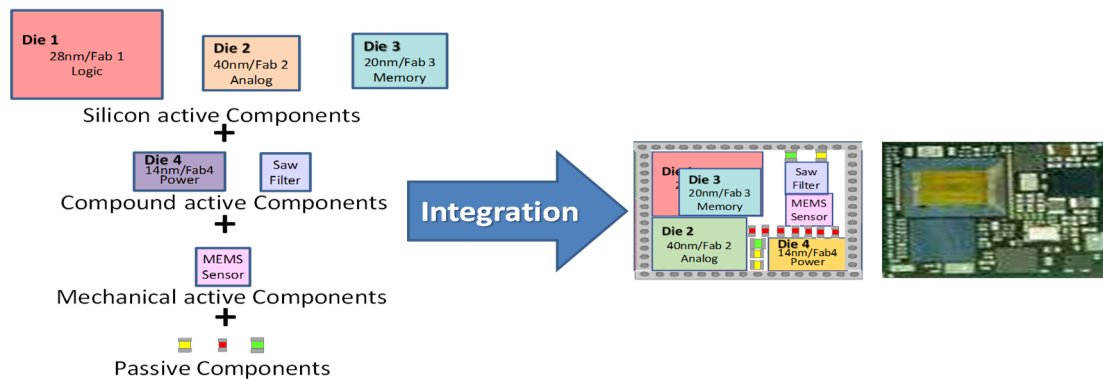

Figure 1. Heterogeneous Integration & System in Package (SiP) (Source: ASE)

This Single Chip & Multichip chapter has been designed to understand the current status for key technology building blocks from Knowledge Base and Data to Manufacturing and Physical Infrastructure, and to ask the questions:

- What are the challenges ahead looking into the future?
- What are the potential solutions?

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

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

Difficult Challenges:

While each section has articulated its 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; testing for KGD; Interface Protocol; Testing 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 the many packaging materials including the laminate substrate. Warpage and stress control for advanced nodes (3nm) 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: High Volume and High Mix manufacturing and assembly
- **Issue**: Consumer drive for "custom" products.
- **Potential Solution**: integrate additive manufacturing into the 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 & defense, mobile infrastructure, wearables & health.
- **Potential Solution**: Full utilization of AI- and sensor-driven data across the production line, field trial and ramp-up.

Acknowledgments

Edited by Paul Wesling