2021 Edition

Chapter 8: Single Chip and Multi Chip Integration

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

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

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

Table of Contents

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

Chapter 2: High Performance Computing and Data Centers

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

Chapter 4: Medical, Health and Wearables

Chapter 5: Automotive

Chapter 6: Aerospace and Defense

Chapter 7: Mobile

Chapter 8: Single Chip and Multi Chip Integration

Section 1: Executive Summary and Scope 1

Section 2: Electrical Analysis and System Requirements 10

Section 3: Thermal Management 18

Section 4: Mechanical Requirements 22

Section 5: Thinning and Singulation 26

Section 6: Wire Bonding 30

Section 7: Flip Chip and Interconnects 38

Section 8: Substrates 42

Section 9: Board Assembly 49

Section 10: Additive Manufacturing 53

Section 11: Electromigration 57

Section 12: Summary and Difficult Challenges 65

Chapter 9: Integrated Photonics

Chapter 10: Integrated Power Electronics

Chapter 11: MEMS and Sensor Integration

Chapter 12: 5G, RF and Analog Mixed Signal

Chapter 13: Co-Design for Heterogeneous Integration

Chapter 14: Modeling and Simulation

Chapter 15: Materials and Emerging Research Materials

Chapter 16: Emerging Research Devices

Chapter 17: Test Technology

Chapter 18: Supply Chain

Chapter 19: Cyber Security

Chapter 20: Thermal

Chapter 21: SiP and Module

Chapter 22: Interconnects for 2D and 3D Architectures

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

Chapter 24: Reliability

Chapter 8: Single Chip and Multi Chip Integration

Section 1: Executive Summary and Scope

Send corrections, comments and suggested updates to the TWG chair, using our HIR SmartSheet:

https://rebrand.ly/HIR-feedback

1. 2020 – 2021 Disruptions, Dislocations and Innovations

History will mark the years 2020 and 2021 as a period of rapid global disruptions, dislocations and global innovations on a scale seldom seen in history in peacetime. Disruptions included COVID-19, extreme wild fires, and flooding across the globe, accompanied by climate change and semiconductor shortages.

In January 2020, the HIR team and the EPS Santa Clara Valley Chapter, together with SEMI colleagues, were planning the 3rd HIR symposium and annual meeting, February 23-24, with all the excitement towards celebrating the release of the 1st HIR edition in October 2019. We moved the annual meeting site from the new Samsung showcase facility in San Jose to SEMI's Global Headquarters in Milpitas when the Samsung site was closed to visitors. This meeting was hugely successful, with much networking in formalized work sessions, during luncheons and a wine tasting event, and business card exchanges. The day after this HIR conference, some of our industry colleagues started to receive notices from their companies restricting conference travel attendance to mitigate COVID-19 risk.

Little did we imagine the extent and ravages of the COVID-19 pandemic that we know today. Shown below are data from the dashboard of World Health Organization October 6, 2021.

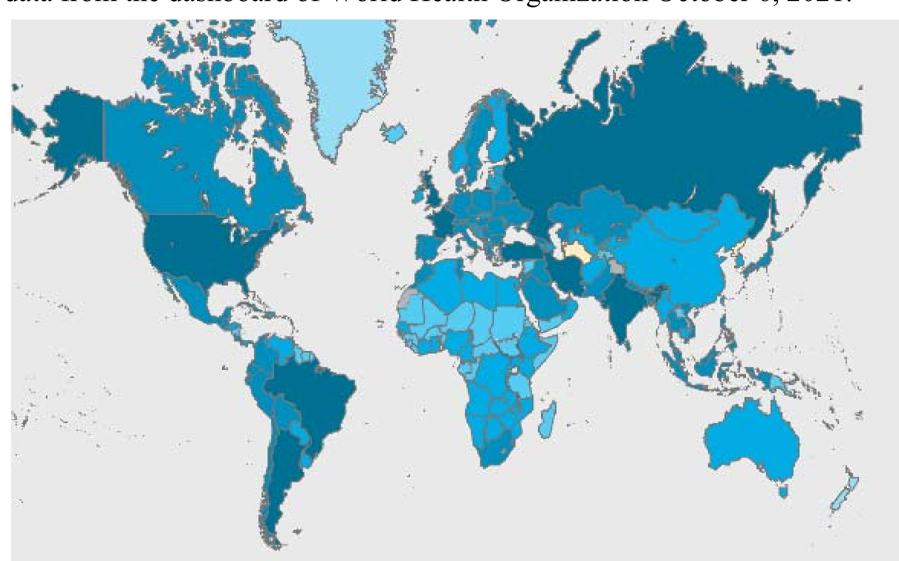


Figure 1A: Globally, as of 6pm CEST, 8 October 2021, there have been 236,599,025 confirmed cases of COVID-19, including 4,831,486 deaths, reported to WHO. As of 6 October 2021, a total of 6,262,445,422 vaccine doses have been administered.

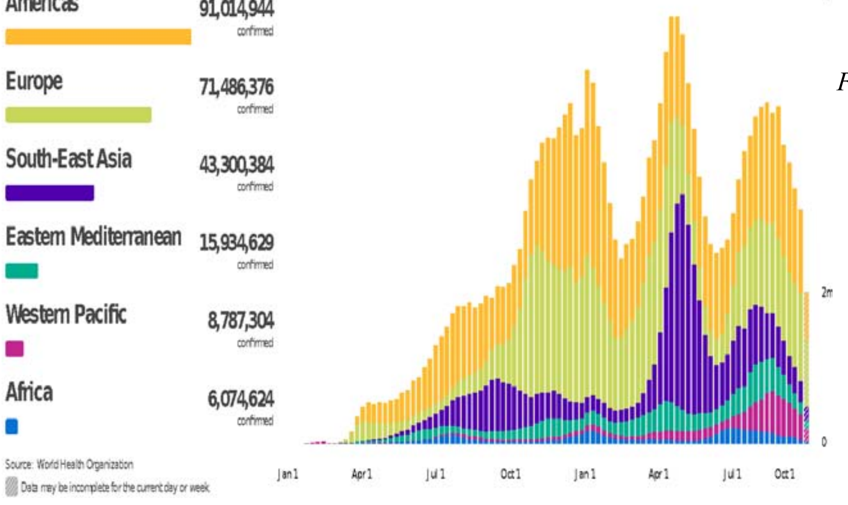


Figure 1B: Regional distribution of confirmed cases of Covid 19 Globally, as of 8 October 2021, reported to WHO. (Source: WHO dashboard)

How do electronics and semiconductors contribute to fighting the COVID-19 pandemic crisis? Medical workers in hospital ICUs and emergency rooms depend upon medical equipment/devices powered by highly sophisticated electronics for diagnostics, monitoring and patient care. Good examples are portable semiconductor ultrasonic devices for patient lung imaging diagnostics, and life-saving ventilators for automated oxygen level monitoring and oxygen delivery when patients contract the COVID-19 virus.

The pandemic generated huge amounts of data in different data formats that researchers use for vaccine development, biometrics, etc, in different formats and languages. Cutting-edge computing capabilities for rapid response and accelerated development, data searching, discovery, and analysis were provided. One such effort is the COVID-19 HPC Consortium spearheaded by the U.S. White House, Department of Energy, national laboratories and industry. (Source link: covid19-hpc-consortium.org)

As the WHO recorded 6,262,445,422 vaccine doses administered, it is worth noting that the accelerated vaccine development went from laboratory to manufacturing and distribution to the population in a short period of days, weeks, and months thanks to many dedicated people in the knowledge supply chain and public health institutions. An article titled "The Tangled History of mRNA Vaccine" in Nature (Volume 597 page 318, September 16, 2021) by Elie Dolgin noted that hundreds of scientists have toiled to advance and to expand the science knowledge base in mRNA to realize its health potential. Not until the COVID-19 pandemic were there calls to urgent action for collaboration and sources of funding leading to breakthroughs in vaccine design, development and trial globally." In commercial product terms, product volume drives innovation, operational learning and business – true for the vaccine and true for semiconductors and electronics.

Throughout the many months of pandemic-driven lock-down, the IT infrastructure (semiconductor, packaging and electronics system-based) has been highly robust in maintaining communication between businesses and customers, colleagues and families, teachers and students, in remote working, shopping, and virtual everything. The pandemic has accelerated the transformation to the digital economy while keeping global commence spinning. Hospitals can order goods online, and stay connected. Critically, scientists could develop treatments and vaccines to begin making the world healthy again. Without semiconductors and the resultant products and services, and using the world's most advanced supercomputers, for example, the historically rapid development of COVID-19 vaccines would not have been possible.

In June 2021, the White House issued a report titled "Building Resilient Supply Chains, Revitalizing American Manufacturing, and Fostering Broad-Based Growth". This report is in response to the Presidential Executive Order 14017 issued in February 2021, "America Supply Chains", directing the administration to perform a 100-day review, due June 2021, to assess supply chain vulnerabilities across four key product areas. They are:

- Semiconductor Manufacturing and Advanced Packaging
- Large Capacity Batteries
- Critical Materials
- Pharmaceutical and Active Pharmaceutical Ingredients

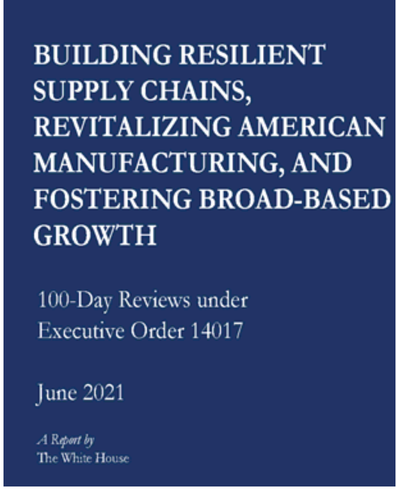


Figure 2: Cover page of White House Report "Building Resilient Supply Chains Revitalizing American Manufacturing, and Fostering Broad-Based Growth". June 4 2021

The report begins with:

> "The COVID-19 pandemic and resulting economic dislocation revealed long-standing vulnerabilities in our supply chains. The pandemic's drastic impacts on demand patterns for a range of medical products including essential medicines wreaked havoc on the U.S. healthcare system. As the world shifted to work and learn from home, it created a global semiconductor chip shortage impacting automotive, industrial, and communications products, among others."

The US Government and Congress has legislation underway to invest US$52 billion on semiconductor and advanced packaging research. Companies including Intel, TSMC, Samsung and Texas Instruments have announced plans to invest many billions in wafer fab manufacturing capacity expansion.

Going forward, we are expecting expanding advances of electronics markets and applications in Internet of Things, Mobile Devices and Network Infrastructure, Automotive, Aerospace and Defense, and of course Medical, Health and Wearables technologies and products. They will be connected by 5G communications and further empowered by AI and ML.

HIR Designed for Market & Application Driven

IoT to IoE

Smart Mobile Everywhere

Data to the Cloud & Data Centers

Autonomous Vehicles

Wearables & Health

Aerospace & Defense

These Six Markets & Applications will be further transformed through the power of AI

Figure 3: Six Major Markets and Applications

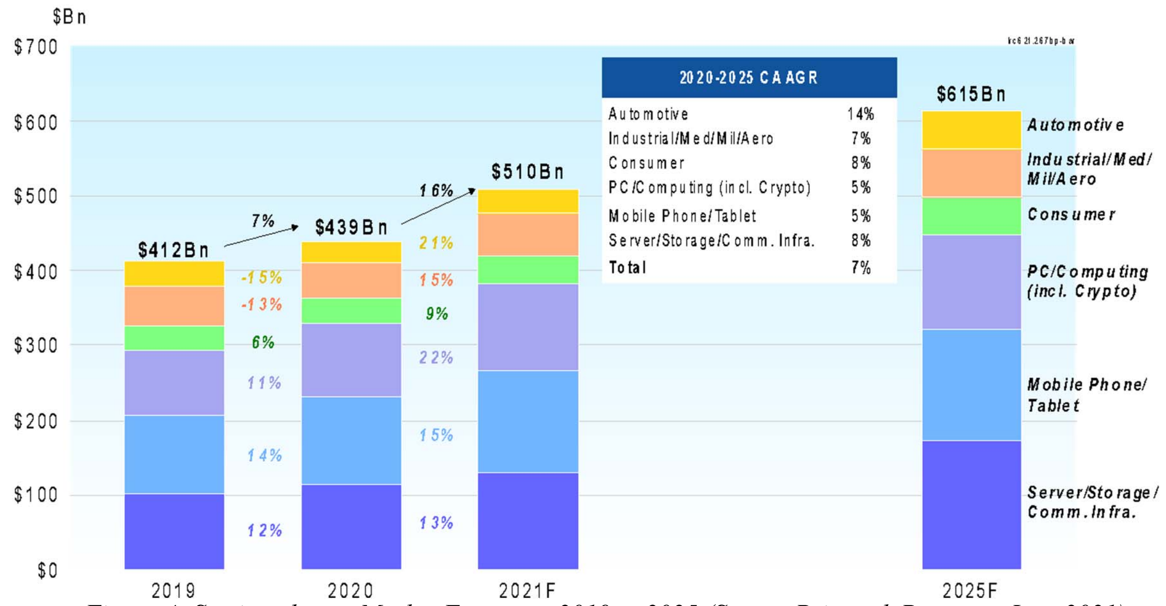


Figure 4: Semiconductor Market Forecast– 2019 to 2025 (Source Prismark Partners June 2021)

Shown above is the Semiconductor market size forecast for different market applications in Figure 3 from 2019 to 2025.

2. Heterogeneous Integration Roadmap

The Heterogeneous Integration Roadmap comprehensively covers the entire semiconductor and electronics technology ecosystem (see Figure 5). It serves as the knowledge-based roadmap for future electronics technology. The Roadmap is market- and application-driven, starting with six specific market segments: High Performance Computing and Data Center, IoT, 5G Communications and Beyond, Smart Mobile, Automotive, Wearable and Health, and Aerospace and Defense. They are followed by 5 Heterogeneous Integration Components, 7 Cross Cutting Technologies, 3 Integration Processes areas, and last but not least, Co-Design and Simulation. Each Technical Working Group contributes a chapter of the Roadmap. Together they work in collaboration to produce this roadmap. We represent the total technology ecosystem – industry, academia, and the research community.

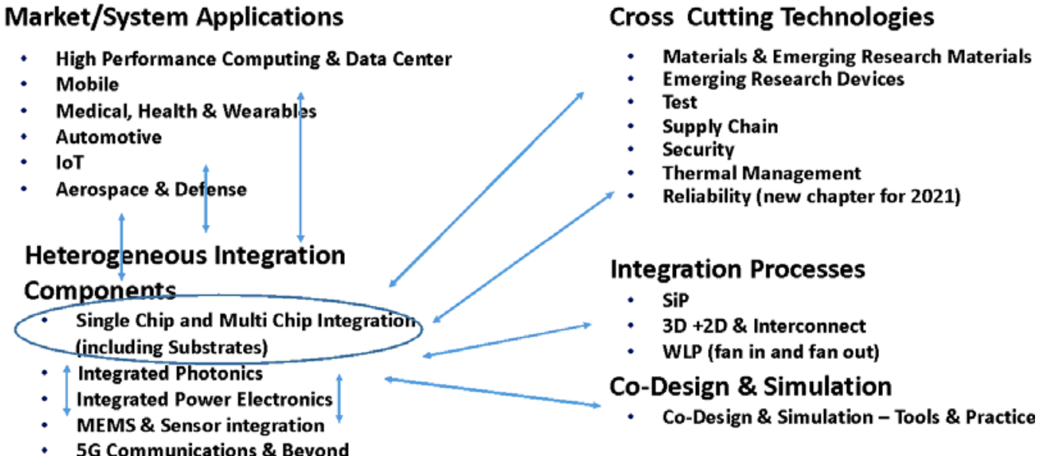


Figure 5: Heterogeneous Integration Roadmap structure

This Single and Multichip Integration Chapter covers the basic knowledge, base tools and physical manufacturing infrastructure tools across all market segments. This chapter serves as the packaging "tool box" to the other HIR technical working groups for roadmap readers. The eleven sections form the key technology building blocks from manufacturing processes and physical infrastructure to packaging knowledge base and data.

IC devices start as wafers from foundries, thinned and singulated into "chips". While wirebond remains the interconnect workhorse of the industry, there is very strong growth in Flip Chip (BGA-CSP) and WLCSP, wafer-level packaging (fan-In and fanout) and a rapid growth rate for advanced packages as shown in Figure 6.

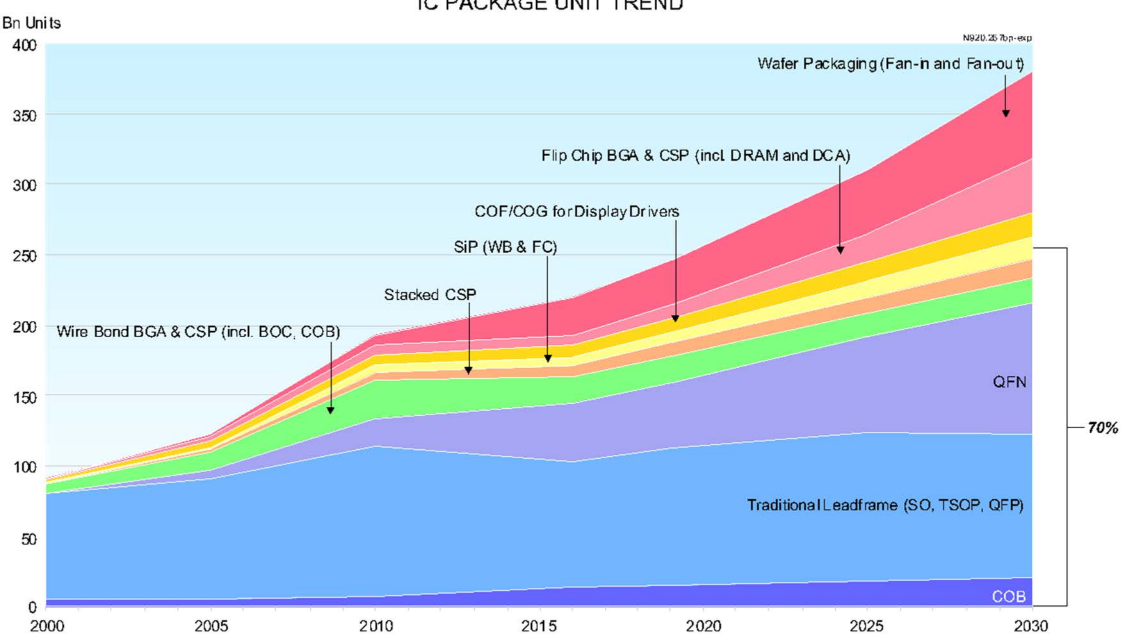


Figure 6: Chip-package interconnect technology trends (Bn units) (Source: Prismark Partners 06-2021)

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

A basic set of tool box for SiP and Heterogeneous Integration is illustrated in Figure below. 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 and through heterogeneous integration from Home Assistants, Smart Phones, Data Centers, Automotives, Avionics, and many other products in the trillion dollars global electronics market.

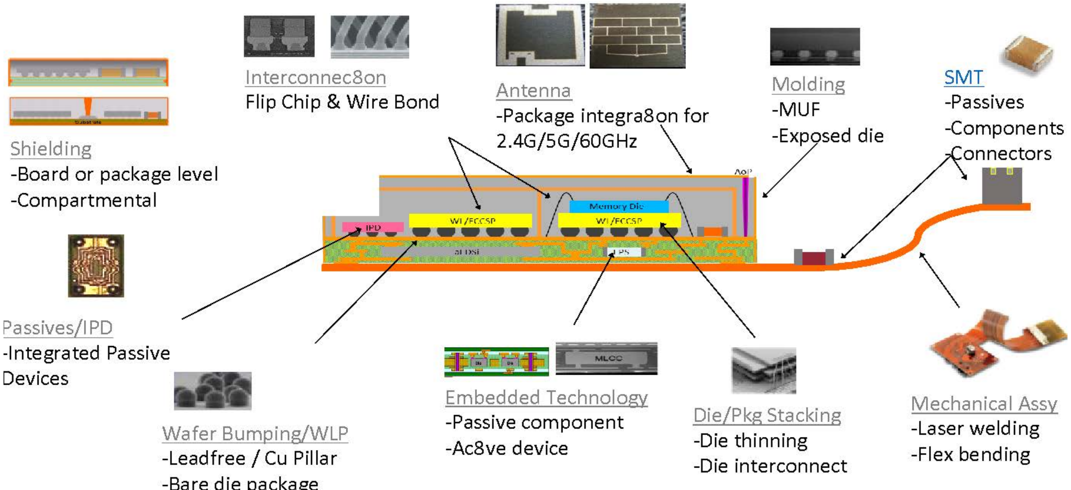


Figure 7: Tool Box for Heterogeneous Integration Technologies (Source: ASE)

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

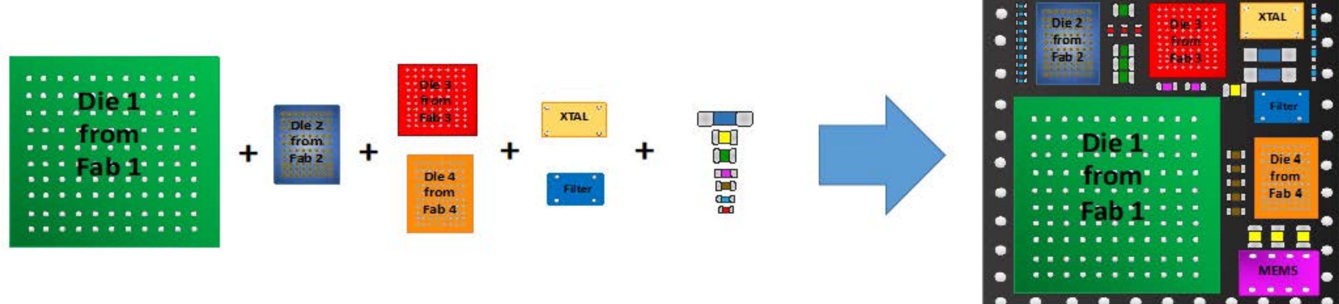
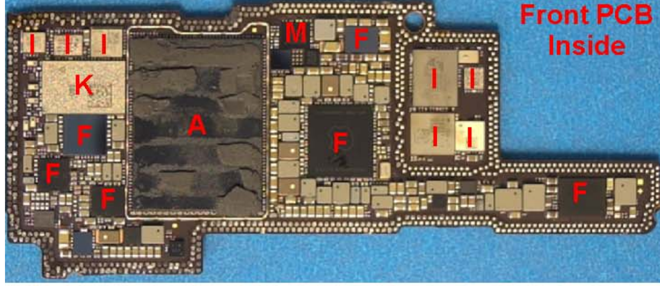
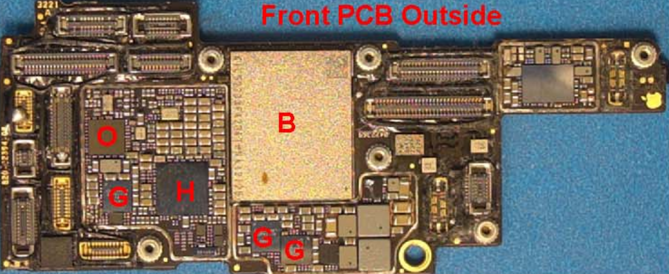
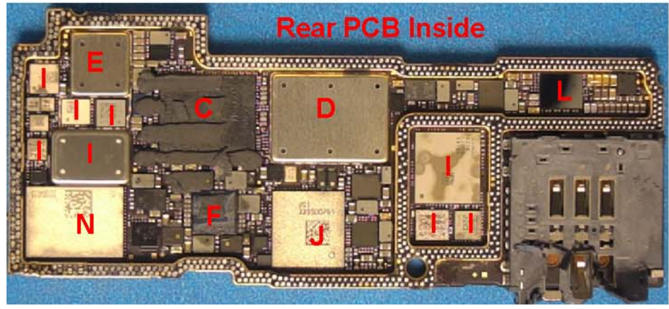


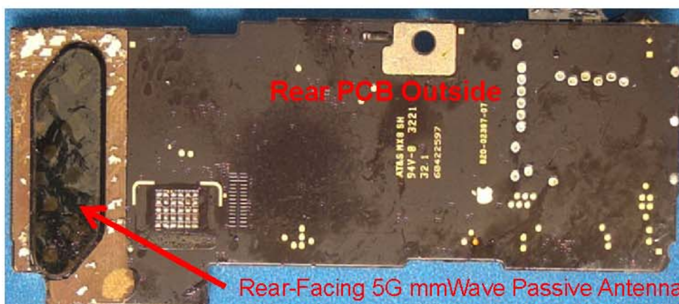
Figure 8: Heterogeneous Integration and System in Package (SiP) (Source ASE)

The smartphone industry has been an effective adopter of Heterogeneous Integration Technology in the use of SiP for its advantage in miniaturization, with its modularity allowing product rollout to subsequent generations in the market place. The application processor, housed in a PoP package with the memory component stacked in close proximity on the top, is almost always of the most advanced node. They are the premier examples of SiP, incorporating the most advanced-node processor integrated with a memory die in close proximity, for the demanding consumer market.









A. Apple A15 Applications Processor
B. Flash Memory
C. Qualcomm SDX60M 2G-5G Baseband
D. Qualcomm SDR868 2G-5G TRx (sub-6GHz)
E. Qualcomm SMR526 5G IF TRx (mmWave)
F. Power Manger (Apple, Qualcomm, TI, ST, etc)
G. Apple/Cirrus Logic Audio Amplifier

H. Apple/Cirrus Logic Audio Codec
I. FEM (Broadcom, Murata, Skyworks, etc)
J. USI WiFi Module
K. USI UWB Module
L. Broadcom Wireless Charging
M. NXP Display Manager
N. 5G mmWave Module
O. NXP NFC Controller

© Prismark Partners LLC

Figure 9: Apple iPhone 13 Pro main board assembly (Source Prismark Partners)

Shown above is the main board of the iPhone 13 Pro. The main processor die is housed in a PoP package, flip-chip assembled on an advanced substrate together with a wirebonded memory component on top. There are multiple modules, including 2G-5G Baseband, 2G-5G TRx (sub-6GHz and mmWave), multiple power managers from different sources, logic, audio amplifier, logic audio codec, FE modules, WiFi modules, UWB modules, wireless charging, Display manager, 5G mmWave module, and NFC controller. They are 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 (eg, a smartphone) to design and construct models from one product generation to another, both rapidly and economically, for the hugely competitive consumer market.

Let us now consider the High Performance Computing and Data Center market application. Shown in Figure 10 is an example of Heterogeneous Integration through SiP, with integration of the processor and HBM memory stack on a silicon interposer platform. The package provides massively parallel high-bandwidth connectivity to the HBM, significant power savings, and greater than 50% shrinkage in X-Y form factor.

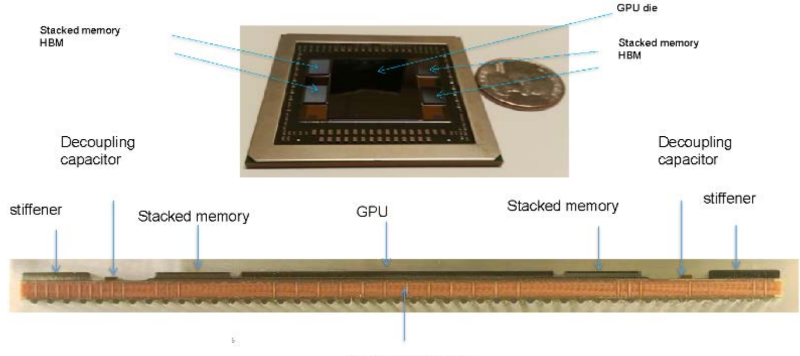


Figure 10: AMD Fiji GPU – HDM on Si Interposer 2.5 D Package (Source ASE)

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

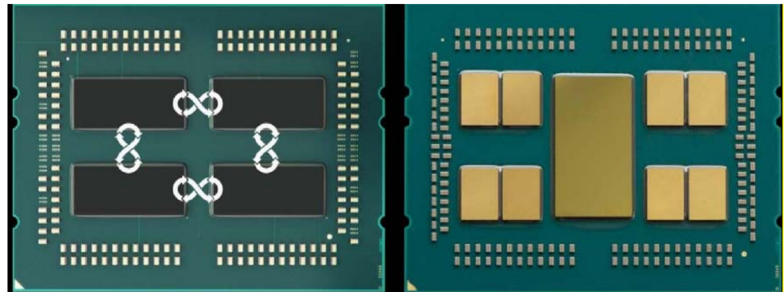


Figure 11 A: Examples of System Integration – 1st and 2nd Generations EYPC Server Processors (Source AMD)

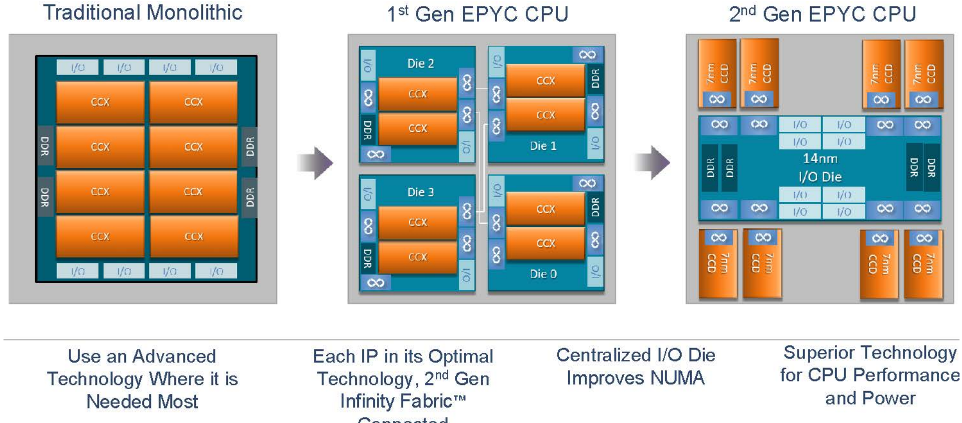


Figure 11 B: Chiplets in Heterogeneous Integration - EYPC Server Processors (Source AMD)

The emergence and growth of chiplet technology, as shown in Figure 11A and 11B, is transitioning the industry into the post-Moore semiconductor era. Using chiplets is the game-changer in this new era for dis-aggregation of monolithic SOCs for economics, time to market, and integration advantages. Components of different nodes or from different companies may be heterogeneously integrated together in one SiP, such as implemented in Intel's Kaby Lake G card, which incorporates a CPU with an AMD GPU linked to 4GB of HBM2. Silicon bridge on organic substrate (EMIB) developed by Intel was used to link multiple die together in close proximity in the package [3].

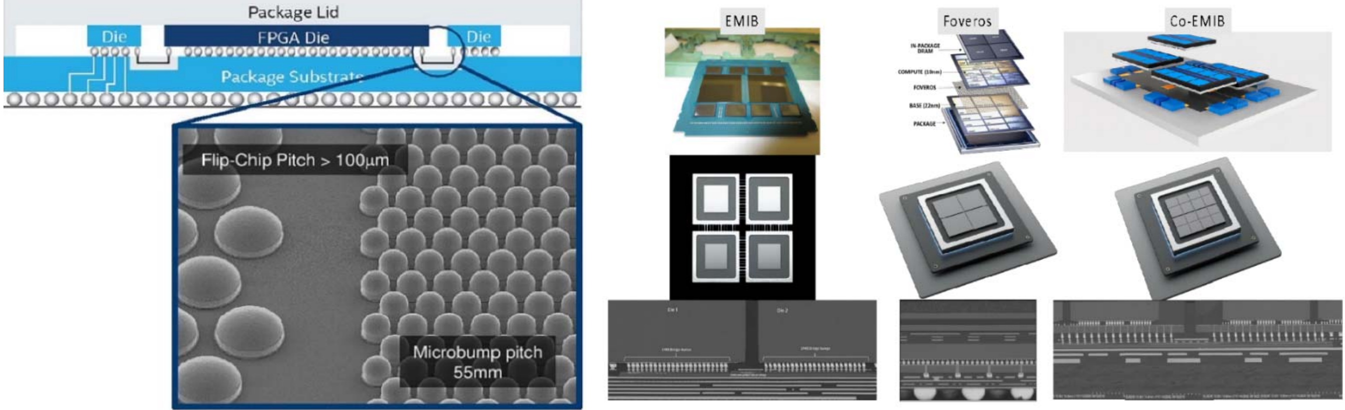


Figure 12A: Silicon bridge on substrate technology (EMIB) from Intel for Heterogeneous Integration. Figure 1.12B: Advance Package Chiplet Integration Technologies at Intel: EMIB, Foveros and Co-EMIB (Source Intel)

Extending the EMIB technology, Foveros and Co-EMIB designs form the advanced technology family in Intel to link multiple die (chiplets) on substrates in the package, as shown in Figure 12B.

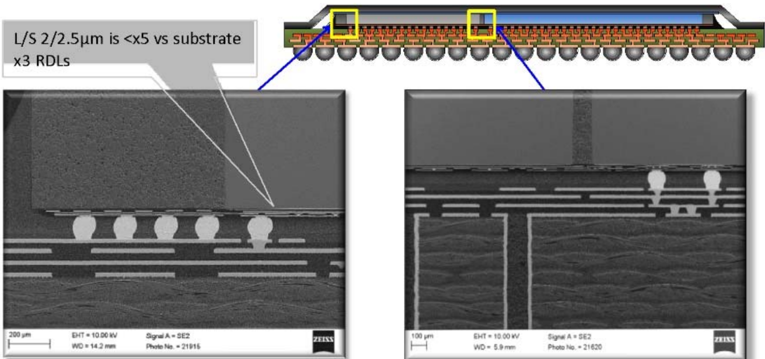


Figure 13: Wafer Level Fan-Out integrating 14 nm and 22 nm die in multichip package (Source ASE)

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 shown in Figure 13.

The previous examples demonstrate the growing momentum for SiP and chiplets in high-performance computing – taking a new look at system architecture and expanding innovations in our packaging 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, time to market, high bandwidth, and performance per watt make multi-die (SIP and chiplets) heterogeneous integration an important trend going forward [4-10].

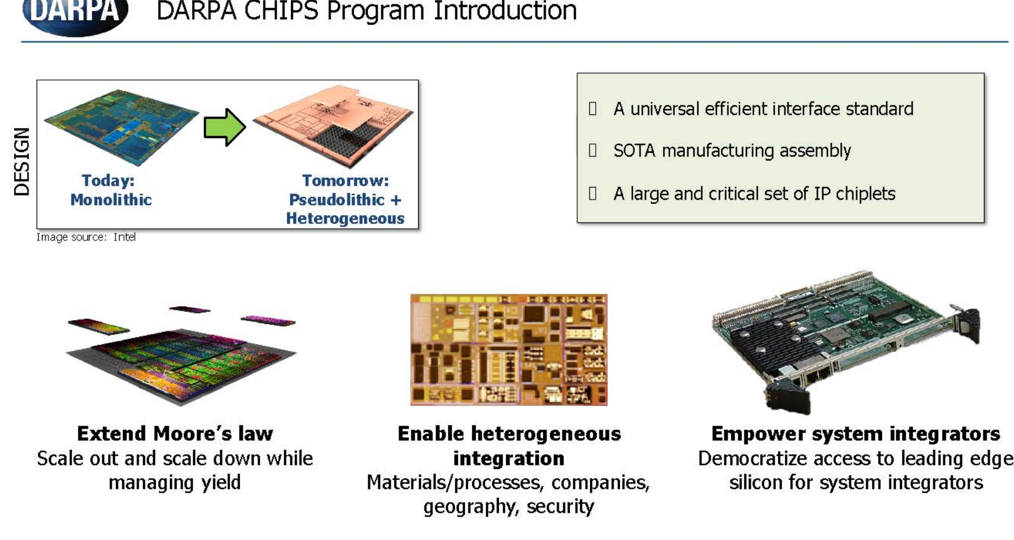


Figure 14: Chiplets in DARPA CHIPS Program: (Source DARPA)

At SEMICON West, July 2019, Andreas Olofsson, DARPA Microelectronics Technology, presented a talk illustrating chiplets in the DARPA CHIPS program. Using the chiplets concept, the program is aimed at three goals (Figure 14): (a) extending Moore's Law; (b) enabling heterogeneous integration; and (c) empowering system integrators.

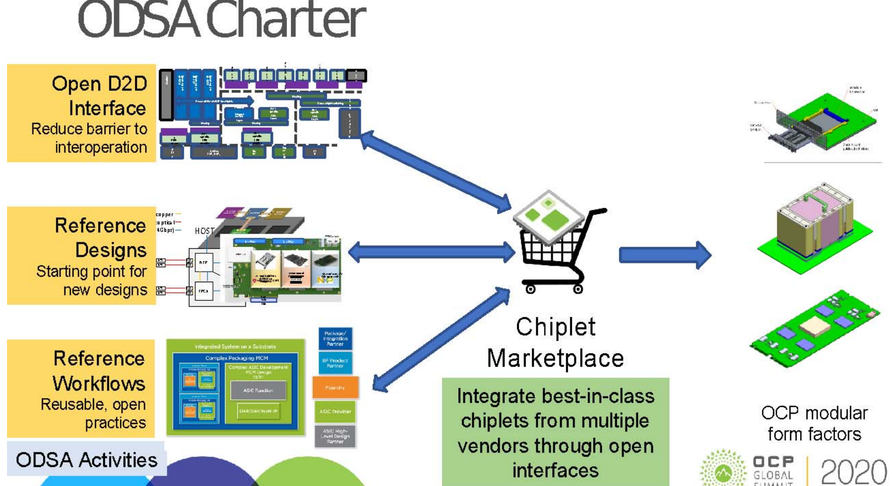


Figure 15: Chiplets in Commercial Perspective (Source: OCP – ODSA)

At the OCP Summit in May 2020 and HIR Workshop in February 2020, Bapi Vinnakota presented the charter for ODSA, which projects a commercial perspective vision of chiplets, with a commercial chiplets marketplace, enabling integration of best-in-class chiplets from multiple sources with heterogeneous integration for commercial electronics products (see Figure 15).

The scope of this Single Chip and Multichip chapter is designed to articulate the current state of the art of two fundamental tool boxes:

(a) Knowledge Base and Data for Package and System Architecture and Co-Design

(b) Package Assembly Manufacturing Flow for Substrates and Boards

We ask the questions: *What is the status of leading edge technologies in our tool box for SiP, chiplets and Heterogeneous Integration? What are the challenges ahead? What are the potential solutions?* We shall address these questions, and include in the roadmap the electrical, thermal and mechanical technology issues from device packaging to subsystem and system packaging, from system-package-device architecture and co-design to manufacturing, inclusive of the total ecosystem. Following Moore's words, our purpose in Heterogeneous Integration is to build large systems out of smaller functions – chiplets, SiP modules, and other functional packages – which are separately designed, packaged, interconnected and manufactured. This is the purpose and theme of the Roadmap. This Chapter starts with a section on Scope followed by 11 sections on key technology building blocks from Knowledge Base and Data to Manufacturing and Physical Infrastructure.

- Knowledge Base and Data for Package and System Architecture and Co-Design
  - Electrical Analysis and System Requirements
  - Thermal Management
  - Mechanical Analysis
  - Electromigration
  - Reliability
- Package Assembly Manufacturing Flow Substrates and Boards
  - Wafer Singulation and Thinning
  - Wirebond
  - Flip Chip
  - Substrate
  - Board Assembly
  - Additive Manufacturing

The Knowledge Base and Data and Manufacturing and Physical Infrastructure form the base foundation for Advanced Packaging and Integration technologies from 2D, 2.5D and 3D, for SiP and wafer level packaging. This chapter links to the other chapters in the Roadmap. Working together with other TWGs, we describe the current state of the art, and the roadblocks in the path going forward, to stimulate pre-competitive research and innovation as much as 15 years ahead.

1.7. References

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

Acknowledgments

The Single and Multichip Integration TWG has over 35 volunteer contributors from industry and academia. We are very fortunate to have this very knowledgeable team making our chapter a reality for our profession and our industry.

TWG Chair William Chen and Co-Chair Annette Teng

Section 1	Background and Scope	Bill Chen (ASE)
Section 2	Electrical Requirements	Lei Shan (Tekollect)
Section 3	Thermal Management	Bahgat Sammakia, Scott Schiffres, Srikanth Rangarajan (BU)
Section 4	Mechanical Requirements	Benson Chan (BU)
Section 5	Wafer Singulation &Thinning	Annette Teng (Promex)
Section 6	Wire Bond	Ivy Qin (KnS)
Section 7	FlipChip	Mark Gerber (ASE)
Section 8	Substrate	Kyu-Oh Lee (Intel)
Section 9	Board Assembly	Jim Wilcox (UIC)
Section 10	Additive Manufacturing	Kris Erickson (Facebook), Dishit Parekh (Intel)
Section 11	Electromigration	Paul Ho (UT), Valeriy Sukharev (Mentor) Eric Cotts (BU), Faramarz Hadian (BU)
Section 12	Summary & Difficult Challenges	Bill Chen

We are pleased to announce that the Reliability Section from previous Roadmap editions has been spun off to form a separate Technical Working Group as Chapter 24 of the Roadmap. Our best wishes to the new Reliability TWG team of Abhijit Dasgupta (UMD), Richard Rao (Marvell) and Shubhada Shasrabudhe (Intel).

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: Single- and Multi-Chip Integration Technology Requirements

Year of Production	*2021*	*2022*	*2023*	*2024*	*2025*	*2026*	*2027*
On-chip feature size (nm)							
Memory (DDR/HBM)	10	7	7	5	5	3	3
Smart Phone / Laptop	5	5	3	3	2	2	2
High-performance (note1), chiplet/monolithic	7	5	5	3	3	2	2
Maximum Average Power Density (W/mm^2)							
Memory (DDR/HBM) (note2)	0.06/0.12	0.05/0.11	0.05/0.11	0.04/0.1	0.04/0.09	0.04/0.08	0.04/0.07
Smart Phone / Laptop	0.5	0.5	0.55	0.55	0.6	0.6	0.65
High-performance	0.9	0.95	0.95	1	1	1.05	1.05
Core Voltage (Minimum Volts)							
Memory (DDR/HBM)	1.2	1.1	1.1	1	1	0.9	0.9
Smart Phone / Laptop	0.75	0.75	0.75	0.75	0.7	0.7	0.7
High-performance	0.8	0.75	0.75	0.75	0.75	0.7	0.7
Package Pin count Maximum							
Memory (DDR/HBM)	288/3200	288/3200	288/3200	326/4100	326/4100	350/4700	350/4700
Smart Phone / Laptop	1212/7000	1212/7000	1275/7600	1275/7600	1396/8400	1396/8400	1396/8400
High performance (note3)	6400	7800	7800	9600	9600	11200	11200
Minimum Package Dimension (mm)							
Memory (DDR/HBM)	133	133	133	133	133	133	133
Smart Phone / Laptop	50	50	55	55	60	60	60
High-performance	79	87	87	95	95	103	103
Performance: On-Chip							
Memory (DDR/HBM), MHz	800	800	1000	1000	1200	1200	1600
Smart Phone / Laptop, GHz	3.2	3.2	4	4	4.8	4.8	5.2
High-performance, GHz	6.4	8	8	9.6	9.6	11.2	11.2
Interconnect: Chip-to-Chip *(note4)*							
Memory (DDR/HBM), Gb/s	3.2/2.0	4.8/4.0	5.2/4.0	6.4/4.0	6.4/4.0	7.8/6.4	8.0/6.4
Smart Phone / Laptop, Gb/s	50	100	100	100	200	200	200
High-performance, Gb/s	32	32	56	56	64	64	112
Interconnect: Pkg-to-Board							
Memory (DDR/HBM), Gb/s	6.4/3.2	8/6.4	8/6.4	9.6	12.8	16	25
Smart Phone / Laptop, Gb/s	50	100	100	100	200	200	200
High-performance, Gb/s	32	56	56	64	64	112	112

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 evolving from current to foreseeable future (6 years). In summary, the demand of power reduction continuously drives lower voltage and leakage current (on-chip feature size). Meanwhile, power density in high-performance applications tends to grow with emerging 3D and 2.xD packaging technologies. Particularly, both artificial intelligence and accelerator architectures involve heterogeneous devices/components, which further drives the increase in package pin count as well as dimensions.

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

a workflow and business processes to enable companies to assemble products from the marketplace. More interconnect options and specifications are to be detailed in the following sections.

1. Chiplet Trend

With the continuous decrease in on-chip feature size, high-end semiconductor manufacturing cost increases exponentially with chip dimensions due to significant yield impact, and therefore, instead of continuously growing, integrated circuits will be divided into "chiplets" with respect to functional regions to take full advantages of semiconductor technologies for the yield optimization and rely on package-level integration to deliver the most cost-effective system performance.

What makes chiplet design different from other SoC design methodologies that have existed for many years is that many of these new chiplet-based parts are putting together pieces that are made on different process technologies. So, for example, a chiplet design might link a 5 or 7 nm CPU with a 10 nm or 14nm I/O element over high-bandwidth die-to-die interconnect. The reason for making these kinds of changes gets to the very heart of some of the transformational developments now impacting the semiconductor business. First, as has been widely discussed, traditional Moore's Law advances in shrinking transistor size have slowed down tremendously, making it difficult (and very expensive) to move all the elements inside a monolithic chip design down to smaller process geometries. Plus, even more importantly, it turns out that some important elements in today's chip designs, such as analog-based I/O and some memory technologies, actually perform worse (or simply the same, but at a significantly higher cost) in smaller feature-sized chips. Therefore, some semiconductor components are better off staying at larger technology nodes. In addition, the processing requirements for different types of workloads (such as AI acceleration) are expanding, leading to the need to combine even more types of processing technology onto a single component. Finally, there have been some important advances in chip packaging and interconnect technologies that are making the process of building these multi-part chiplets more efficient.

In order to efficiently deliver power and data to these various chiplets, leading semiconductor companies have been working on advancing their packaging technologies for the last several years. For instance, Intel announced some important new additions to its arsenal of chip packaging capabilities at Semicon West 2020, all designed to enable even more sophisticated, more flexible, and better yielding chiplet-based products in the years to come. They logically combine EMIB and Foveros, so called "Co-EMIB", that enables both 2D-horizontal and 3D-vertical connections of components in a single package. They also developed a technology called ODI (Omni-Directional Interconnect), which works through and across chips to provide the low-power and low-latency connections needed to attain performance closer to monolithic chip designs, together with a new version of their AIB (Advanced Interface Bus) standard called MDIO that provides the physical layer connect for die-to-die connections used in EMIB.

On the other hand, AMD has been successfully implementing the chiplets concept into their EPYC-family product for years, by leveraging existing organic build-up substrate technology and off-the-shelf die-to-die interconnect IPs. Though this represents a significant divergence from traditional semiconductor advances, it's become abundantly clear that the future of the semiconductor industry is going to be driven by chiplets. With more and more similar implementations, there's no question that the flexibility that chiplets enable is going to be critically important for advances in semiconductors and computing overall. As a result, heterogeneous packaging solutions will dictate chip and system architectures, and become essential to the advancement of the semiconductor industry.

2. On-package interconnections

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

In addition, to take full advantage of multi-chip packaging, it is critical to provide high bandwidth, low latency connections among functional components. Specifically, for connecting multicore processor die with stacked memory dies, point-to-point interconnections are needed and the number of memory dies will be proportional to the number of cores on the processor die. Conservatively assuming that core counts scale by a factor of 1.4X per

generation, 1.4 times as many memory dies need to be accommodated per generation in the SiP. Simultaneously, if we assume that advances in the stacked memory technologies enable twice as many data bits to be delivered per generation and assuming that the clock rate on the processor-memory link remain unchanged, the number of bit links between the multicore die and the stacked memory dies will have to grow by a factor of 2.8X with each process generation.

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

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

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

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

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

Short-term (0 - 5 years):

As mentioned above, driven by the demands of on-package performance and functionality scaling, high-bandwidth interconnects are experiencing an explosive growth. Both I/O speed and density 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.

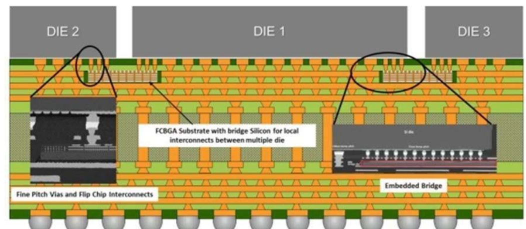


Figure 1: Embedded Multi-die Interconnect Bridge (EMIB) [1]

- **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.

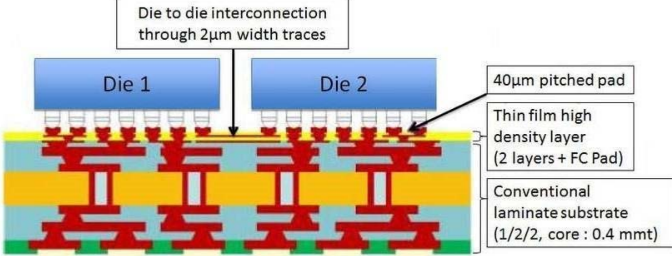


Figure 2: Emerging high-density organic substrate [2]

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

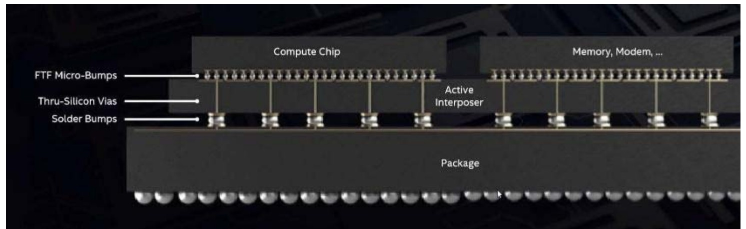


Figure 3: FOVEROS 3D chip stacking by Intel [3]

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

Longer term (5 – 15 years):

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

- Integrated photonics with polymer waveguides and improved optical transceiver stability;
- Plasmonic interconnections;
- Optical vias for 3D integration;

- Other technologies, protocol-specific, such as embedded components.

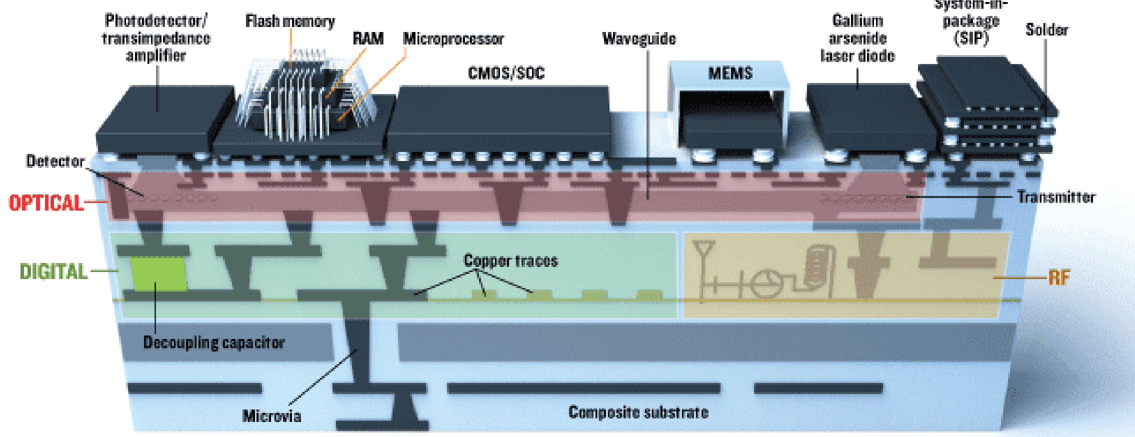


Figure 4: 3D heterogeneous integration

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

Connections to off-package interfaces and DRAM controllers on the SiP substrate can continue to rely on the PCIe standard, and the evolution path for multi-lane PCIe 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 requirements 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.

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, since techniques like wavelength division multiplexing can be used to implement several connections concurrently on a single photonic link. Here again, limiting wavelength drifts becomes 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 for solutions for PCIe Gen5, which is a clear indication of package-level system integration advancement. PCIe Gen5 is expected to carry 32Gbps per data channel without changing the Tx/Rx specifications. IBM and Amphenol Corporation jointly developed a new PCIe connector and demonstrated PCIe Gen5 bandwidth in early 2018, which significantly accelerates the availability of the new standard. The upcoming PCIe Gen6 will further double the data rate to 64Gbps by adopting PAM4 signaling.

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

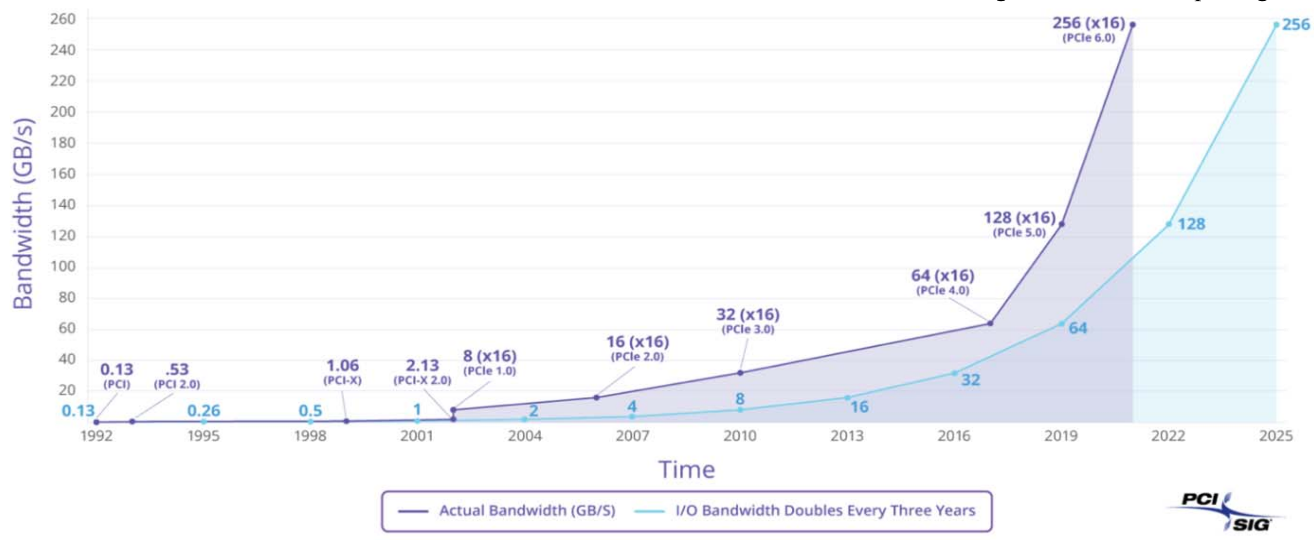


Figure 5: PCIe standard evolution [5]

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

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

- **Electrical/optical “flyover” cabling:**

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

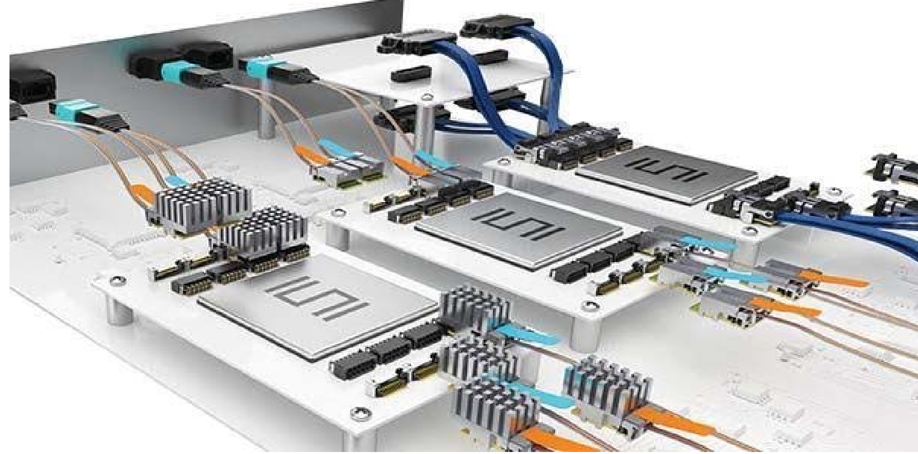


Figure 6: Electrical/optical “flyover” cabling [6]

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 vias 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 for excessive loss, ISSI, and cross-talk. For data rates beyond 50Gbps, PAM4 signaling will prevail, for much lower Nyquist frequency.

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 and 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

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. Smith R., "PCI Express Bandwidth to Be Doubled Again: PCIe 6.0 Announced, Spec to Land in 2021", in AnandTech, June 18, 2019
6. Samtec Micro FireFly Flyover System (samtec.com)

Section 3: Thermal Management

The Longevity of Moore's law has been under constant debate [1]. Many packaging experts still believe that Moore's law can hold good for at least the next 10 years due to recent technological efforts to combine functionalities horizontally and vertically into a single-chip module, commonly referred to as heterogeneous integration (HI). HI is the basis of future-generation computer systems. It refers to the assembly and placing of multiple separately manufactured components onto a single module to improve functionality and enhance operating characteristics. In such application, heat sinks are typically designed to cool multiple dies with different powers, sizes, heights, and maximum allowable temperatures in specified arrangements on electronic boards simultaneously. For several decades, electronic industries and relevant academic research communities globally have made constant efforts to develop and commercialize Moore's law [1,2], leading to many technological breakthroughs and revolutions in packaging technologies and skyrocketing thermal management challenges. These challenges are attributed to a decrease in device size and increased power consumption. With simultaneous shrinkage in size and increased heat dissipation, the flux values have increased exponentially, yet the required temperature difference coined as the thermal budget for thermal management is also reduced or unchanged, making it even more challenging for effective thermal management. Thermal management of electronic devices is important to improve the reliability of devices by maintaining silicon junction temperatures below critical safe operating temperatures. While the academic efforts on optimization of thermal solutions have been predominantly streamlined to theoretical studies, the constraints on manufacturability and reliability are not widely considered part of the optimization problem. Furthermore, owing to the multiple hotspot boundary conditions, the physics based computational modeling becomes more complex due to loss of symmetry. The challenges due to multiple hotspots will be discussed in this section, and are further analyzed in Chapter 20.

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

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

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

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

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

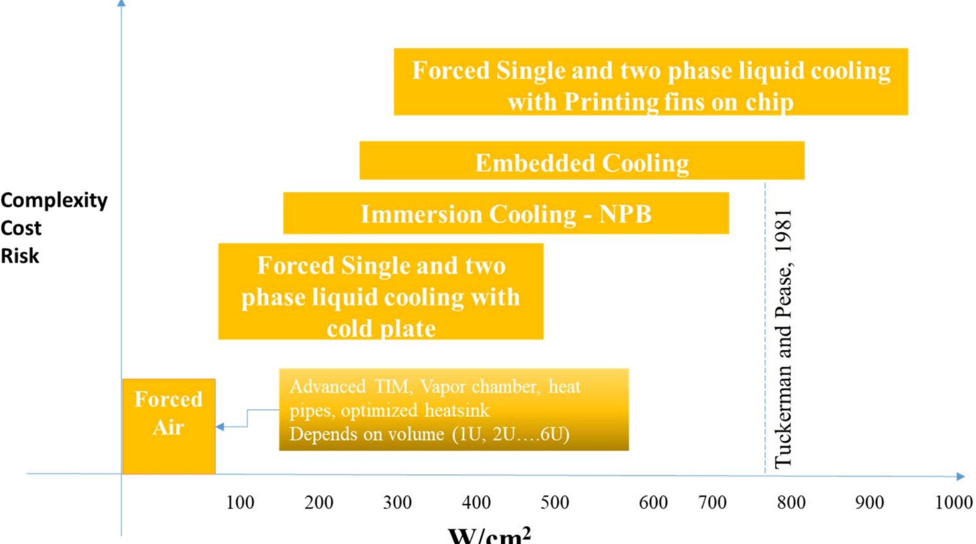


Figure 1: Heat transfer limits of different thermal management technologies

The associated complexities and risks in implementing each of these technologies are marked on the ordinate of Figure 1. In practice for today's ongoing data center power demand, the most common cooling method takes the form of air-cooling [3]. Air-cooling is often claimed to be the most reliable technology that is of practical interest. However, in practice, air cooling suffers from low heat transfer performance and acoustic-related problems. The low heat transfer performance is attributed to low heat transfer coefficient, high inlet air temperature, and low thermal spreading due to the requirement of a larger heat sink base area. The noise issue is addressed by imposing an engineering constraint on the maximum fan velocity [4]. To address the thermal spreading issue, technologies like vapor chambers integrated with heat pipes are employed. The acoustic constraints and the allowable pressure drop dictate the limit on the maximum airflow in the system. The constraint on minimum inlet air temperature is dictated by the Telcordia GR63 standard and ASHRAE for telecom application (55°C) and data center application (45°C), respectively. With regard to easier manufacturing, plate-fin parallel-channel heat sinks are still the most widely adopted fin design. With most of the stated factors associated with air cooling having reached standard limits, only the geometric optimization of air-cooled heat sinks could help the research community define the limits of the technology under practice. Unlike the air-cooled heat sink modules, the liquid-cooled heat sinks (commonly referred to as cold plates) can perform exceptionally well due to the superior thermal properties of liquid water. However, the superior thermal performance balances with the reliability and technological risks associated with using water closer to the circuit (e.g. unfavorable dielectric strength). In a liquid-cooled module, the heat sink resistance is 10-fold lower than the spreading and the TIM counterparts.

For a highly non-uniform and intense hotspot power map, there are two possible thermal solutions:

- Highly efficient thermal spreading by employing advanced vapor chamber thermal solutions
- Embedding the cooling solutions directly onto the die (hotspot-targeted)

The highly non-uniform chip power map [5] from a mechanical perspective leads to high warpage, which becomes a critical challenge. This issue also pertains to multi-chip modules. Localized heat (hotspot) is one of the significant influencers to both parametric and catastrophic reliability issues on electronic devices. Emerging chip hotspots are a major reliability concern, with heat fluxes as much as 5-6 times greater than those found elsewhere on the chip. Chip hotspots also augment thermo-mechanical stress [6] at chip-package interfaces, leading to failures during cycling. Highly localized transient chip cooling is technically challenging and costly. Furthermore, the non-uniform distribution of the localized zones leads to a much more substantial impact on the mechanical stability of the package [7]. Each material in a package responds to temperature changes differently according to its thermal expansion coefficient and generates thermomechanical stress between different materials due to a mismatch in thermal and mechanical properties. Though advanced vapor chamber heat spreading solutions can help minimize the thermal spreading resistance, they still demand a thermal interface material (TIM) between the chip and the spreader. Eliminating the TIM and other sequential conduction resistances from the chip to the coolant becomes critical from a thermomechanical perspective. In the case of multi chip modules, significant challenges involve the TIM design.

Tuckerman and Pease [8] dissipated heat flux up to 790 W/cm^2 for a substrate temperature rise of 71°C by etching a microchannel on the backside of the active device. However, there are practical limitations, challenges, and risks in implementing the work in practice and it requires a massive change in the existing infrastructure. Understanding the scaling limits of air and liquid cooling is essential to understand better the potential of embedding the cooling solutions onto the silicon die. Figures 2(a) and 2(b) show the limits of air and liquid cooling with increasing area ratio (AR = ratio of area of the spreader to the area of the chip). AR=1 translates to a scenario of embedded cooling. The results from the scaling analysis make it evident that the liquid cooling with AR=1 has the minimum total external resistance. Hence liquid cooling with printed fins on-chip with no additional spreader area ideally can extend the heat transfer limit of single-phase liquid cooling. However, air cooling requires a large area for effective heat dissipation and does not perform well when embedded.

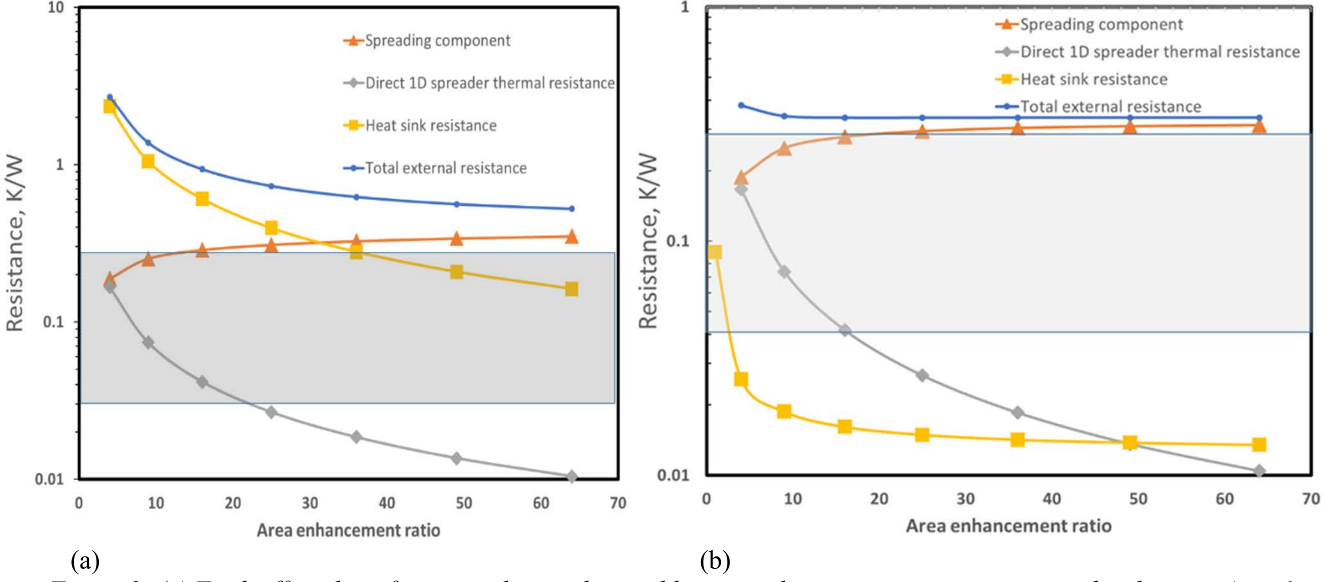


Figure 2: (a) Tradeoff analysis for air cooling with variable area enhancement ratios, assuming the chip size Ac = 1cm^2, metal conductivity, k = 150W/mK, spreader thickness t = 10mm, wch = 1mm, hch=25mm, Vair= 10 m/s. This figure excludes the TIM resistance. (b) Tradeoff analysis for a single phase liquid cooling scenario for a chip size Ac = 1 cm^2, metal conductivity k = 150 W/mK, Spreader thickness t = 10 mm, wch = 0.05 mm, hch=0.3 mm, Q= 0.5 liters/min. The plot excludes the TIM resistance. [9]

With growing heterogeneous integration of multiple functionalities onto a single chip, there are growing regions of intense hotspots. 3D printing of fins onto the chip [10,11] is seen to be a potential solution to mitigate hotspots and achieve a minimum of the maximum chip temperature. Vahideh et al [7] numerically demonstrated single-phase heat transfer enhancement of a liquid water-based cooling system employing printing fins directly onto a chip. The fin material was pure silver. The authors demonstrated that careful optimization of the pin fin profile and shape could lead to a total chip-to-ambient thermal resistance of 0.26°C-cm^2/W (compared to a separable cold plate of 0.58 °C-cm^2/W) for a 4cm^2 uniformly heated chip under a constrained pressure drop [11]. The same technique was extended by Vahideh et al [10] to a chip with simultaneous hotspots and background heating conditions for a minimum total case-to-fluid resistance of 0.21 °C-cm^2/W for four 4cm^2 chips under similar flow conditions. The rationale behind the low resistance was attributed to the fact that each core/hotspot had a dedicated inlet on top, thus minimizing the total resistance's caloric resistance component.

While recent efforts to explore two-phase liquid cooling using separable cold plates[12,13] for data center applications have shown remarkable performance in mitigating high heat fluxes with significant energy savings, two-phase cooling is essentially recommended for direct chip cooling using a dielectric coolant (Technology 1).

Two-phase immersion cooling could be a potential solution for highly intense and non-uniform heated devices. Boiling enhancement coatings on top of the heated surface are seen to increase the critical heat flux (CHF) and the boiling heat transfer coefficient (h_b) simultaneously [14]. 3D printing of fin and microporous structures can be beneficial in promoting nucleation and enabling rewetting of the heated surface. The combination of fin and microstructures has been highly beneficial in the simultaneous improvement of CHF and h_b [15]. Wong et al. [16] printed lattice structures on top of a heated surface, performed pool boiling experiments, and demonstrated a critical heat flux of 107 W/cm^2 corresponding to a boiling heat transfer coefficient of 1.5 W/cm^{2}K.

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

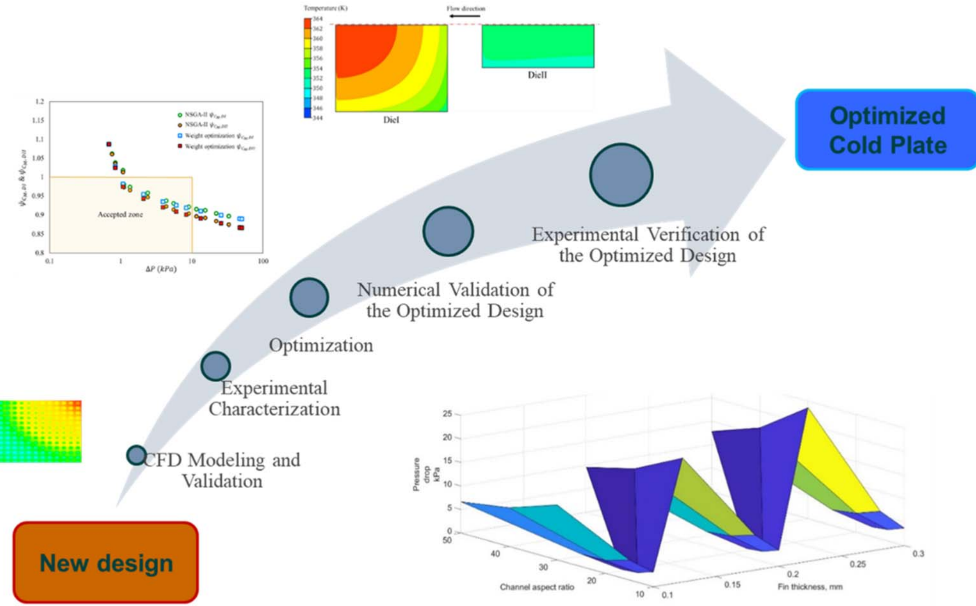


Figure 3: Optimization approach for the cold plate/ heat sink thermal solution

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

Challenges in computational modeling

As the hotspots become more non-uniform in a chip, the symmetry is lost, and hence the computational modeling of the thermal solutions becomes more challenging. Recently, Sharma et al. [5] published an embedded cooling-based thermal management technique for a highly non-uniform thermal power map. The authors proposed a simplified modeling approach for the thermal management solution. However, simulating the complete fluid flow and heat transfer requires detailed modeling of the entire module. Srikanth et al. [17] recently demonstrated an optimal arrangement of the non-uniformly distributed hotspot in a 3D chip package. The optimization was assisted by complete physics-based modeling due to the loss in symmetry. The modeling gets even more challenging when dynamic/transient non-uniform power maps evolve.

References

[1] R. Courtland, "Transistors could stop shrinking in 2021," IEEE Spectrum, vol. 53, no. 9, pp. 9–11, Sep. 2016, doi: 10.1109/MSPEC.2016.7551335.

[2] dcadmin, "2015 International Technology Roadmap for Semiconductors (ITRS)," Semiconductor Industry Association. https://www.semiconductors.org/resources/2015-international-technology-roadmap-for-semiconductors-itrs/ (accessed Oct. 15, 2020).

[3] "High Powered Chip Cooling -- Air and Beyond," Electronics Cooling, Aug. 09, 2005. https://www.electronics-cooling.com/2005/08/high-powered-chip-cooling-air-and-beyond/ (accessed Oct. 15, 2020).

[4] "G. Refai-Ahmed, H. Do, Y.Hadad, S. Rangarajan, B. Sammakia, T. Cader `` ESTABLISHING THERMAL AIR-COOLED LIMIT for High Performance Electronics Devices'', In 2020 IEEE 22nd Electronics Packaging Technology Conference (EPTC), IEEE, 2020".

[5] C. S. Sharma, M. K. Tiwari, and D. Poulikakos, "A simplified approach to hotspot alleviation in microprocessors," Applied Thermal Engineering, vol. 93, pp. 1314–1323, 2016.

[6] © Stanford University, Stanford, and California 94305, "Limits of hotspot detection and prediction in microprocesors [sic]." https://purl.stanford.edu/jc805vt4819 (accessed Oct. 28, 2021).

[7] J.-H. Choy, V. Sukharev, and A. Kteyan, "Advanced methodology for assessing chip package interaction effects on chip performance and reliability after chip assembly and during chip operation," Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena, vol. 38, no. 6, p. 063205, 2020.

[8] D. B. Tuckerman and R. F. W. Pease, "High-performance heat sinking for VLSI," IEEE Electron Device Letters, vol. 2, no. 5, pp. 126–129, May 1981, doi: 10.1109/EDL.1981.25367.

[9] Rangarajan s, Schiffres S, Sammakia Bahgat, “Scaling limits, challenges, opportunities in embedded cooling,” Encyclopedia of Thermal Packaging: Thermal Packaging Tools/Ed.-in-chief Avram Bar-Cohen. World Scientific.2021.
[10] V. Radmard et al., “Performance Analysis of Impinging Chip-Attached Micro Pin Fin Direct Liquid Cooling Package for Hotspot Targeted Applications,” in 2021 20th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (iTherm), Jun. 2021, pp. 220–228. doi: 10.1109/ITherm51669.2021.9503295.
[11] V. Radmard et al., “Multi-objective optimization of a chip-attached micro pin fin liquid cooling system,” Applied Thermal Engineering, vol. 195, p. 117187, Aug. 2021, doi: 10.1016/j.applthermaleng.2021.117187.
[12] “Khalili, Sadegh, Srikanth Rangarajan, Vadim Gektin, Husam Alissa, and Bahgat Sammakia. ‘An Experimental Investigation on the Fluid Distribution in a Two-Phase Cooled Rack Under Steady and Transient Information Technology Loads.’ Journal of Electronic Packaging 142, no. 4 (2020).”.
[13] “Hoang, Cong Hiep, Srikanth Rangarajan, Sadegh Khalili, Bharath Ramakrisnan, Vahideh Radmard, Yaser Hadad, Scott Schiffres, and Bahgat Sammakia. ‘Hybrid microchannel/multi-jet two-phase heat sink: A benchmark and geometry optimization study of commercial product.’ International Journal of Heat and Mass Transfer 169 (2021): 120920.”.
[14] G. Liang and I. Mudawar, “Review of pool boiling enhancement by surface modification,” International Journal of Heat and Mass Transfer, vol. 128, pp. 892–933, Jan. 2019, doi: 10.1016/j.ijheatmasstransfer.2018.09.026.
[15] C. M. Patil and S. G. Kandlikar, “Pool boiling enhancement through microporous coatings selectively electrodeposited on fin tops of open microchannels,” International Journal of Heat and Mass Transfer, vol. 79, pp. 816–828, Dec. 2014, doi: 10.1016/j.ijheatmasstransfer.2014.08.063.
[16] K. K. Wong and K. C. Leong, “Saturated pool boiling enhancement using porous lattice structures produced by Selective Laser Melting,” International Journal of Heat and Mass Transfer, vol. 121, pp. 46–63, 2018.
[17] S. Rangarajan, L. Choobineh, and B. Sammakia, “Supervised machine-learning approach for the optimal arrangement of active hotspots in three-dimensional integrated circuits,” IEEE Transactions on Components, Packaging and Manufacturing Technology, pp. 1–1, 2021, doi: 10.1109/TCPMT.2021.3109662.

Section 4: Mechanical Requirements

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

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 assembly 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 becames essential for successful product engineering, from design to development qualification and volume manufacturing.

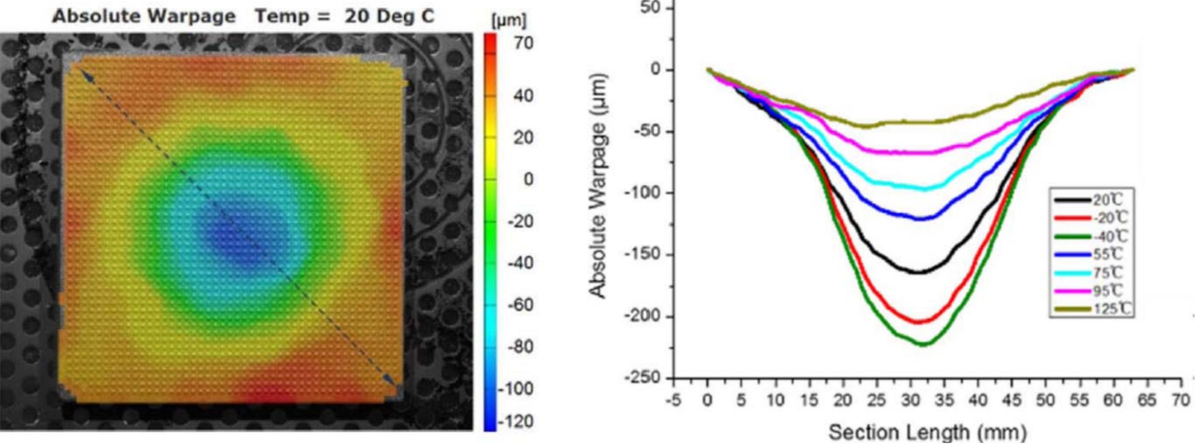


Figure 1. Thermal deformation (warpage) of a high-performance flip chip PBGA package

For 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 modeling and simulation tools? What are the metrology tools?

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

Warpage Metrology tools available today:

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

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

- Optical non-contact method
- No-surface-treatment 3D metrology tool
- Measurement accuracy: < 1 um
- Special resolution: < 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 from -55C to ~250C

Package Stress and contribution to warpage

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

Package stacking and die stacking add layers to the assembly to shorten interconnect and reduce size. Package-on-package and other stacked structures would require insight in their warpage behavior and stress management. A method to help visualize the warpage during reflow is called Digital Image Correlation (DIC). This tool uses stereo cameras pointed into a chamber where the part to be studied is placed. The chamber goes through a temperature cycle from 25°C to 245°C. This data is captured and a deformation map is created to show the movement of the points that were captured.

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.

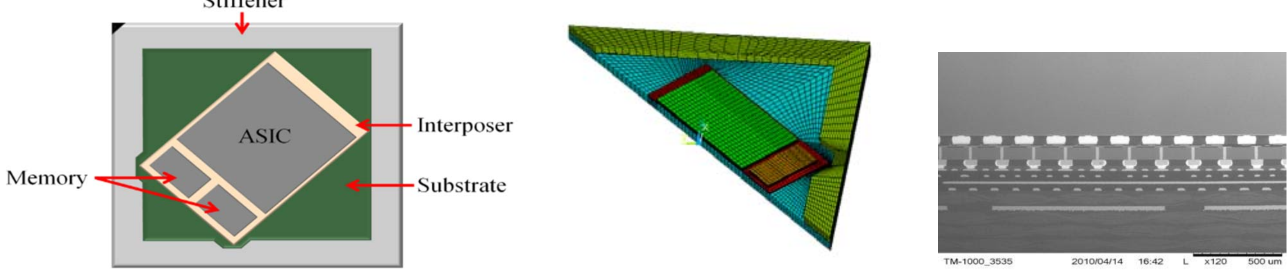


Figure 2. FEA model of 2.5D package and micro bump cross section

Stress Management and Chip Package Interaction

Chip package interaction (CPI) refers to failure modes such as delamination in low-k dielectrics or in solder bump failure due to mechanical stress from temperature excursion 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-node ICs, and develops new package types, stress management and CPI will continue to occupy an essential part of the engineering community tool kit.

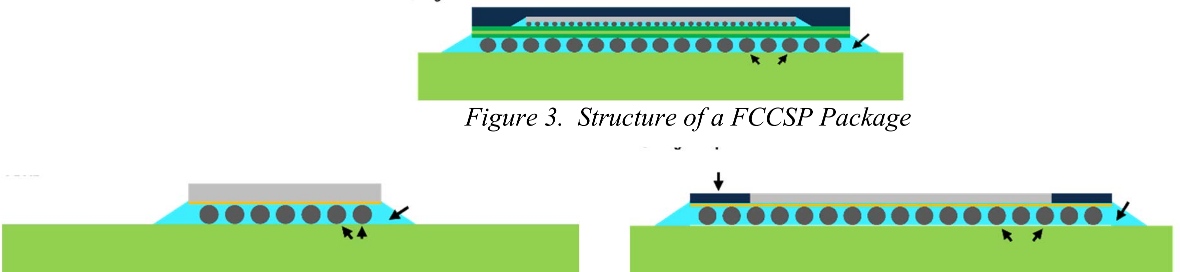


Figure 3. Structure of a FCCSP Package

Figure 4: Structure of a WLP Package

Figure 5: Structure of a FOWLP Package

Table 1. Warpage allowance vs interconnect pitch

	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
HPC	**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

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

interconnect pitches remaining at 1.00mm to 0.8mm. The interconnect pitch is being driven by the complexity of escaping these large packages.

The warpage allowance is governed by interconnect pitch reduction and package size. The smaller the pitch, the lower the warpage allowance. As the pitch gets smaller, the interconnect will also get smaller, so for a solder ball interconnect, the ball size will be smaller so the allowable warpage and co-planarity will be reduced to ensure proper assembly yields. The package size has the same effect: as package sizes get larger, the warpage will increase so more consideration will be given to materials and structures to reduce the warpage during reflow to ensure manufacturing yields.

Future Challenges and Opportunities

There has been a trend for HPC packages to use larger and larger substrates with multiple dies. The current packages being developed for use in the upcoming year range from 75mm to 100mm square. These packages are basically large SIPs with multiple ASIC/CPU, memory and IO devices. Putting these functions together onto a substrate reduces latencies and allow for more efficiencies in the system. The challenge comes when this large substrate package gets integrated into the next-level assembly. As the package gets larger, warpage control gets more important. If the package warpage exceeds a total of 0.006" (per JEDEC JESD22 – B112B) the instances of opens and head-in-pillow (HiP) will increase. Figure 6 shows how warpage causes such a fail.

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

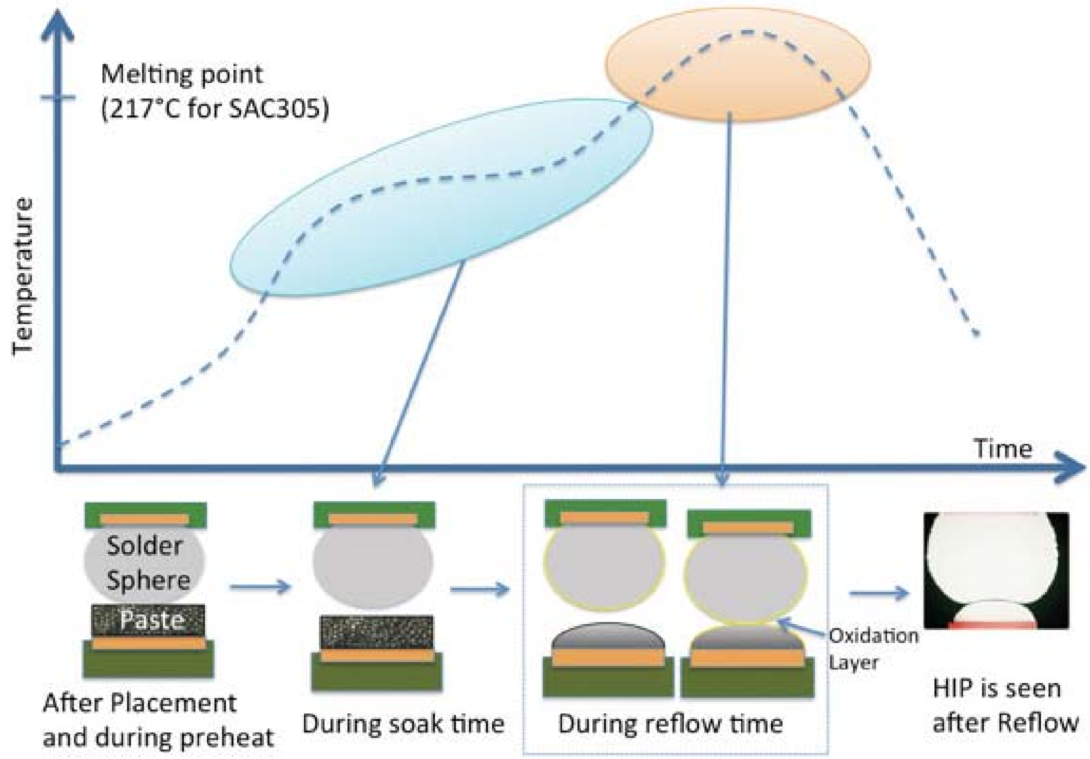


Figure 6. Head in Pillow failure mechanism (source: Circuit Assembly 19 May 2017)

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 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. Table 2: Warpage Allowance across two market segments

Section 5: Thinning and Singulation

1. Introduction to Thinning

All silicon wafers start out at the foundry at between 0.7 to 0.8mm in thickness. To fit inside today's low-profile single and multichip integrated packages, wafers are thinned by processes where the wafer backside is removed leaving the active frontside at a fraction of the original thickness. Thinning semiconductor wafers is a widespread process using abrasive rotary grinding and polishing wheels to ultrathin thicknesses with very good total thickness wafer variation across the entire 300mm wafer. Mechanical abrasive grinding of silicon wafers reaches a limit at around 20um and a gentler chemical removal is required to thin below 50um in the ultrathin realm. Chemical mechanical polishing, wet etching and dry etching are used as very smooth, stress-free and defect-free surfaces become imperative. The same thinning techniques are used for compound semiconductor wafers including silicon carbide, sapphire, InP, GaAs and others. According to Yole, the wafer thinning market activity is expected to increase to $135M in 2025 [1].

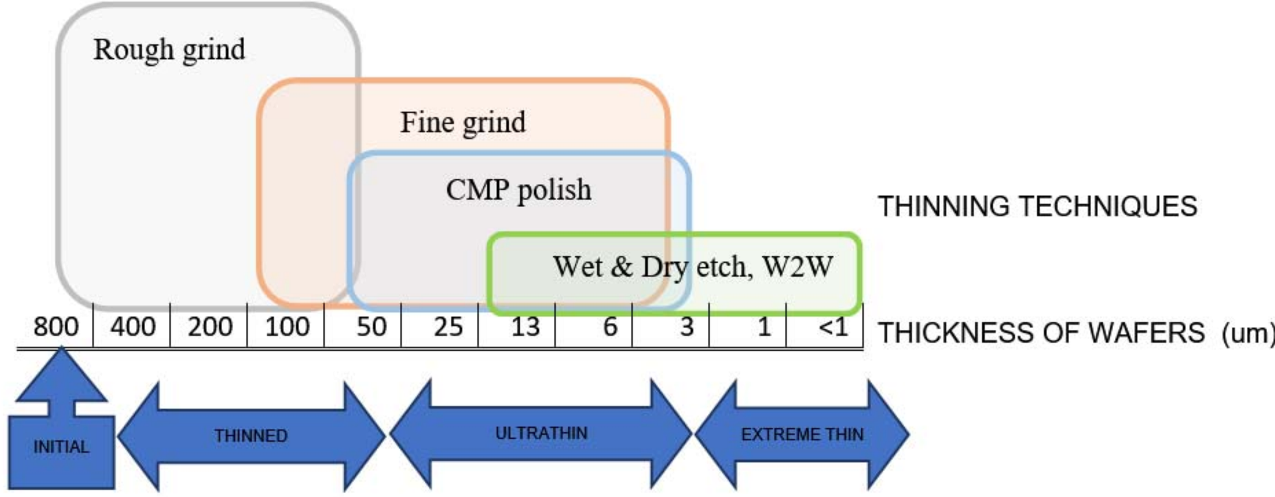


Figure 1. Thinning techniques vs. wafer thickness

Wafer thinning allows for more devices and functions to be fitted in a tight space to create the smart, lightweight and portable devices that are commonplace now. Additionally, thinning and miniaturization can improve and speed up device performance as well as improve thermal and power management. For wearables, the ultrathin stress-free dies at 30um and below allow for the devices to behave like fabric. For image sensors used in cameras, making dies extremely thin by flipping wafer backside to become frontside creates smaller pixel size and better camera performance. Sony and other suppliers have been making ultrathin image sensors in the 3-5um range for over a decade based on concepts patented by Kodak and Tower Semiconductor [2]. For memory dies, multiples of 50um die are interlayered and stacked, allowing astonishing levels of vertical integration. However, there is a push to go even thinner. The WOW Alliance at the Tokyo Institute of Technology comprised of Micron Memory, DISCO Corp and DAICEL Corp. is investigating memory chips to below 10um in thickness using Bumpless Wafer on wafer

(WOW) stacking [3]. They can achieve 3um thickness by fine grinding or by CMP but tests did show grinding defects were lower for the CMP process. One would think this is the limiting thickness for memory devices, as the removed backside is encroaching into the Depletion regions below the deep n-wells of the active regions.

2. Extremely thin dielets

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

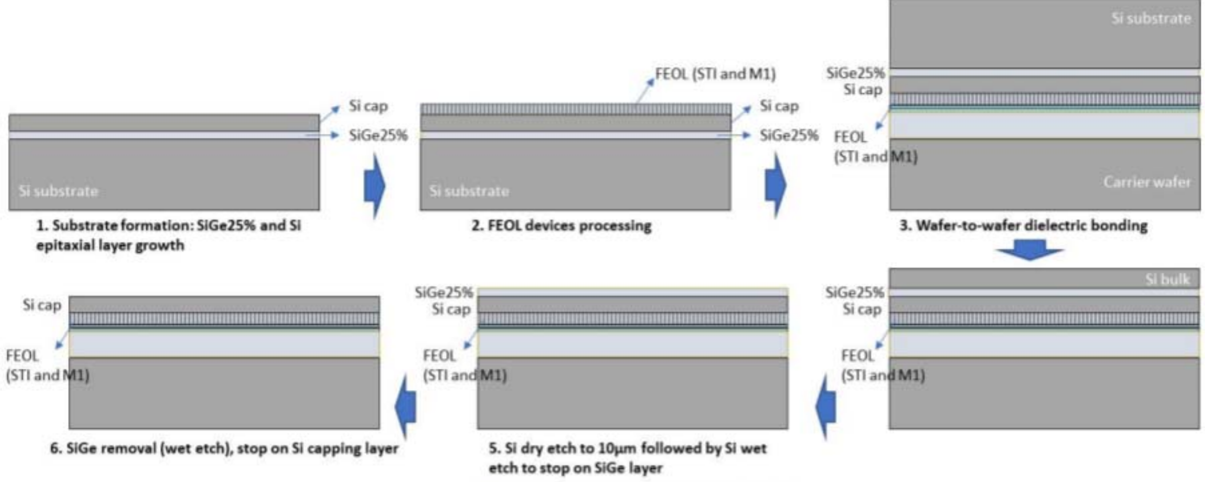


Figure 2. Process flow for making extreme thin dies, courtesy of A. Jourdain et al, imec.[4]

The approach by C. Bower and team at X-Display is a method of fabricating an extreme thin dielet in the 0.5 um thickness range. It does not require backside flipping to front side, but instead extreme thin dielets are fabricated by undercutting the die by chemical etching [6]. Individual die mesas on wafers are formed first, as shown in Figure 3. Subsequent wet etching is both the thinning and singulating process combined into one process step.

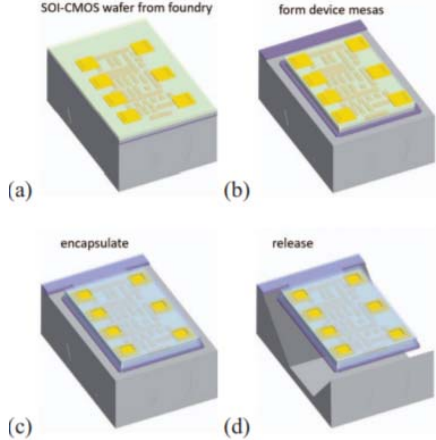


Figure 3: Process flow for making print-ready microICs, courtesy of C.Bower of X-Display[6]

3. Singulation of IC chips

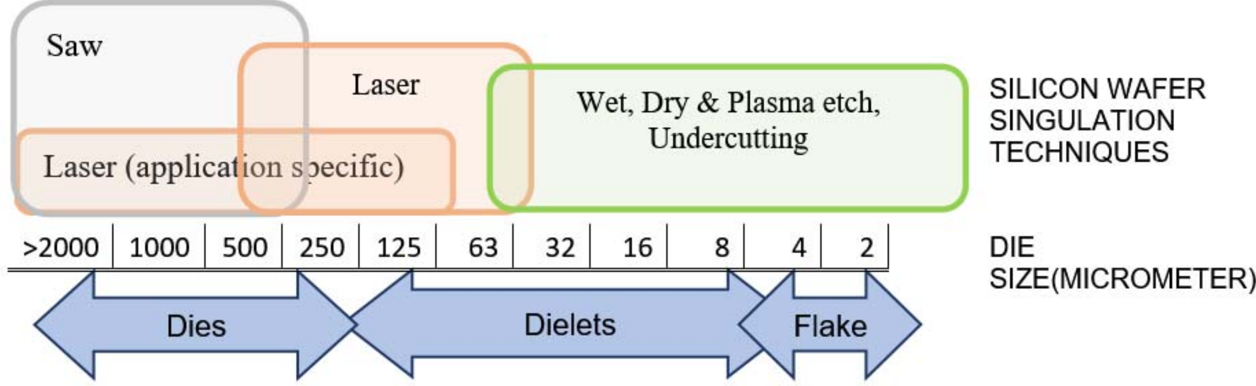


Figure 4. Singulation methods for silicon wafers

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

A new approach is revealed by C. Bower, et al [6], to undercut material under the die surface to fabricate dielets without grinding. A die mesa is formed and after undercutting; each dielet is held by a tether anchored to the mother wafer and ready to be snapped off by the picking action.

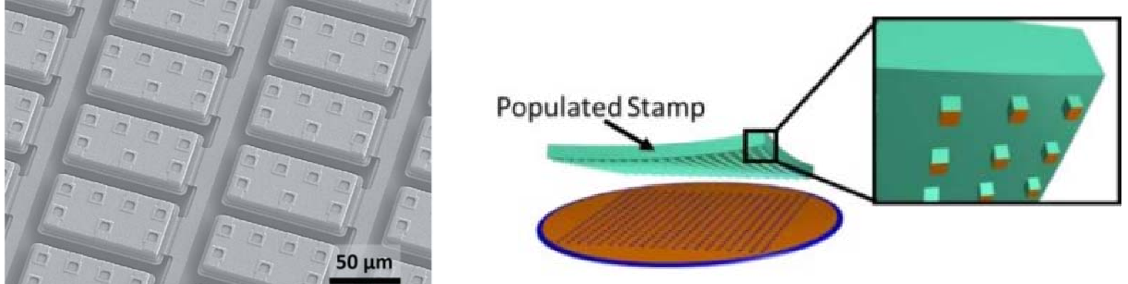


Figure 5. A scanning electron micrograph of 3.5 µm thick 50x90um silicon CMOS dielets [6] ready for picking with an elastomeric array tool [7]

4. Dielet Pick

Die pick is a challenging step for fragile ultrathin dies which are susceptible to mechanical damage. Picking with a traditional vacuum pick tool on the top of die with ejector pins underneath the taped die is no longer viable for ultrathin dies. Consequently, the "push and pick" action has been replaced by "peel and pick" action with the use of timed and graduated pushing and vacuuming. This is particularly neccessary for larger dies. The need for large quantities of MicroLEDs to be efficiently placed on large displays has established the micro transfer printing (µTP) process. The pick tool is made of elastomeric material PDMS with properties of softness and temporary stiction. Preparation at the wafer level through WLP must be developed and implemented to allow wafers to be µTP. This

includes patterning tethers and anchors to hold dies in place after undercutting. The tether of each dielet breaks off during pick and sticks to the tool. After transfer, it is released to stick to the receiving substrate, followed by interconnection to pads. Contact sites can multiply on the tool to pick multiples of dielets at once, tremendously increasing efficiency to >10000 per minute [6]. µTP scaling to larger dies for the future is a possibility. It is also possible for dielets to be harvested from a wafer using a liquid medium. E. Chow and team at PARC are developing a microassembler by making dielets suspended in liquid to self-sort, transport and self-orient to form programmed patterns. The arrangements are then transferred to a final substrate with contact stamping or an electrostatic roller belt [8].

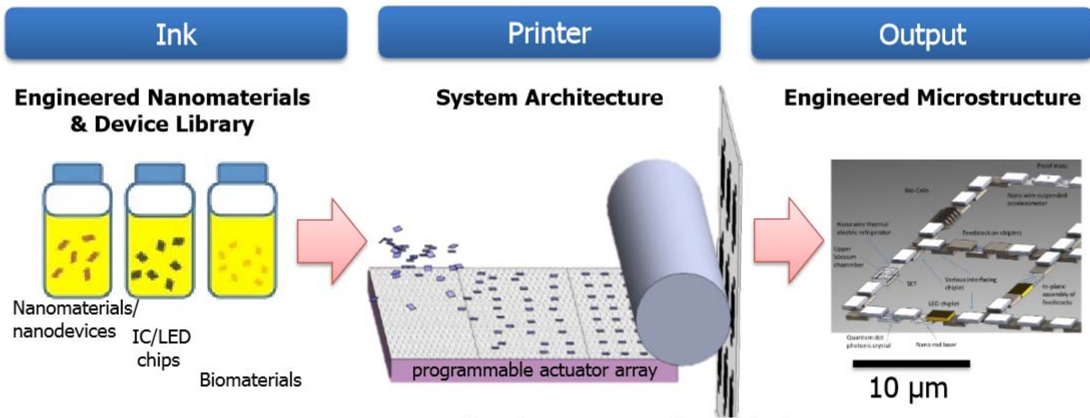


Figure 6. PARC Microassembler Printing concept for dielets; courtesy of E. Chow [8]

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

References

1. YOLE Report “Thinning Equipment Technology and Market Trends for Semiconductor Device, 2020 Market and Technology Report 2020”
2. US 6,168,965 B1 Date of Patent: Jan. 2, 2001 METHOD FOR MAKING BACKSIDE ILLUMINATED IMAGE SENSOR Inventors: Yacov Malinovich, Tivon; Ephie cited by examiner Koltin, Givat Elah, both of (IL)
3. Zhiwen Chen et al., “Ultra-Thinning of 20 nm-Node DRAMs down to 3 um for Wafer-on-Wafer (WOW) Applications”, 2021 IEEE 71st Electronic Components and Technology Conference (ECTC), pp 1131-1137 June 2021.
4. Jourdain et al., “Extreme Wafer Thinning and nano-TSV processing for 3D Heterogeneous Integration”, 2020 IEEE 70th ECTC
5. Jourdain et al., “Extreme thinning of Si wafers for via-last and multiwafer stacking applications”, 2018 IEEE 68th Electronic Components and Technology Conference (ECTC), pp1523-1530 May 2018
6. 6A. Bower et al., “High-brightness displays made with micro-transfer printed flip-chip microLEDs”, 2020 IEEE 70th Electronic Components and Technology Conference (ECTC), June 2020.
7. Gomez, K. Ghosal, M.A. Meitl, S. Bonafede, C. Prevatte, T. Moore, B. Raymond, D. Kneeburg, A. Fecioru, A.J. Trindade and C.A. Bower, “Process Capability and Elastomer Stamp Lifetime in Micro Transfer Printing,” 2016 IEEE 66th Electronic Components and Technology Conference (ECTC), pp. 680-687 (2016)
8. Chow, “Chiplet Micro-Assembly Printer” 2019 IEEE 69th Electronic Components and Technology Conference (ECTC). pp1312-1315 May 2019

Section 6: Wire Bonding

Wire bonding continues to be the most dominant interconnection technology used in the electronics 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, 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.

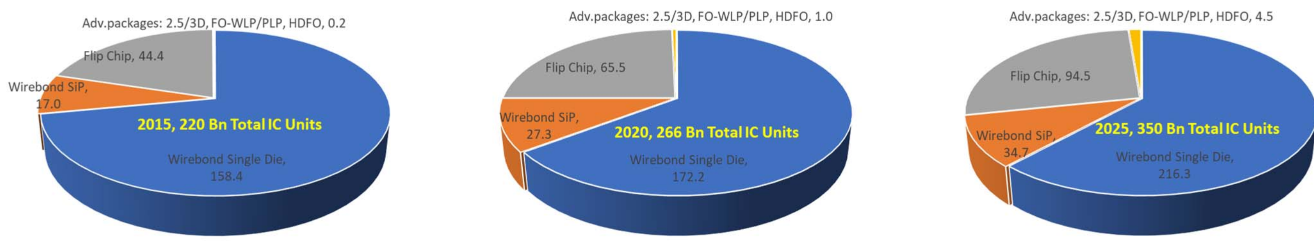


Figure 1: Semiconductor package unit growth by interconnect type. IC Package Shipment excluding Discrete, LED and Opto. (Source: Prismark May 2021)

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].

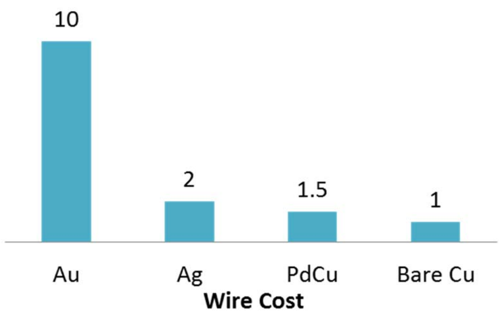


Figure 2: Relative cost for different types of wires, based on 25um diameter wire (Source: Heraeus and Tanaka). 94% Ag alloy is used in the comparison.

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

Another material cost reduction is the development of lower-cost substrates and leadframes, such as PPF (Pre-Plated Frame) QFN. PPF QFN offers reduced cost and simplified assembly by eliminating the 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 modeling 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].

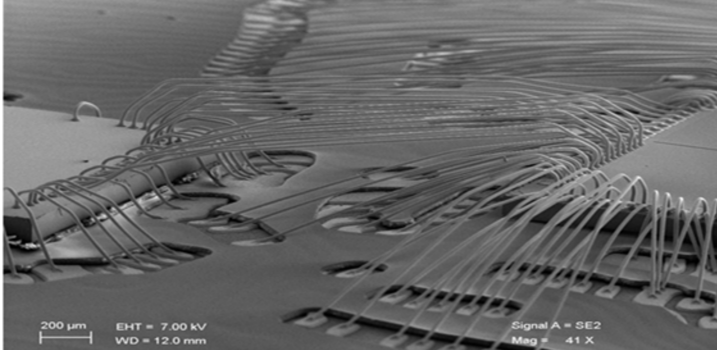


Figure 3. Advanced multiple tier looping and die-to-die bonding loops (K&S)

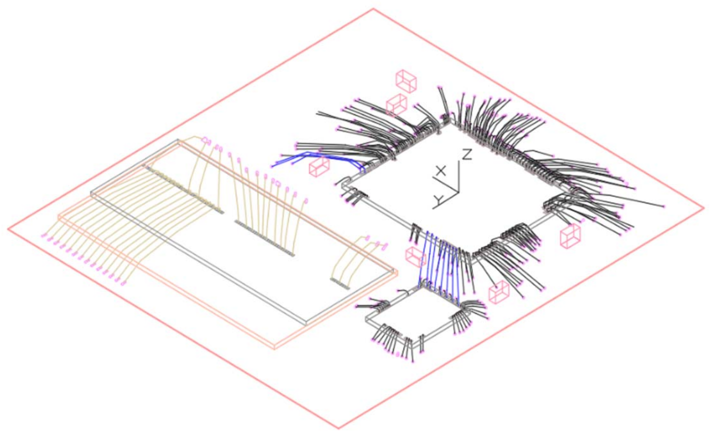


Figure 4. 3D loop design and clearance check software (K&S)

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]. The industry is driving toward finer SSB pitch capability.

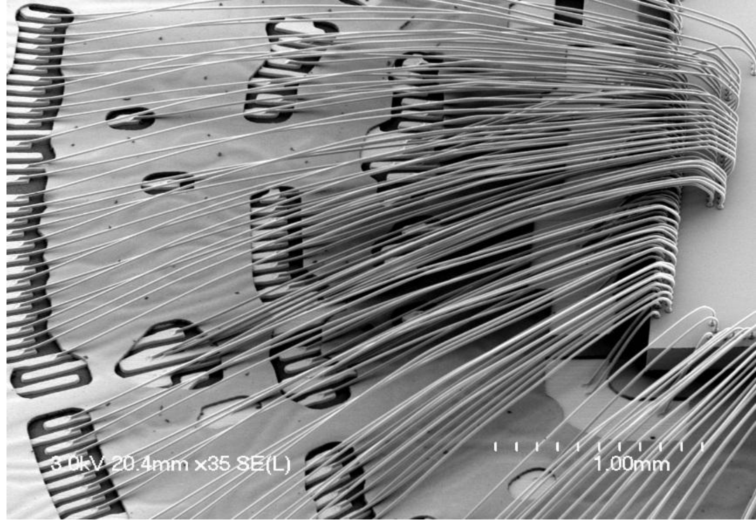


Figure 5. High density multi-tier package (K&S)

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

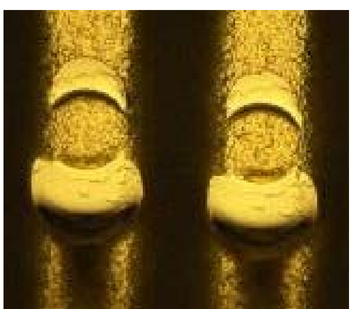


Figure 7. Bonding on 20um width lead fingers (K&S)

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

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

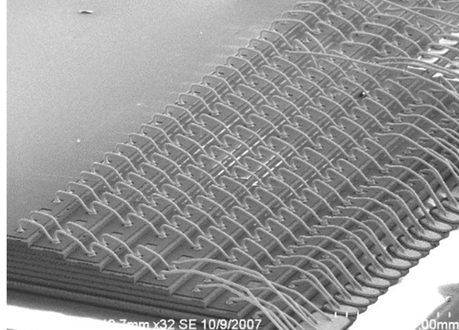
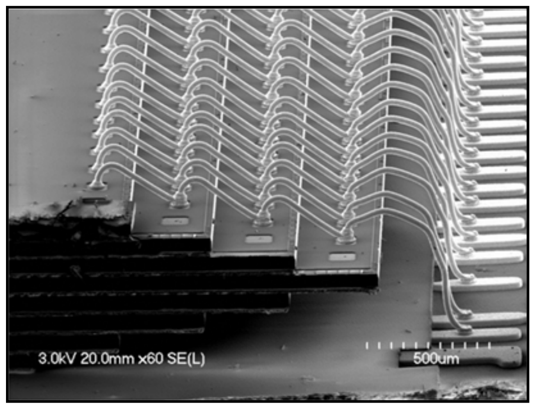


Figure 8. Example of a stacked-memory device (K&S)

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



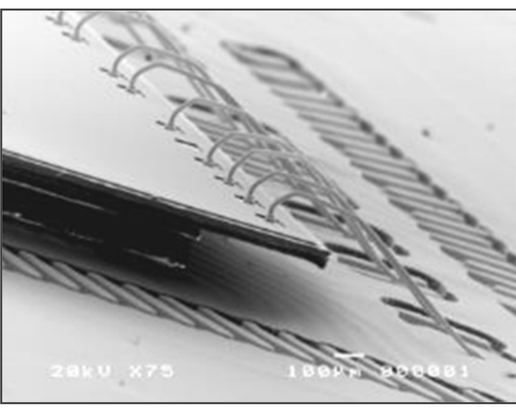
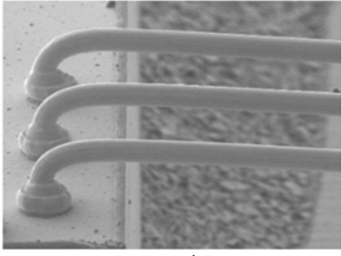
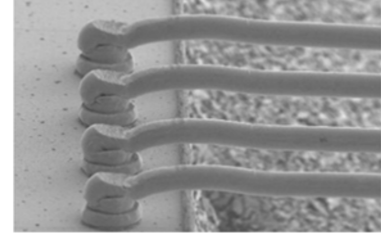


Figure 9. Examples of bonding on overhang die (K&S)

In order to achieve overall package height requirements for multi-stack memory packages, low loop heights of 100 um or 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.





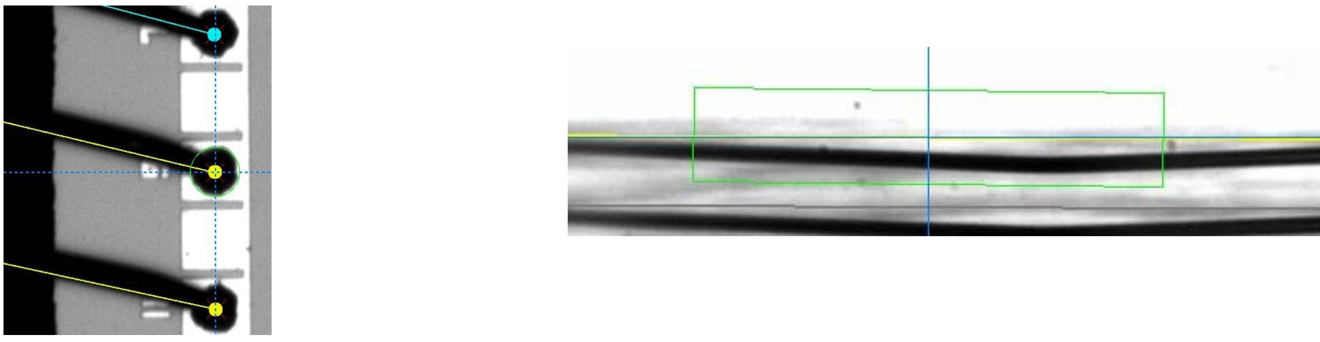
Normal Loop Compressed loop

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

Leading trends and technologies towards smart factory, autonomous wire bonder and I4.0

A key driver for the wire bonder roadmap is to support smart factory and industry 4.0 initiatives. New machine functionalities have been added to meet the desire for factory automation, real-time monitoring, closed-loop optimization and traceability. Similar to other AI applications, the goal for a smart factory with a fleet of wire bonders is to collect data from the equipment, and derive optimal actions from the data analytics. The key benefits of smart factory and smart wire bonders are higher productivity and efficiency, improved capability, portability, ease of setup and optimization, yield loss prevention and quality improvement. Some of the smart functionalities in today’s state-of-the-art wire bonders include:

- Smart process solutions using a response-based approach and closed-loop feedback [14]
- On-bonder automatic setup and calibration solutions to enhance portability and tool matching
- Auto-recovery features to improve bonder up time and reduce the need for operator assist
- On-bonder metrology using machine vision, electrical and servo system: PBI (Post bond inspection), automatic loop height, wire sway measurement. Figure 11 shows the bonder vision-based wire sway and ball diameter and bond placement measurement.
- Machine health monitoring and maintenance tracking features
- Traceability and connectivity: recording and streaming data, data analytics at the edge and on the cloud, and connected through a factory network (Figure 12).



a) Ball diameter and placements PBI b) Wire sway PBI

Figure 11. Using wire bonder vision system for post bond inspection (PBI) to detect defects early and improve yield.

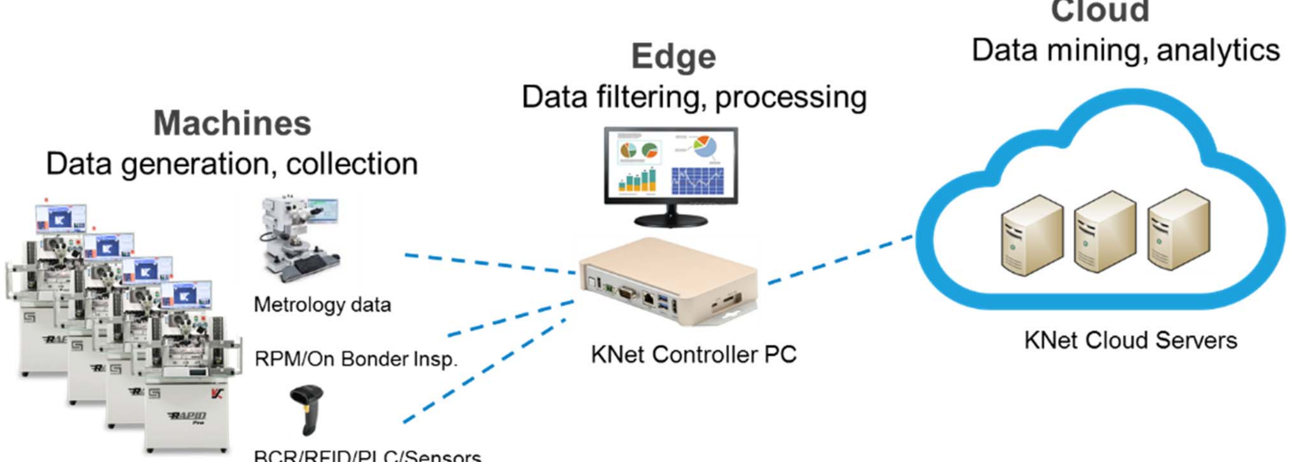


Figure 12. Improve factory efficiency, quality control, traceability using factory connectivity solutions and big data collection and analytics.

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.

Heavy Wire Bonding for Power Electronic Devices

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

Higher Temperature Capability

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

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.

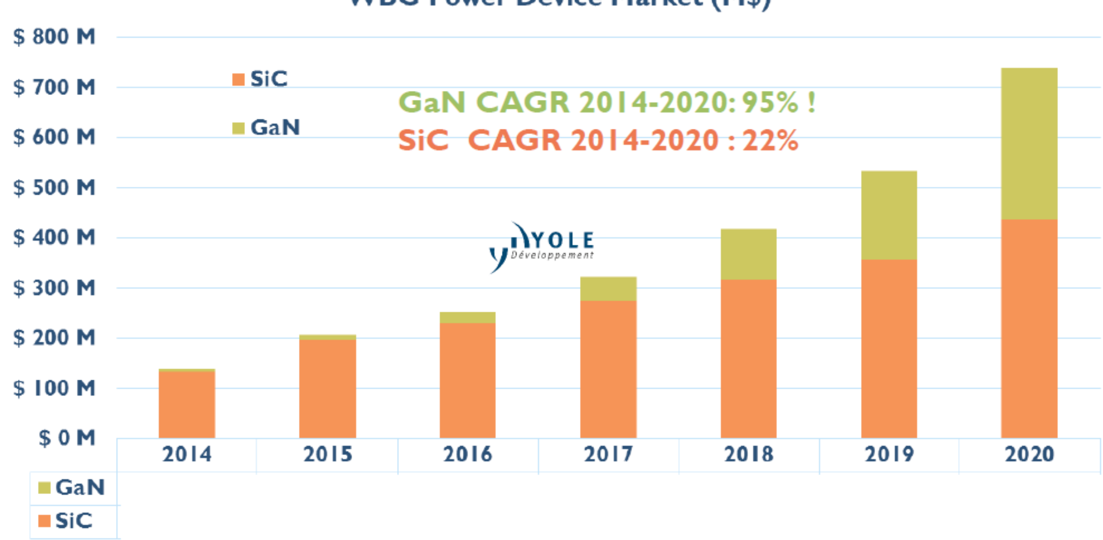


Figure 11: SiC and GaN Power Device Market. (Source Yole Research 2016)

Material Thickness Increase

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

Automotive Electrification

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

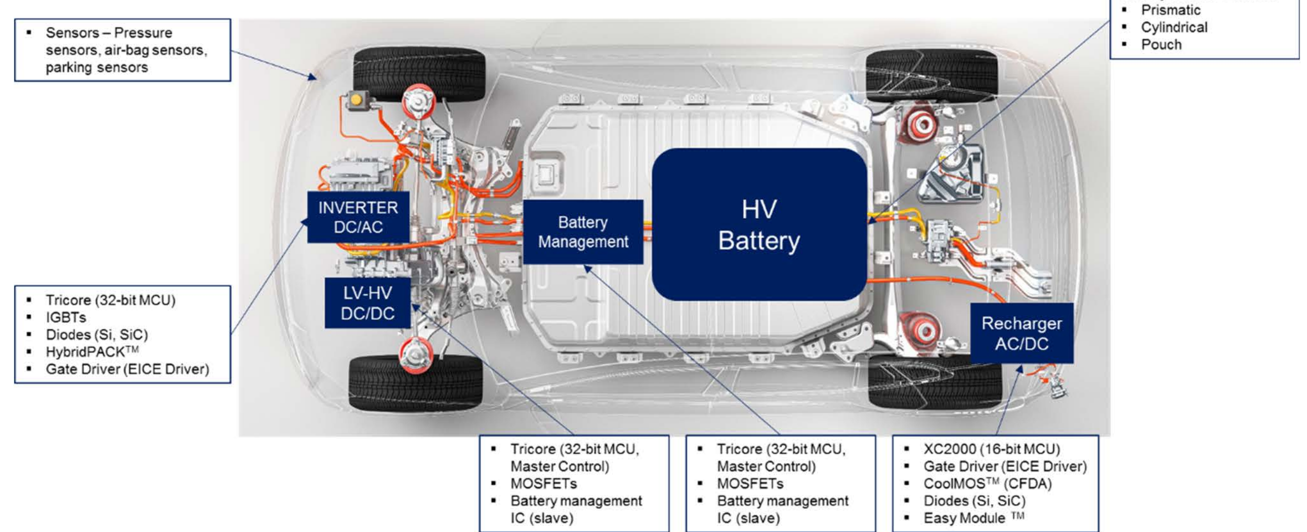


Figure 12: Examples of Automotive Wedge Bond applications [15]

Future Challenges and Roadmap towards Autonomous Wedge Bonder

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

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

Other Emerging Wire Bonding Applications

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

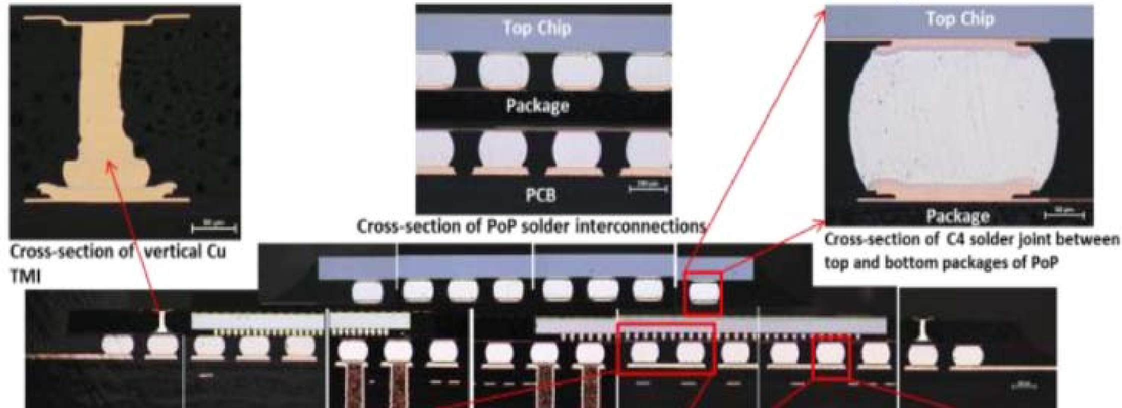
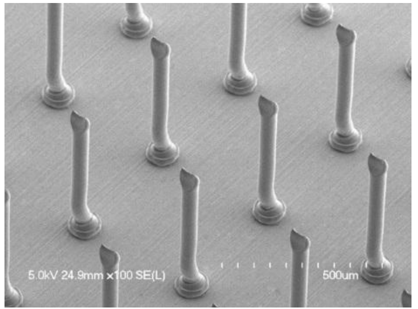
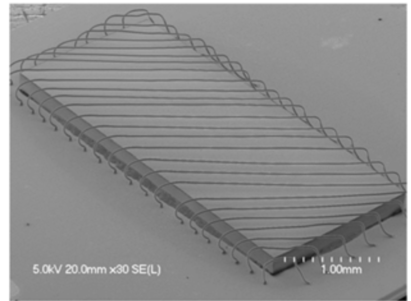


Figure 13. Cross section of assembled FOWLP package with wire bonded Vertical Interconnects [17] (Source: IME).

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





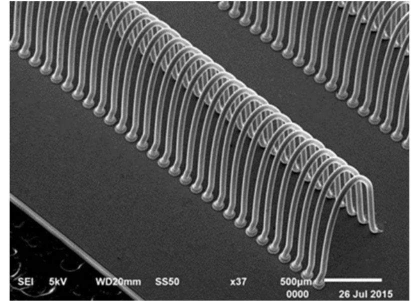


Figure 14. Wire bonded loops used for EMI shielding (Source: K&S).

REFERENCES

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

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

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

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

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

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

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

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

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

[10] F. Carson and M. Karnezos, "Advances in Stacked –Die Packaging", Global SMT and Packaging, March 2003. www.globalsmt.net.

[11] M. Karnezos, "3-D Packaging: Where All Technologies Come Together", Proceedings of 2004 IEEE/SEMI International Electronics Manufacturing Symposium, 2004.

[12] A. Shah et al., "Advances in Wire Bonding Technology for Overhang Applications", IMAPS Conference, Oct., 2016

[13] Hoet, J, and Hiebert S, "Automotive Apps are Driving Inspection Requirements for Advanced Nodes," Chip Scale Review, Volume 22, Number 4, 2018, pp. 8-9.

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

[15] G. Prasad, P. Klaerner, J. Fu "High Volume Manufacturing Requirements For EV Battery Interconnect And Power Management Applications", Charged Electrical Vehicles Magazine, Sept. 2nd Conference, 2020

[16] P. Klaerner "The advantages and limitations of wire bonding in electric vehicle applications", Charged Electrical Vehicles Magazine, Jul. 16th, 2019

[17] V. S. Rao et. al., "Development of High Density Fan Out Wafer Level Package (HD FOWLP) with multi-layer Fine Pitch RDK for Mobile Applications", Proceedings of IEEE Electronic Components and Technology Conference (ECTC), Las Vegas, NV, Jun 2016, pp. 1522-29

Section 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 at how best to leverage heterogeneous integration.

FLIP CHIP IC PACKAGE UNIT GROWTH

Bn Units	2016	2017	2018	2019	2020	2025F	CAAGR 2020-2025
Flip Chip QFN/MIS	3.8	4.2	3.9	4.4	5.0	7.0	10%
Flip Chip SiP on Laminate*	3.4	3.8	3.7	4.4	5.8	7.1	10%
Flip Chip CSP	4.4	5.0	4.5	5.2	6.3	8.2	10%
Flip Chip CSP for DRAM	3.8	4.5	4.4	5.7	7.0	10.5	13%
Flip Chip BGA/PGA/LGA	1.0	1.1	1.1	1.3	1.5	1.7	6%

Figure 1: Flip Chip IC Package Unit Growth (Prismark 2021)

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. Another solderless technology – hybrid bonding or Cu-to-Cu interconnect – was originally used for memory-to-memory wafers but is being considered for die-to-wafer as well. This interconnect requires a chemical adhesion to bring the die/wafer together so that the Cu-to-Cu pads come into contact. This is followed by an anneal process to complete the metal-

to-metal connection. Potential design benefits include interconnect pitches that are well below 30um. One of the key challenges is planarity of the wafers or die being bonded, as an ultra-flat surface is required to ensure metal to metal contact during the metallic bond formation. Tighter pitches require more precise optical alignment for bonding and can impact the overall bond cycle time. For wafer-to-wafer bonding, wafer alignment is used for a large number of die in a batch format, vs a die-to-wafer process. For die-to-wafer hybrid bonding (figure 3a), a flip chip-like bond process is used. Most hybrid bonders are modified Flip Chip bonders with additional capabilities, and there are multiple suppliers on the market today.

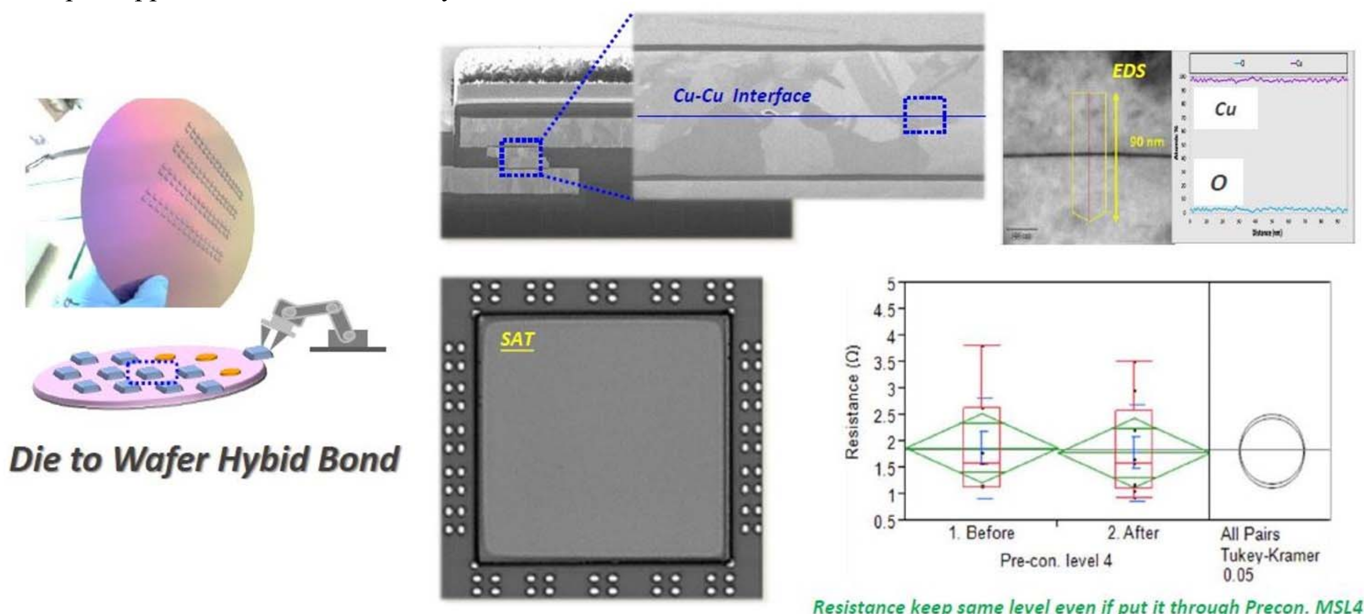


Figure 3a-Die-to-Wafer hybrid bonding (Provided by ASE – Advanced Semiconductor Engineering)

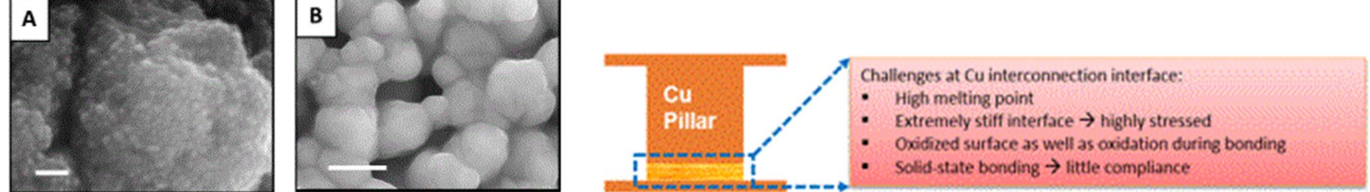


Figure 3b: SEM micrographs of copper nanoparticles isolated from the synthesis (A) and material after fusing at 210°C for4 minutes (B). The scale bars represent 100 nm and 500 nm for panels A and B respectively (Lockheed Martin)

Figure 4: Direct Cu to Cu Challenges (Georgia Tech)

3. Difficult Challenges

Challenges for flip chip depend on the application that 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.

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 pillars 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 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.

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

In this smaller-body FCCSP space, key considerations of the interconnect system can also influence high speed/frequency 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-frequency skin effect can also drive the need for good etching control of the copper lines in a substrate. As other applications, such as 77Ghz radar for automotive, come to the market, 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.

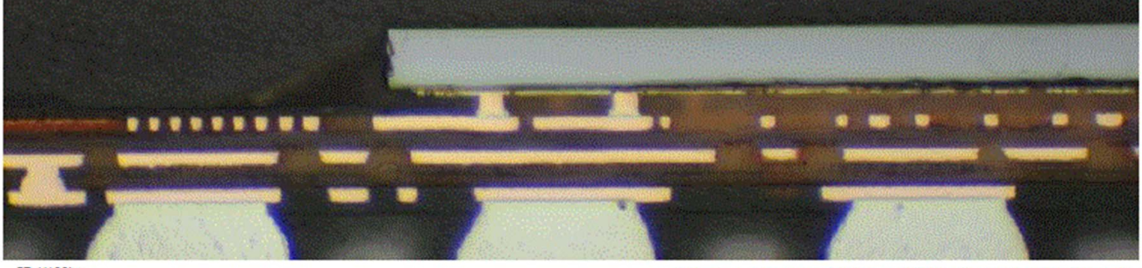


Figure 6: 3 Layer Coreless Flip Chip CSP with Copper Pillar

For larger body flip-chip devices, a number of new challenges are emerging as heterogeneous integration is 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 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 issue. 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 a new die system.

2.5D Design, Characterization & Modeling

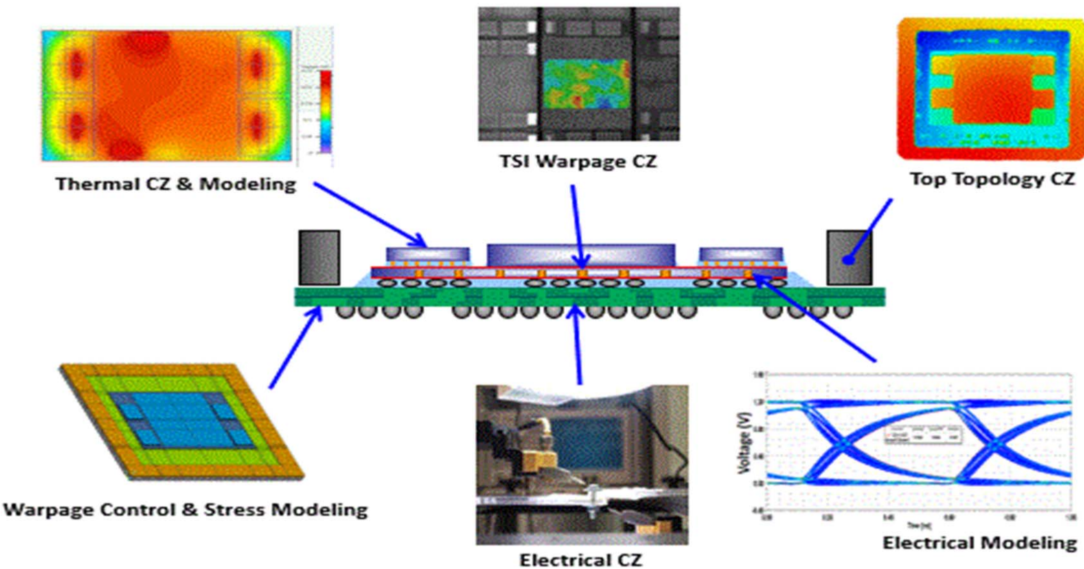


Figure 7: Modeling and Characterization key considerations for High End Flip Chip (ASE Group)

Table 1: Flip chip interconnects

Table HI-4 Chip-to-package Substrate Technology Requirements (Updated Sept 2020)									
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 (Periphery Staggered, Inline Same as large Body Cu Pillar)	40/80	30/60	30/60	30/60	30/60	20/40	20/40	20/40	20/40
Flip Chip- Cu Pillar Large Body >12mm Sq Die	110	110	100	100	90	80	80	70	70
Flip Chip Solder - COW	50	50	50	50	50	50	50	50	50
Flip Chip Cu Pillar -COW (Chiplets)	40	40	40	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
Die to Wafer Cu to Cu Interconnect (Hybrid)	30	30	30	30	30	30	30	30	30
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 HI-4:
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.
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.

Substrate technology is a key enabler for the more advanced flip-chip package solutions. Integrating jumper chips with finer lines/spaces 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, metallizations 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 challenging task. Figure 7 shows some of these considerations for a complex 2.5D Flip Chip package.

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.

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. Package. Mater. (APM 2013), vol. 2, no. Paper 8, pp. 93–101, 2013.

Section 8: Substrates

Scope

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

1. High performance computing

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

Table1. Substrate interconnect scale roadmap

Materials	Application	Features	2018	2019	2020	2021	2022	2025	2028	2031	2034
Organic laminate	FC-BGA	Min. Bump Pitch (um), Full grid array w , w/o a single route btw bumps	130,110	130,110	110,100	110,100	110,100	100,90	100, 90	90, 80	90,80
		Min. Bump Pitch (um), Periphery/ Staggered w, w/o a single route btw bumps	40/80, 30/60	30/60, 20/40	30/60, 20/40	30/60, 20/40	30/60, 20/40	20/40, 15/30	20/40, 15/30	20/40, 15/30	20/40, 15/30
		Min. Line width/space (um)	9/12	9/12	9/12	8/8	8/8	5/5	5/5	5/5	5/5
		Min. uVia diameter (um)	50	50	50	40	40	30	30	20	20
	Chiplet (Fan-out, Organic interposer)	Min. Bump Pitch (um)	50	50	50	45	45	40	40	30	30
		Min. Line width/space (um)	2/2	2/2	2/2	1.5/1.5	1.5/1.5	1/1	1/1	0.5/0.5	0.5/0.5
		Min. uVia diameter (um)	30	30	30	20	20	10	10	5	5
Silicon	Chiplet (Si Interposer, 3D)	Min. Bump Pitch (um)	40	40	40	35	35	30	30	20	20
		Min. Line width/space (um)	0.6/0.6	0.6/0.6	0.6/0.6	0.6/0.6	0.6/0.6	0.5/0.5	0.4/0.4	0.3/0.3	0.2/0.2
		Min. uVia diameter (um)	0.6	0.6	0.6	0.6	0.6	0.5	0.4	0.3	0.2

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

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

uVia scaling for organic laminates is challenging below around 40um using a CO_2 laser and 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), photo-imageable dielectric materials, dry desmear, PVD seeding processes, or damascene-like via-reveal processing.

Products with bump pitch below 150um have transitioned from solder-paste printing to ball placement or electrolytic plating technologies. It is expected that ball placement technology can be extended to about 70um bump pitch, below which electroplating technology is expected to be a potential solution. The coreless substrate with embedded traces can enable much tighter bump pitch with periphery-staggered bump layout. This would be an alternate way to tighten the bump pitch and reduce layer count by scaling the critical dimension only on the surface layer, since the bump pad is formed as part of semi-additive process without overlay penalty due to solder resist opening to capture pad alignment, while cored substrates normally have recessed solder resist defined pad.

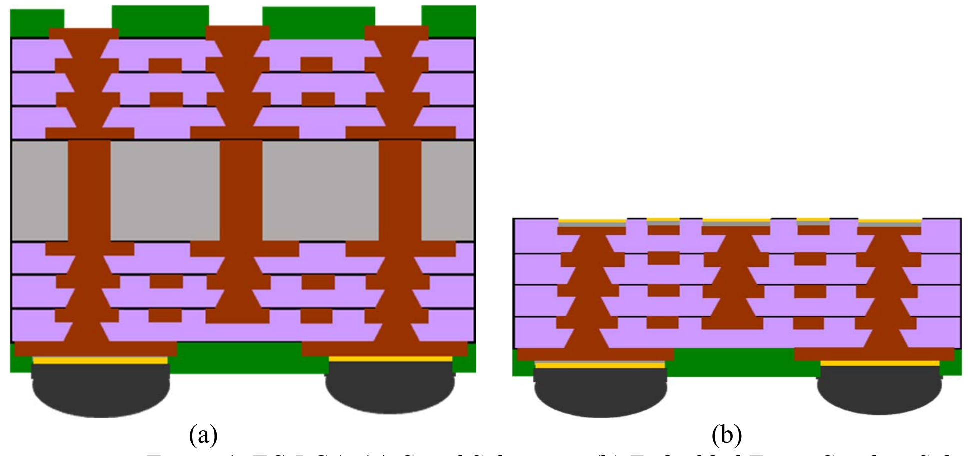


Figure 1. FC-BGA (a) Cored Substrate (b) Embedded Trace Coreless Substrate

One of the challenges in fine-pitch bumping is to enable mixed bump pitch with different pad sizes and uniform bump height. However, due to the very fine pitch, the solder bump stand-off height must be low and its variation has to be very tight. As the bump pitch scales aggressively, conventional solder-based flip chip packages are on the verge of migrating to advanced diffusion bonding using Cu-Cu and Au-Au. [2]

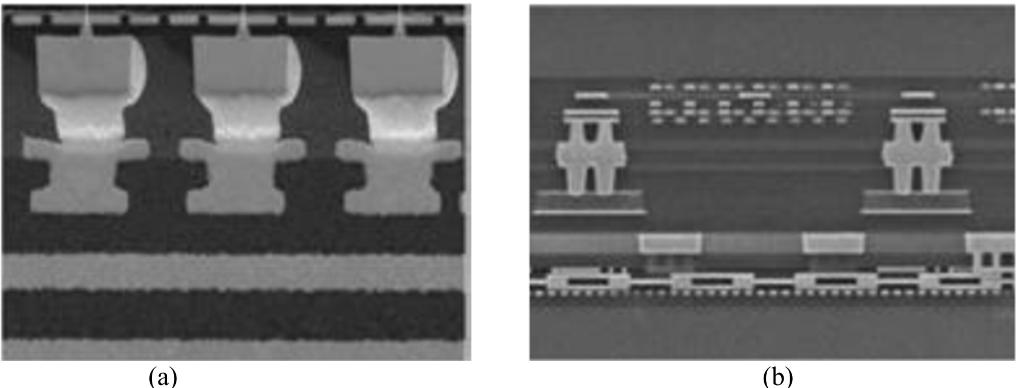


Figure 2. First Level Interconnect roadmap (a) Flip Chip Solder Based Interconnect (1995+)
(b) Cu-Cu Diffusion Based Interconnect (2021+)

With the full-grid-array fine-pitch products moving from mass reflow to thermal compression bonding (TCB) for the die assembly process, a critical parameter for assembly will be the substrate thickness uniformity within the die. The challenge is more crucial for larger form factors and higher stack-up organic-based substrates given that the thickness variation within the die increases with layer count and form factor. 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, since the probes need to scale with bump diameter and pitch and have better alignment to cover larger areas of the die field. A scalable and cost-effective test solution is required in terms of design, probe technology, and probe materials development.

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

High performance computing that includes chiplets must do more, be faster, and burn less energy, given that it has extra layers of interconnect interface for chip-to-chip communication compared to the monolithic SOC system. High-density interconnect and high-speed signal transfer rate are key enablers to match the performance of the monolithic SOC by increasing the bandwidth [3]. In general, the network switch application is leading the way in the data transfer rate within the package. As shown in the connectivity roadmaps in Table2 [4], it is expected that the link speed for serial speeds goes up to 200 Gb/s by 2027, which is a significant challenge for package and system design. Electrical losses are becoming increasingly important as data transfer rate increases. Insertion loss consists of four components of loss: dielectric, conductor, leakage, and radiation. Radiation loss and leakage loss are not an issue

when properly designed. Dielectric loss is mostly related to the loss tangent of the laminate build-up material, while conductor loss is primarily from conductor surface roughness.

Low dielectric-loss material typically has a low polar molecular resin system which is resistant to de-smear 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 de-smear and electroless Cu processing and has good adhesion to the conductor.

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

	2019	2023	2027
Link speed (Gbps)_Serial Speeds [4]	50	100	200
BU dielectrics loss, Df	0.007	0.004	0.002
BU dielectric roughness (Rq)	300~400 nm	150~200 nm	100~150 nm
Cu roughness (Rq)	350~400 nm	200~250 nm	50~100 nm

Traditional power-delivery systems use off-chip voltage regulators to supply the required voltage and currents, but power delivery networks (PDN) that connect them to the chip have a longer length and this results in significant parasitic effects. This becomes a significant issue, especially for high-performance computing systems with higher total design power (TDP) [5]. Recently, fully integrated voltage regulators (IVRs) are gaining traction since this enables dedicated voltage domains. Embedding passive components inside the package is an efficient and responsive way to reduce the ohmic resistance to improve the power delivery due to the shorter interconnect length between the passive components and the die compared to the discrete component attached on the land side or die side. Along with the development of high-density capacitance and low ESL and ESR passive components, it is essential to develop an innovative way to embed the larger number of passive components (capacitors and inductors) inside the package.

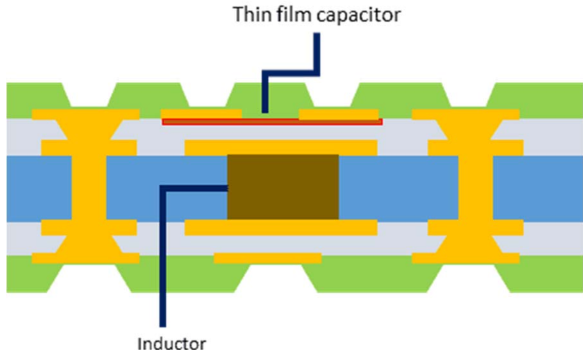


Figure 3. A schematic to demonstrate thin film capacitor and inductor embedding inside substrate

Typically, an inductor is too bulky to embed inside the buildup. The core may provide enough space to embed the inductor, but this requires development of innovative inductor materials that are compatible with the substrate manufacturing process. Embedding a few capacitors inside the core has been commonly used in industry, especially for mobile applications where it requires a thin profile for the package. However, it is quite challenging to embed hundreds of passive components in the core for high-performance computing. High density (> 100nF/mm^2) thin film capacitor (< 20um) embedding inside the buildup can be an alternate way to avoid the core space issue, as illustrated in Figure 3.

2. Mobile and Wearables

The smartphone has become the major consumer electronics product across the globe. The main trends from smartphone consumers are the continuous demand for more usability, longer battery life and affordability. At the same time, the smartphone industry rolls out new models every year incorporating new functions, better performance within the same form factor and faster processing speed. The strategy has been to develop different sets of different functional System-in-Packages (SiPs) or modules, each miniaturized and qualified. Going from one smartphone product generation to the next, the product design and release cycle would be contained within each of these functional SiPs. Smartphones today may have twenty or more SiPs. Perhaps the most important SiP is the Package-on-Package (PoP) technology. In this PoP approach, two packages are stacked on top of each other. Each package

is fully assembled on its own substrate. The PoP design, with electrical and mechanical interconnect platform built in, enables integration of ASIC/logic with memory into a single PoP package. PoP approaches are most common for the baseband or application processor assembled in FC-CSP format, with memory mounted on top of the package. A major advantage is that the devices can be fully tested before assembly. This is critical since the processors are often at the newest nodes where the wafer yield needs to be closely watched as the foundry product ramps up.

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

Shown in Figure 4 is the Qualcomm 855 package in the Samsung Galaxy 10 Smart Phone based upon a 12.5 x 12.4mm MCeP style PoP. It has a 100 um thick die with 25 um Cu pillars at < 100 um pitch. The top substrate is for mounting the memory stack. The processor die, 100 um thick, with 25 um Cu pillar at < 100 um pitch, is assembled on the lower FC-CSP substrate. The embedded trace substrate (ETS) is 10 um L/S, 130 thick with 55 um diameter via [6].

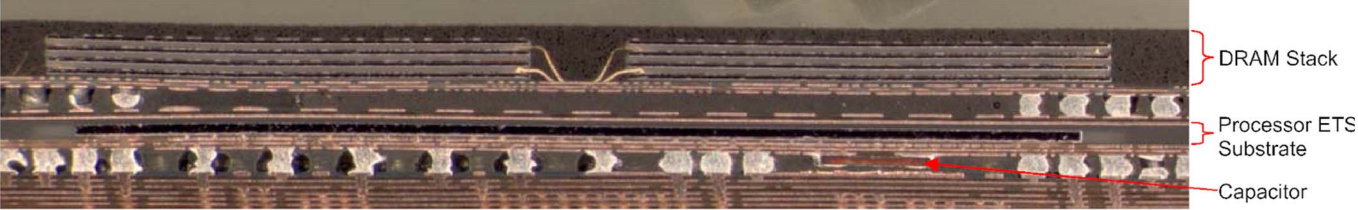


Figure 4. PoP cross-section for Qualcomm 855 processor in Samsung Galaxy 10 Smart Phone (source Prismark Partner, 2019)

Fan-out technology is another effective z-height reduction solution for semiconductor devices. This can 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 in wafer level and panel level fan-out roadmap, please refer to the WLP – Fan-in and Fan Out Chapter (Chapter 23).

In the Apple iPhone, the A12 processor is packaged in PoP architecture in TSMC InFO Fanout technology [7], as shown in Figures 5 and 6. This process has been implemented in Apple iPhone for its application processors since the adoption for the A10 processor in 2016, the A11 in 2017, and the A12 in 2018, and also used for the application processor in the Apple S4 watch in 2018. Samsung has implemented its Panel Level Fanout technology [8] in its Galaxy phone and smart watch since 2018. See Chapter 23 for in-depth discussion.

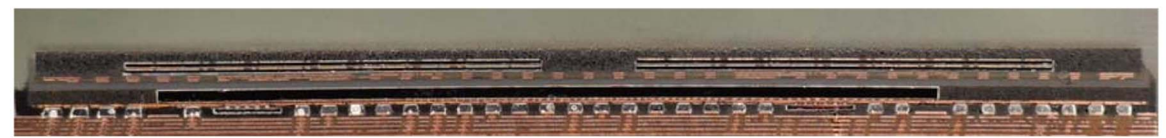


Figure 5. Cross-section of Apple iPhone A12 processor (source Prismark Partner, 2019)

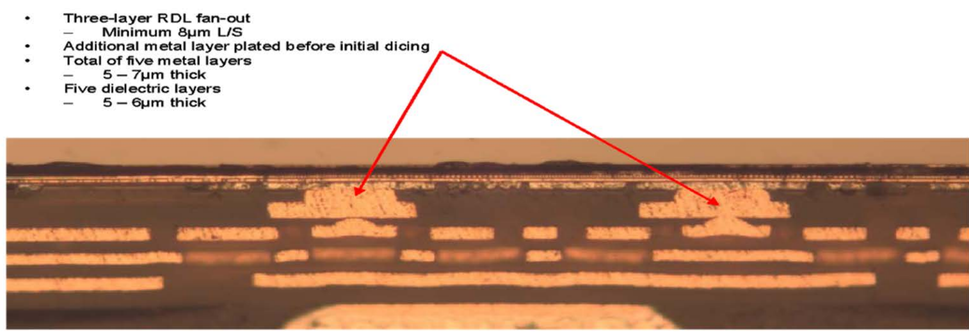


Figure 6. Expanded Fanout RDL cross-section (source Prismark Partner, 2019)

3. Network Connectivity

With the exponential growth of data traffic in the global digital economy and the arrival of 5G [9], 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, increasing bitrate requirements towards 10 Gbps, which can be only enabled through 5G mmWave. This is described in depth in Chapter 12 on 5G.

Table 3. 5G/mmWave bandwidth and substrate materials requirements

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

The network systems application requires large FC-BGA packages for the coming 5G mobile broadband, with its wider spectrum, for higher bandwidth data communication. It will require higher I/O counts and lower signal transfer loss. High I/O counts require substrates with fine line and space and fine vertical interconnect pitches (uVia and plated through hole). Low-loss signal transfer rate requires low-loss dielectric materials and smoother conductor surfaces. These substrate requirements are similar to the network switch roadmap covered in the previous section. Each of the network systems would have different operating environment and operating life requirements, different from the High Performance and Data Center needs.

4. Autonomous vehicle

With the increasing capabilities in Advanced Driver Assist Systems (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) [10] has published standards for verification of electronic components reliability requirements: AEC-Q100 Rev H (ICs), AEC-Q101 (Discrete), AEC-Q102 (Discrete Opto), and AEC-Q104 (Multichip Module). AEC-Q100 Rev H “Failure Mechanism Based Stress Test Qualification for Integrated Circuits”, published in 2017, provided requirements for Grade 0, Grade 1, Grade 2, Grade 3 based upon their function. AEC-Q104 “Failure Mechanism Based Stress Test for Multichip Modules” was published in 2017 to address increasingly complex heterogeneous integration packages such as SiPs and MCMs that includes components that passed AEC-Q100, 101 and 200 before being integrated into a package. It includes board-level reliability (BLR) and ESD requirements.

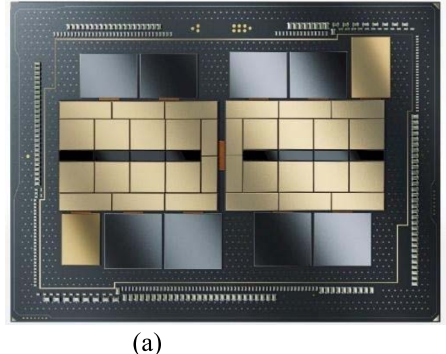
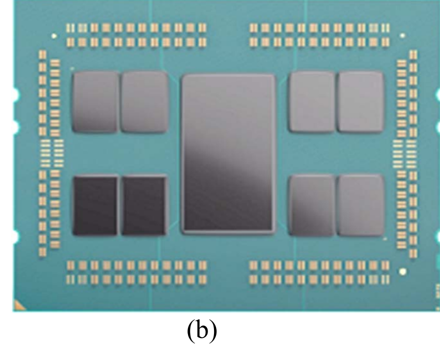
ADAS and the Autonomous Vehicle require a high level of computing power, in-vehicle networking and sensor integration, requiring multi-die large form factor package integration with advanced packaging. Developing advanced packages to meet the stringent requirements of the automotive application is challenging. Substrate process design and materials are at the front line for innovation, including core materials, buildup materials, and solder masks. Materials properties such as elastic modulus, glass transition temperature, thermal expansion coefficients (CTE1 and CTE2) and their interlayer adhesion need to be carefully examined to meet Grade 1 and 0 requirements [8].

Development of new core, build up, and passivation layer materials is needed to meet the stringent reliability and durability requirement of automotive components in heterogeneous integration. With fast transients in voltage and current causing noise which impacts nearby electronic devices, cost-effective and reliable EMI shielding technology is needed for the automotive spaces [11].

5. Difficult Challenges

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

Advanced package substrates are used for a wide range of processors, from CPU to GPU, TPU, and AI. These processors are commonly packaged with multiple chiplets, such as compute dies, ASICs, and in-package memory. Figure 7 shows state-of-the-art advanced processors that have been built using organic-based advanced substrates.

(a) (b)

Figure 7 Chiplet using advanced organic substrate (a) Ponte Vecchio with 47 active tiles (Intel) (b) EPYC Gen2 with 9 active tiles (AMD)

A substrate serves as a space transformer between die and motherboard. Combined with the emerging trend of high performance computing and high-speed in-package interconnects between die and off-package I/O, and power delivery, package form factors and layer counts of the advanced multilaminate substrates are trending to significantly increase, which drives substrate capacity constraint. Figure 8 shows the maximum form factor and layer count for high end computing.

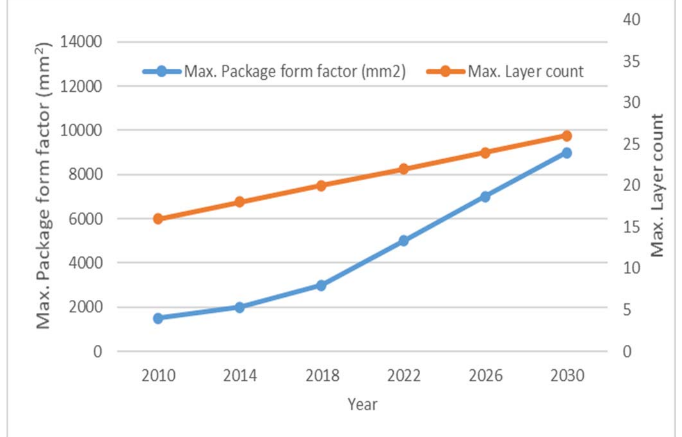


Figure 8. Package maximum form factor and layer counts

The higher layer-count and large form-factor demands have impacted manufacturing processes and technology scaling with several challenges. The biggest challenge is to handle thick and heavy panels in the current line, as the current maximum stackup is already reaching its limit (< ~2~3mm). This would cap the maximum layer count of the substrates for a while until a new factory with higher layer count handling capability is available. Across the IC industries, the theme is clear: We are time-strapped for expediting the time-to-market of the new product. However, we are seeing considerable increases in substrate lead time. This is getting worse with the increased complexity of the highly integrated substrates. A lack of substrate manufacturing capacity is the biggest challenge that IC industries are facing. The advanced ICs utilizing the larger organic panels have more room to increase the package form factor compared to a silicon-based substrate, which is typically limited by reticle size. Reticle stitching has been pursued to increase the silicon interposer, but wafter utilization will still be very low. This further exacerbates substrate capacity constraint due to not only lower panel utilization but also lower yield. With the continuous increase of layer count and form factor increase due to integration densities, and complexities driven by the integration of multiple chips, the challenges of meeting the specification have become much more acute, mainly because the panel shrinkage variation and accumulated stack tolerance is getting significantly worse while the specification requirements tighten.

With the overarching challenges associated with larger layer count and form factor, multi-pronged solutions need to be explored in parallel. First of all, substrate suppliers and equipment vendors must make sure all the future lines and tools can handle thicker and heavier panels, considering future demand. One solution to avoid higher capex is to simply boost the yields on an existing substrate line, but vendors would need to invest or upgrade their line and tooling to have a better manufacturing environment, and high-precision control in substrate manufacturing processing such as panel-level laser drilling, lithography and electrolytic plating and developing. Materials innovation is essential for critical-dimension scaling and better substrate yield, since direct build-up materials and other materials are playing essential roles not only for high resolution patterning capability for trace and via, but also good thermo-

mechanical reliability. Scaling the critical dimension for bump pitch, line width and space, and vertical interconnect would be effective ways to help flatten the curve of the capacity demand increase. However, given that scaling would increase the risk of substrate yield and quality, design optimization needs to be carefully executed by balancing the yield and benefit of scaling. As the critical dimension of the substrate is approaching the silicon one, manufacturing techniques, materials, and processes are increasingly having more in common with silicon foundry processes. The tightened tolerance to current process variation is also driving higher levels of inspection, and in more cases the development of automated repair of shorts and opens will help increase yield

References

1. "Flip Chip & WLP: Trend & Market Projection", TechSearch International, May 2019
2. "Heterogenous Integration An Advanced Packaging View" Hamid Azimi, ISES February 2020
3. "Intel Advances Architecture for Data Center, HPC-AI and Client Computing" Raja Koduri, Intel architecture day 2021
4. "IEEE 802 Plenary Meeting presentation" Ethernet Alliance, IEEE 802.3 New Ethernet Applications Adhoc meeting, March 14, 2017
5. "Power Delivery for High-Performance Microprocessors—Challenges, Solutions, and Future Trends" K. Radhakrishnan et.al. IEEE Transactions on components, packaging and manufacturing technology, VOL. 11, NO. 4, APRIL 2021
6. "Semiconductor Packaging Report" Prismark Partners, December 2018
7. "INFO (Wafer Level Integrated Fan-Out) Technology" Chien-Fu Tseng et al ECTC 2016
8. "Study of Advanced Fan-Out Package for Mobile Application" Taejoo Hwang et al ECTC 2018
9. "The Mobile Economy 2019", GSM Association, February 2019
10. Automotive Electronics Council (AEC) standards publications: AEC – Q100 Rev H (ICs), AEC – Q101 (Discrete), AEC – Q102 (Discrete Opto), & AEC Q104 (Multichip Module), 2014
11. "Automotive EMI shielding controlling automotive electronic emissions and susceptibility with proper EMI suppression methods", John, N., Gary, F., Colin, T., 2010

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.

1. Board Level Interconnect Density

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

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

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

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

The attributes of leading-edge packages drive the assembly innovations required for next-generation products. 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

2. Difficult Challenges

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

2.1. Temperature Induced Distortion

Complex Module Warpage

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

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

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

Via-in-Pad, Plated Over BGA Pads

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

Potential Solutions for Temperature-Induced Challenges

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

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

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

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 achieve in proximity to large-body BGA SiP footprints.

High Accuracy Device Placement

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

Extreme Proximity Rework

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

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

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.

2.4. Flexible Substrate Assembly

Substrate Temperature Capability

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

Printed Ink Joint Integrity

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

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.

3. Board Assembly Supply Chain Requirements

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 (<200°C) 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.

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 provide 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.

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 for print resumption. The PP/R approach is especially interesting as other electronics manufacturing methods (i.e. wire-bonding) could possibly be used on the printed partial part in between the pause and resume steps. AE solutions like this are under development both in the literature and commercially. [1-4]

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.

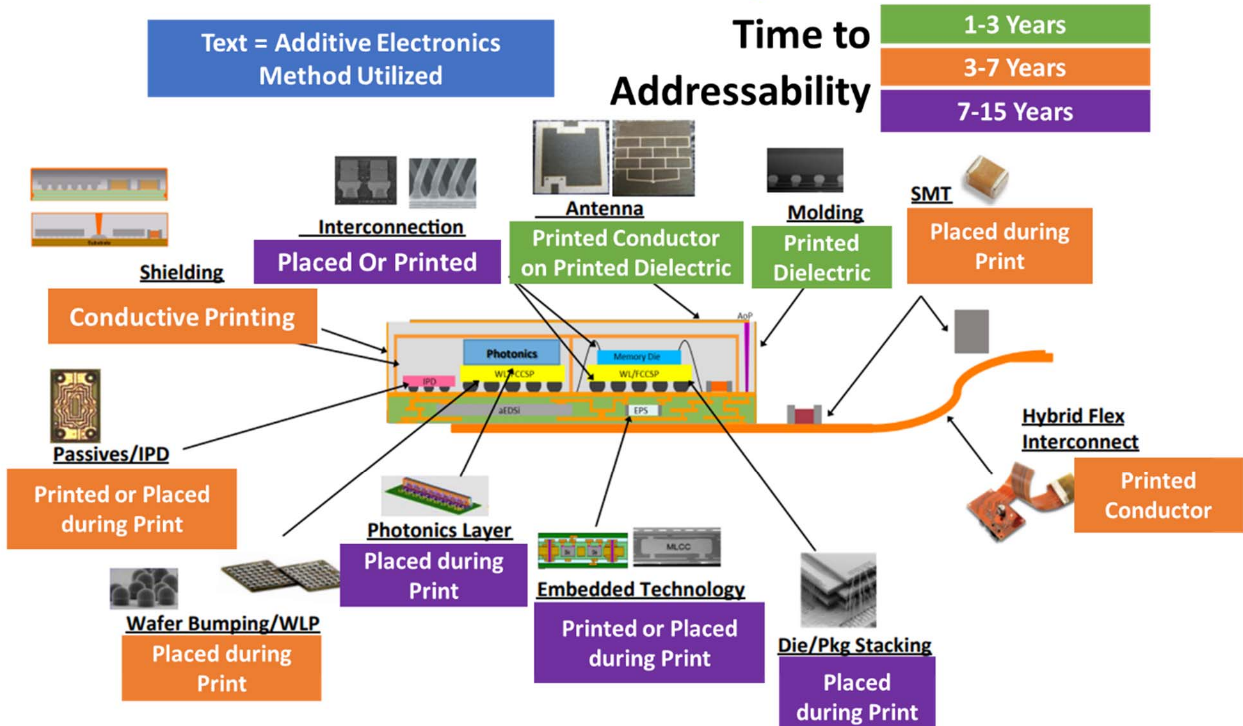


Figure 1: Possible uses for AE within SiP applications for heterogenous integration with estimated time horizons for development. See section 1, figure 7. [5]

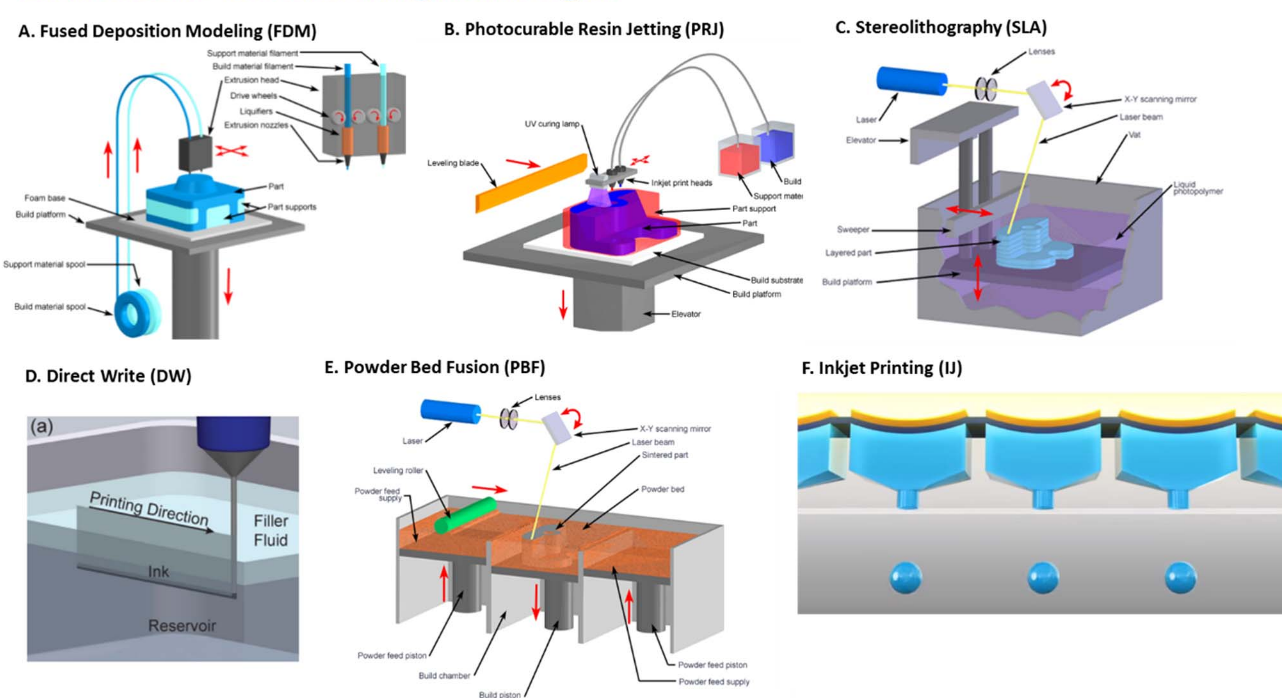


Figure 2: Basic AM technologies, which form the basis for AE [9,10]

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 2A). 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.

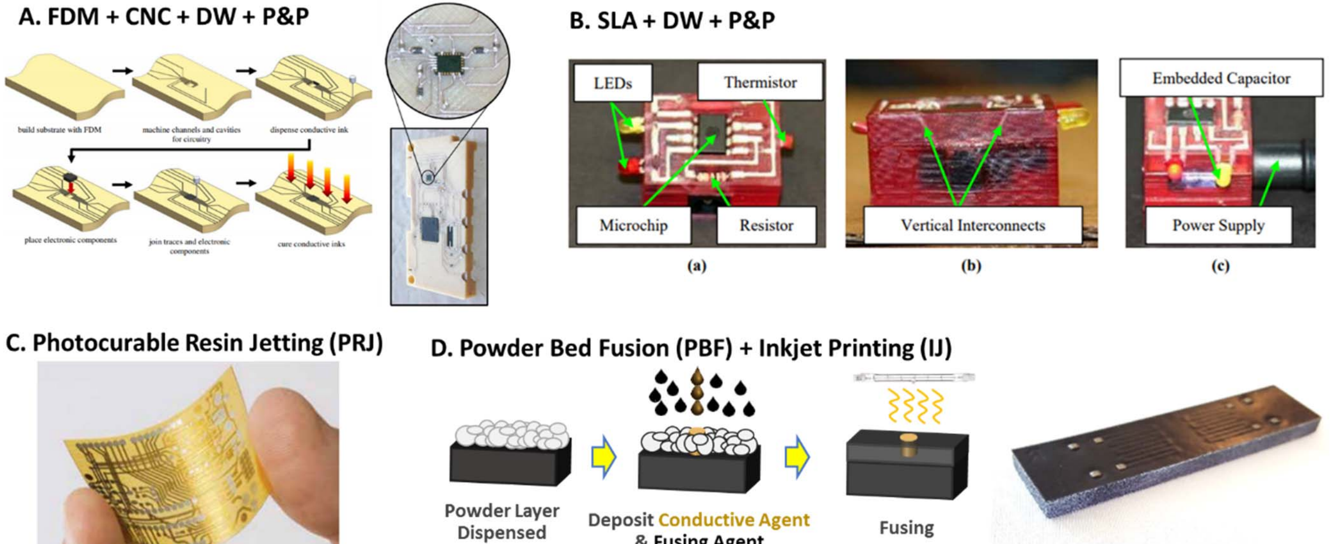


Figure 3: AE technologies, which often combine one or more AM approaches and conventional manufacturing approaches. A[3] B[2] C[12] D[18]

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 11.8z32^{-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.

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

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

For printing attributes, the line width, space width and conductivity are of primary importance. Resolutions down to 20 µm and conductivities of up to 11.8×10^{-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 for 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.

References

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 Chapter 24 on 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

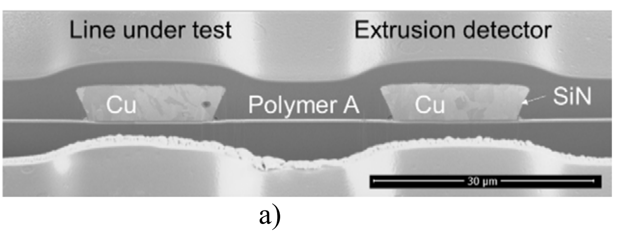
1. EM in Redistribution Layer (RDL)

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

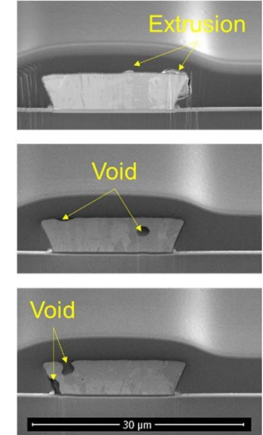
under-bump metallurgy (UBM) structures with significant resistance increases. These phenomena are responsible for structural damage of redistribution line (RDL)/UBM and UBM/solder interconnects due to ionic diffusion or electromigration.

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].



a)

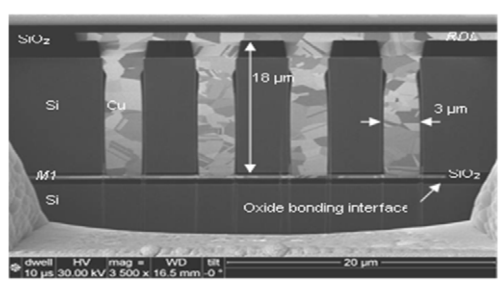


b)

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.



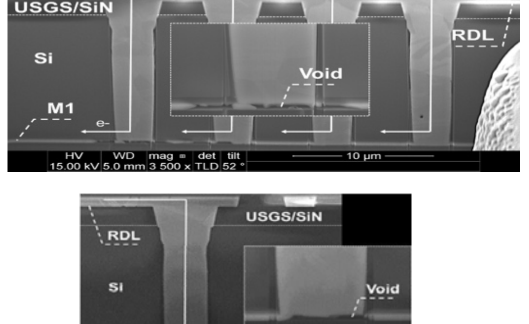


Figure 2: DD TSV and EM damage at the interface between TSV and BEOL interconnect [8].

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.

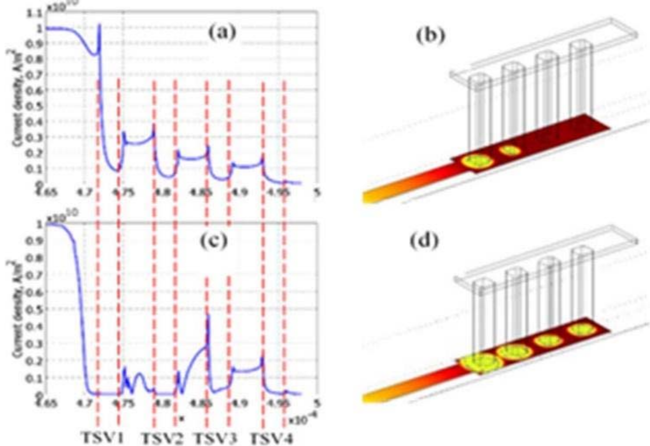


Figure 3: FEA results for up-stream configuration showing (a) initial current distribution, (b) first TSV failure, (c) current distribution after failure of two TSVs, and (d) the circuit failure. Color maps in (b) and (d) show the hydrostatic stress level: yellow is zero, darker red shows higher tensile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

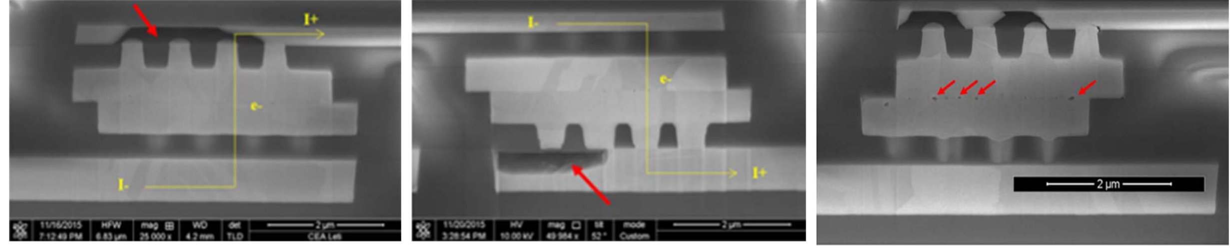


Figure 4: EM-induced voiding in the NIST-like structures [4].

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.

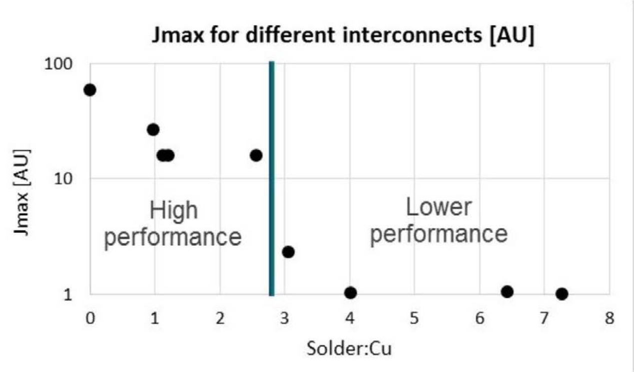


Figure 5. Classification of Jmax behavior for different solder-to-Cu volume ratios for solder interconnect structures [10].

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.

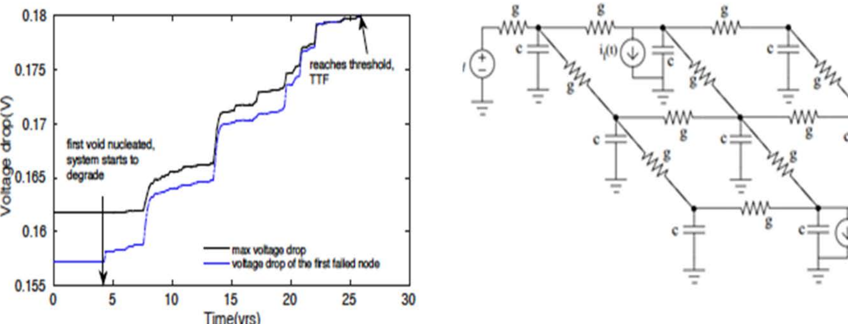


Figure 6. (a) Voltage drop of the first failed node and the maximum voltage drop in a power grid as a function of time [13], (b) Power grid schematics with user-provided current sources and voltages [14].

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 combine them 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.

Electromigration in Low temp solders

Interest continues to grow in the replacement of near-eutectic SnAgCu solder with Sn-Bi based, solder, but more needs to be known about the effect of current stressing on such low-temperature solder joints. There is an ongoing

need for assembly of electronic components that cannot safely withstand high reflow temperatures of typical near-eutectic SnAgCu (SAC) Pb-free solders [1–4]. Furthermore, lowering reflow temperatures can reduce warping in printed circuit boards (PCB) [5]. Great interest has been directed toward the Sn-Bi system, a simple binary eutectic system with no intermetallic compounds [6]. The SnBi eutectic temperature is 139°C, and this alloy has a relatively low coefficient of thermal expansion and relatively low cost [7]. These advantages make Sn-Bi solder a promising choice for low temperature soldering application in the electronics industry [1]. However, the mechanical properties of Sn-Bi alloys depend strongly on Bi concentration. Bi diffuses by a vacancy mechanism at a relatively high rate in Sn, particularly when subjected to current stresses in the kA/cm^2 range, at typical operating temperatures [8–11][9,19,21,22]. A number of different studies have observed the accumulation of a significant fraction of the Bi at the anode of a Sn-Bi based solder joint during current stressing at a current density up to approximately 5 kAcm^2 and temperature up to 125°C. For instance, Figure 1 reveals the accumulation of Bi at the anode of a near eutectic, Sn-Bi based solder joint after current stressing only 200 h at 125°C. Similar results have been observed for longer times, at lower temperatures and current densities. It's clear that Bi accumulates in significant amounts at the anode of these low-temperature solder joints, potentially changing their mechanical properties and reliability. Therefore, we require a clear understanding of the effect of current stressing on the microstructure of these low-temperature alloys in order to optimize the reliability of such solder joints [12,13]. In fact, some previous investigations have focused on electromigration failure in solder, in particular SnPb solder joints [32]. While the physical mechanism of electromigration has been most extensively investigated for pure metals [30,31], electromigration was deemed a reliability issue in packaging technology by the International Technology Roadmap of Semiconductors (ITRS) in 1999 [33]. Since then, electromigration has been identified as a major potential failure mechanism in interconnects and solder joints in semiconductor devices in the presence of high direct-current densities.

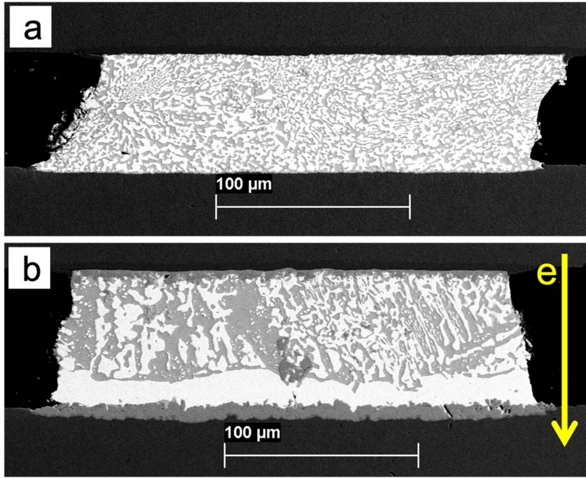


Figure 1: A scanning electron microscopy micrograph of a cross section of a near eutectic Sn-Bi based solder joint (a) as received, and (b) which was exposed to a current density of 4 kA/cm^2 for 200 h, at a temperature of 125°C. A significant fraction of the Bi in the solder has accumulated in a twenty micron thick layer at the anode.

A study of current stressing in low-temperature solder joints must encompass not only eutectic SnBi, but other Sn-Bi based alloys. Often, small percentages of alloying elements such as Ag, Cu, Ni, Sb, and In [1] are added to SnBi to improve its mechanical properties. For instance, adding a small amount of Ag (up to 2%) to Sn-58Bi was found to enhance its creep resistance and also to increase solder strength [23–25]. While many researchers have tried to improve the mechanical properties of Sn-Bi solder alloy by addition of one or two minor elements, a clear concentrated effort is needed in order to understand the effect of such tertiary and quaternary elements on the electromigration of Bi in Sn-Bi based alloys.

Besides near eutectic Sn-Bi based solder joints, examination of electromigration in SnBi/SnAgCu mixed assemblies is needed. Since the majority of solder-bumped components available in the electronics supply chain have SAC solder balls attached, near-term studies of current stressing should include examination of mixed assembly SAC/Sn-Bi-Ag solder joints. These solder joints (e.g., SAC305/Sn-57Bi-1Ag) are fabricated in order to provide lower temperature assembly for a wide range of conditions. The mixed solder joint after reflow would be either partial mixed or total mixed (homogeneous), depending upon the initial paste-ball volume ratio, reflow peak temperature and time above liquidus [31].

The motion of Bi due to electromigration (electric wind force) and due to any Bi concentration gradients must be carefully considered in a study of current stressing. If these two forces dominate, then one can model the Bi flux simply as

$$J = \frac{CD}{kT} Z^* e\rho j - D \frac{\mathrm{d}C}{\mathrm{dx}}$$

In fact, such a model that contains only the two terms of this equation has been found to generally be effective for temperatures up to 125°C and current densities up to approximately 5 kA/cm^2. But clearly care is needed; when a direct current is applied to a solder joint, a thermal gradient is established by the associated joule heating. There is some indication that for current densities approaching 10 kA/cm^2 and above, these gradients may transport atoms from hot spots toward the surrounding area. Continued consideration of such possibilities is important as geometries change, particularly with decreases in length scale, as well as of the development of significant stress gradients. [42–44]

Only the first term of the equation is needed to model the prodigious migration of Bi reflected in the example of the data of Figure 1. In this case, Bi precipitates are present at all times and they maintain the concentration of Bi, C, in equilibrium with the SnBi phase diagram (i.e. 20%wt percent at a temperature of 125°C). Thus the second term is zero, and the first term is constant. Linear variations of the accumulated Bi thickness versus time are found in such cases, with linear fits to the data resulting in values of the parameters (i.e. DZ*) in reasonable agreement with previous measurements of D, and estimates of Z* = -50 based upon measurements of Z* for Pb in SnPb. Perhaps the most important point here is that, given the very large electrical resistance of Bi as compared to Sn and near-eutectric SnBi, a simple model for the variation of the electrical resistance changes of these low-temperature solder joints with Bi accumulation results, and is proven to work well. Four terminal measurements (e.g. Figure 2) of the electrical resistance of low temperature solder joints such as those of Figure 1 during current stressing show a linear increase in electrical resistance that agrees well with measurements of Bi thickness and simple theory. Measurements of the electrical resistance provide measurements of the Bi thickness at the anode during current stressing.

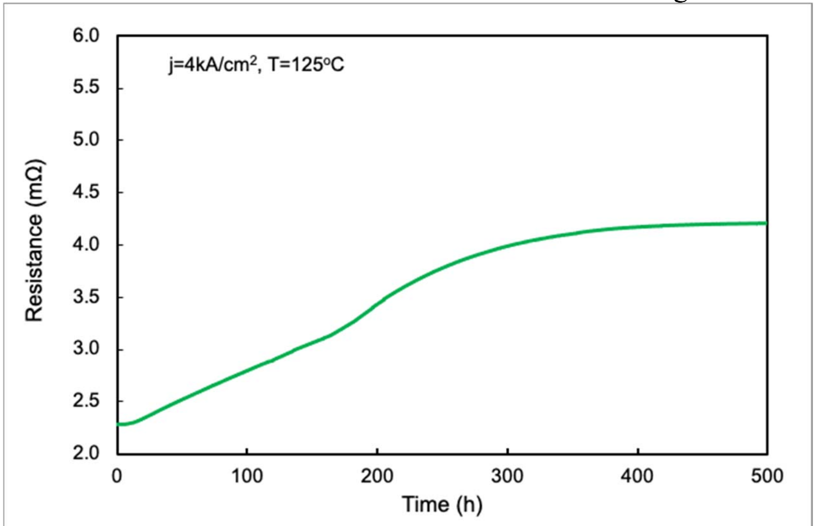


Figure 2: A plot of average resistance change of SnBi solder joint under current stressing of 4 kA/cm^2 at 125°C as a function of time

Models of measurements of the thickness of Bi accumulating at the anode during current stressing of samples like those of Figure 1, and corresponding measurements of changes in electrical resistance, lead to specific predictions of electrical resistance changes during current stressing at different values of current density and temperature. Thus, one may arrive at a prediction for the mean time to failure.

The relative influence of both terms in the equation is clearly illustrated with a current-stressing experiment performed on a SnBi/SAC mixed assembly. The reflow of a component with SAC solder balls attached on pads (with Cu or Ni metallization) and SnBi-based solder paste resulted in SAC/SnBi mixed solder joints with a clear demarcation between SnBi-rich region and the SAC phase (Figure 3a). The microstructure of these SAC/SnBi mixed solder joints was characterized before and after prolonged application of a current stress to investigate Bi movement. Figures 3b and 3c demonstrate the microstructures of representative SAC/SBA mixed solder joints after current stressing of 4 kA/cm^2 at 125°C for 200 h. In this experiment, alternate joints were stressed in the opposite direction, with the same current magnitude. The arrows labeled 'e' indicate the direction of electron flow. Due to the existence

of large amounts of Bi near the board interface and almost no Bi in the SAC region, the initial concentration gradient generated will always drive Bi in the upwards direction. Therefore, in solder joints with upward electron flow, both driving forces including electromigration and concentration gradient are in the same direction and push Bi towards the component interface. In contrast, in solder joints with downward electron flow, two driving forces act in opposite directions. For relatively short current stressing time (200 h), it can be seen that for solder joints with an upward current (Fig. 3b), Bi diffused towards the component interface. In the solder joints with a downward current (Figure 3c), a significant amount of Bi was observed to have segregated at the interface with the IMC at the board side. In solder joints with an upward electron flow, Bi precipitates are closer to the component interface than that in the joints with a downward current (Figures 3b and 3c). This indicates that concentration gradient has a dominant effect on Bi distribution.

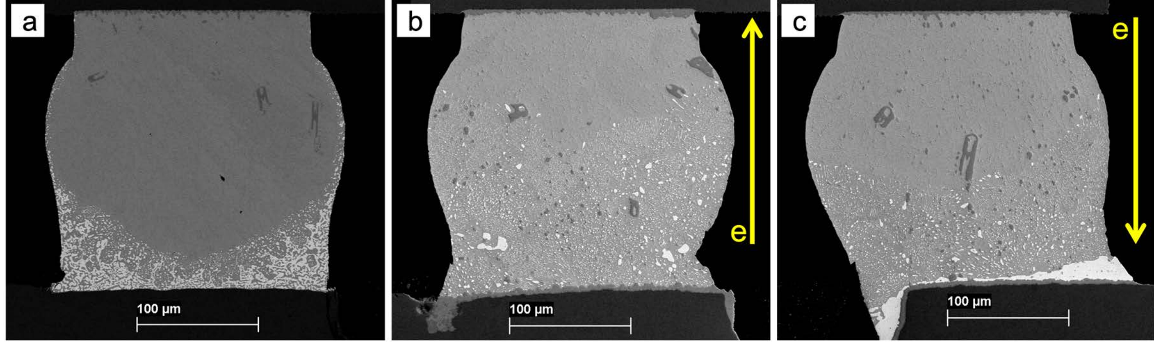


Figure 3: Microstructure of SAC/SnBi mixed solder joint (a) after reflow, after 200 h current stressing of 4 kA/cm^2 at 125°C under (b) upward electron flow direction (c) downward electron flow direction

References

1. F. Wang, H. Chen, Y. Huang, L. Liu, and Z. Zhang, J. Mater. Sci. Mater. Electron. 30, 3222 (2019).
2. H. Fu, R. Aspandiar, J. Chen, S. Cheng, Q. Chen, R. Coyle, S. Feng, M. Krmpotich, R. C. Lasky, S. Mokler, J. Radhakrishnan, I. Corporation, M. Hill, M. Corporation, I. Corporation, and I. Corporation, 17 (2017).
3. S. Sahasrabudhe, S. Mokler, M. Renavikar, S. Sane, K. Byrd, E. Brigham, O. Jin, P. Goonetilleke, N. Badwe, and S. Parupalli, in 2018 IEEE 68th Electron. Components Technol. Conf. (IEEE, 2018), pp. 1455–1464.
4. H. Fu, J. Radhakrishnan, M. Ribas, R. Aspandiar, B. Arfaei, K. Byrd, A. Caputo, J. Chen, Q. Chen, R. Coyle, D. Daily, S. Feng, I. Corporation, M. A. Electronics, I. Corporation, S. C. Corp, and I. Electronic, 513 (2019).
5. S. Mokler, D. Ph, R. Aspandiar, D. Ph, K. Byrd, O. Chen, and S. Walwadkar, 25 (2016).
6. H. Okamoto, J. Phase Equilibria Diffus. 31, 205 (2010).
7. F. Hua, Z. Mei, and J. Glazer, 277 (1998).
8. F. Wang, L. Liu, D. Li, and M. Wu, J. Mater. Sci. Mater. Electron. 29, 21157 (2018).
9. X. Zhao, M. Saka, M. Muraoka, M. Yamashita, and H. Hokazono, J. Electron. Mater. 43, 4179 (2014).
10. C. M. Chen, L. T. Chen, and Y. S. Lin, J. Electron. Mater. 36, 168 (2007).
11. X. Gu, D. Yang, Y. C. Chan, and B. Y. Wu, J. Mater. Res. 23, 2591 (2008).
12. R. Coyle, R. Aspandiar, M. Osterman, C. Johnson, R. Popowich, R. Parker, D. Hillman, N. Bell-labs, M. Hill, C. Park, and R. Collins, Proc. SMTA Int. Rosemont, IL, USA 17 (2017).
13. P. J. Shang, Z. Q. Liu, D. X. Li, and J. K. Shang, Scr. Mater. 58, 409 (2008).
14. Y. S. Lai, K. M. Chen, C. L. Kao, C. W. Lee, and Y. T. Chiu, Microelectron. Reliab. 47, 1273 (2007).
15. K. N. Tu, Springer Ser. Mater. Sci. 117, 245 (2007).
16. Y. C. Chan and D. Yang, Prog. Mater. Sci. 55, 428 (2010).
17. P. Su, M. Ding, T. Uehling, D. Wontor, and P. S. Ho, Proc. - Electron. Components Technol. Conf. 2, 1431 (2005).
18. P. Liu, A. Overson, D. Goyal, I. Corporation, W. C. Blvd, C. Az, and M. Stop, 473 (2020).
19. L. T. Chen and C. M. Chen, J. Mater. Res. 21, 962 (2006).
20. S. B. Liang, C. B. Ke, W. J. Ma, M. B. Zhou, and X. P. Zhang, Proc. - Electron. Components Technol. Conf. 2016-Augus, 264 (2016).
21. Y. C. Tan, C. M. Tan, and T. C. Ng, Microelectron. Reliab. 50, 1352 (2010).
22. X. Gu and Y. C. Chan, J. Electron. Mater. 37, 1721 (2008).
23. S. Liu, T. Song, W. Xiong, L. Liu, Z. Liu, and S. Huang, J. Mater. Sci. Mater. Electron. 30, 6701 (2019).
24. H. Sun, Q. Li, and Y. C. Chan, J. Mater. Sci. Mater. Electron. 25, 4380 (2014).
25. R. M. Shalaby, Mater. Sci. Eng. A 560, 86 (2013).
26. M. Mccormack, H. S. Chen, G. W. Kammlott, and S. Jin, J. Electron. Mater. 26, 954 (1997).
27. J. Shen, Y. Pu, H. Yin, D. Luo, and J. Chen, J. Alloys Compd. 614, 63 (2014).
28. L. Shen, Z. Y. Tan, and Z. Chen, Mater. Sci. Eng. A 561, 232 (2013).

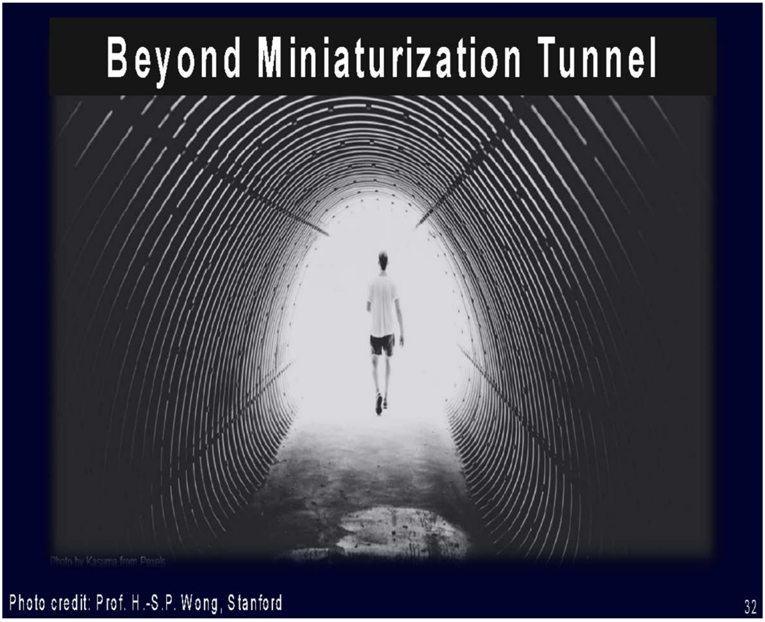
29. S. Ismathullakhan, H. Lau, and Y. C. Chan, Microsyst. Technol. 19, 1069 (2013).
30. G. Xu, F. Guo, X. Wang, Z. Xia, Y. Lei, Y. Shi, and X. Li, J. Alloys Compd. 509, 878 (2011).
31. and E. J. C. Mohammed Genanu, Faramarz Hadian, Randy Owen, J. Electron. Mater. (2020).

Section 12: Summary and Difficult Challenges

1. Introduction

It has been 57 years since Dr. Gordon Moore Published the paper "Cramming More Components onto Integrated Circuits" in the journal Electronics on April 19, 1965. In a brief summary for the Fairchild lawyer, who would review his draft, Moore wrote "The promise of integration is extrapolated into the wild blue yonder, to show that integrated electronics will pervade all electronics into the future. A curve is shown to suggest that the most economical way to make electronics systems in ten years will be of the order of 65,000 components (transistors) per integrated circuit." [1]. At the time of writing his paper in early 1965, the industry integrated circuit capability was 65 transistors. He foresaw markets and applications with integrated circuits becoming more and more powerful, with applications not yet imagined, that would expand throughout the electronics industry, for society and humanity. For half a century, year after year, the exponential rate in Moore's Law continued and the proliferation of tinier transistors has made electronics profoundly affordable and incredibly powerful.

In 2003 at the ISSCC Conference, Dr Moore gave a plenary lecture titled "No Exponential is Forever, But It Can Be Delayed". It has been 18 years since the "No Exponential is Forever" lecture. Heterogeneous Integration will be the critical "DELAY" role as well as for integration of new and emerging devices into the global semiconductor portfolio over the next fifty years. [2]. This Single and Multichip chapter provides the basic building block toolbox articulating the state of the art, projecting future challenges and potential solutions.



At the 2020 ERI Conference the Plenary Speaker, Prof Philip Wong of Stanford University, and TSMC Chief Scientist, gave a keynote talk "The Future is System Integration". He characterized semiconductor research near the end of Moore's Law as being like a person walking out of a long tunnel seeing green fields and sunlight. During Moore' s Law time, the single focus has been miniaturization towards the next set of nodes. As one emerges from the miniaturization tunnel, opportunities for research outlook and innovations become infinitely brighter and broader. This is the promise and challenge for all of us: Building Future Forward.

2. Summary

This Single and Multichip Integration chapter starts with an Executive Summary followed by 11 sections on key packaging and assembly technologies in packaging design and assembly, with building blocks from the Knowledge Base and Data for Package Integration:

A. Knowledge: Base and Data for Package Integration Co-Design

- Electrical Analysis and System Requirements
- Thermal Management
- Mechanical Analysis
- Electromigration
- Reliability

B. Packaging and Assembly Manufacturing Flow with Substrate

- Wafer Singulation and Thinning
- Wirebond
- Flip Chip
- Substrate
- Board Assembly
- Additive Manufacturing

The Knowledge Base and Data for Package Integration Co-Design and Packaging, and Assembly Manufacturing Flow with Substrate, form the foundation pillars of all market applications: High Performance Computing/Data Centers, Mobile, Automotive, IoT and 5G, Wearables and Aerospace and Defense. The Single and Multichip Integration Chapter TWG works collaboratively with other chapters TWGs as illustrated in Figure 1 below.

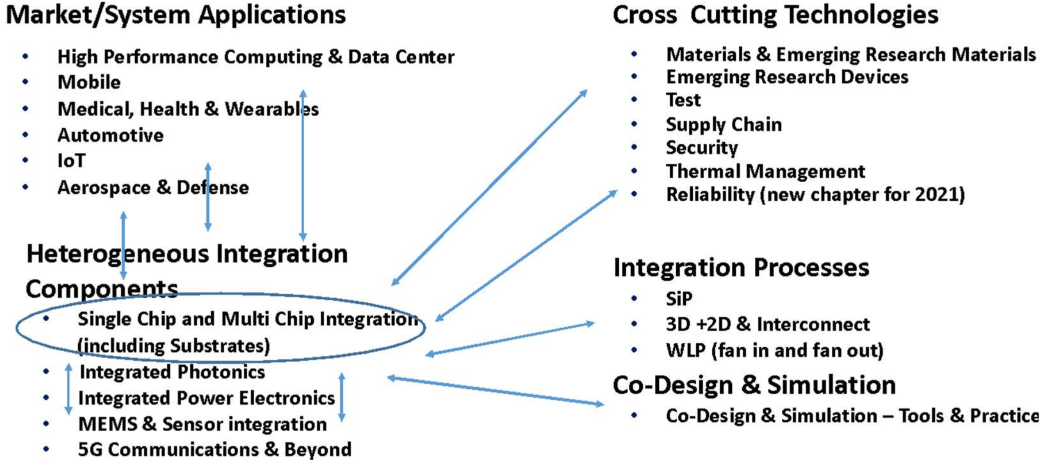


Figure 1 : Heterogeneous Integration Roadmap Chapters and Technical Working Groups (TWGs)

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."

3. Difficult Challenges:

Each section has articulated its difficult challenges and potential solutions; we shall not repeat them here.

Difficult Challenges

A. Next Five Years

1. Integrate artificial intelligence and machine learning into the profession and industry from research, development, design, prototyping to volume manufacturing.
2. Packaging and Assembly for 3D ICs and 3D Packages from, Co-Design, Assembly and Test, Materials, Process and Equipment in a cost-effective way.
3. Packaging Assembly and Test for Backside Power Delivery per IMEC and ARM at VLSI Conference 2021, described as "Power Via" by Intel, for 2024 implementation.

B. Five to fifteen years

1. Packaging and Assembly for Quantum Computing proliferation

2. Packaging and Assembly for 2D semiconductor devices such as based upon CNT (Professor Mitra, February 24th 2021 HIR Symposium.)

References

1. "Moore's Law: The Life of Gordon Moore, Silicon Valley's Quiet Revolutionary", Basic Books 2015, page 254.

5. Single and Multi-Chip Chapter Team

William Chen (ASE)

Baghat Sammakia (BU)

Ivy Qin (KnS)

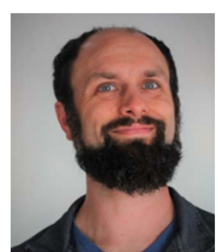
Jim Wilcox (UIC)

Eric Cotts (BU)

Annette Teng (Promex)

Srikanth Rangarajan (BU)

Mark Gerber (ASE)


Kris Erickson (FB)

Valeriy Sukharev (Mentor)

Lei Shan (Tekollect)

Benson Chan (BU)

Kyu-Oh Lee (Intel)

Paul Ho (UT)

TWG Members

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

TWG Chair: William Chen and Co-Chair Annette Teng

Section 1	Background and Scope	Bill Chen (ASE)
Section 2	Electrical Requirements	Lei Shan (Tekollect)
Section 3	Thermal Management	Bahgat Sammakia, Scott Schiffres, Srikanth Rangarajan (BU)
Section 4	Mechanical Requirements	Benson Chan (BU)
Section 5	Wafer Singulation &Thinning	Annette Teng (Promex)
Section 6	Wire Bond	Ivy Qin (KnS)
Section 7	FlipChip	Mark Gerber (ASE)
Section 8	Substrate	Kyu-Oh Lee (Intel)
Section 9	Board Assembly	Jim Wilcox (UIC)
Section 10	Additive Manufacturing	Kris Erickson (Facebook)
Section 11	Electromigration	Paul Ho (UT), Valeriy Sukharev (Mentor) Eric Cotts (BU), Faramarz Hadian (BU)
Section 12	Summary & Difficult Challenges	Bill Chen, ASE

''''''''

''''''''''Gf kxgf 'd{ 'RcwnY gurkpi