HETEROGENEOUS INTEGRATION ROADMAP

2021 Edition

Chapter 23: Wafer-Level Packaging (WLP)

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.

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

Table of Contents

Chapter 1: Heterogeneous Integration Roadmap:
Driving Force and Enabling Technology for Systems of the Future
Chapter 2: High Performance Computing and Data Centers
Chapter 3: Heterogeneous Integration for the Internet of Things (IoT)
Chapter 4: Medical, Health and Wearables
Chapter 5: Automotive
Chapter 6: Aerospace and Defense
Chapter 7: Mobile
Chapter 8: Single Chip and Multi Chip Integration
Chapter 9: Integrated Photonics
Chapter 10: Integrated Power Electronics
Chapter 11: MEMS and Sensor Integration
Chapter 12: 5G, RF and Analog Mixed Signal
Chapter 13: Co-Design 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
1. Executive Summary 1
2. Market Drivers and Applications for WLP 3
3. Wafer Level Packaging Overview: Technologies, Integration, Evolution and Key Players 7
4. Technical Challenges 21
5. Supply Chain Activities and Considerations 27
6. Summary and Final Conclusions 31
Chapter 24: Reliability

Chapter 23: Wafer-Level Packaging

Scope and Chapter Outline

The intent of this chapter is to provide a brief overview of Wafer Level Packaging (WLP), including Wafer Level Chip Scale Packaging (WLCSP) and Fan-Out Packaging, as a background for a roadmap of these technologies going forward. It is not the intent to give a detailed history, nor a detailed description of all possible structures, processes and materials that are associated with these technologies. More detailed information can be found in various articles and books published on the subject. This chapter is an attempt to look at WLP technology as it has developed to date, and project forward to future needs and challenges.

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

https://rebrand.ly/HIR-feedback

Wafer level packaging is where a die is packaged while still in wafer form, either singly or combined with additional dies or other components such as discrete passive devices, or functional components like micro-electromechanical systems (MEMS) or radio-frequency (RF) filters. This allows the production of wafer- and panel-level packaging using heterogeneous integration. Although by definition WLPs have historically been produced using either a 200mm or 300mm diameter round wafer format, multiple suppliers are extending similar manufacturing methods to rectangular panel formats. This will allow the manufacture of heterogeneous packages not only on a wafer level infrastructure (Wafer Level Packages, or WLPs), but also based on a panel level infrastructure (Panel Level Packages, or PLPs). This chapter will include both WLP and PLP formats for the Heterogeneous Integration Roadmap (HIR). The chapter is organized into 7 sections:

1. Executive summary
2. Market drivers and applications for wafer level packaging
3. Wafer level packaging overview: Technologies, integration, evolution and key players
4. Technical challenges
5. Supply chain activities and considerations
6. Summary, final conclusions and acknowledgements
7. References

1. Executive Summary

The world of technology is an ever-expanding part of our daily lives. We have seen, over the past decades, significant innovation in the semiconductor industry which greatly impacted our daily activities – from computing and the internet to the introduction of mobile devices and the evolution of smartphones. While mobile applications have driven a lot of innovation in packaging and semiconductors for the past decade, with a strong adoption of wafer level packaging technologies, new applications are emerging that bring new challenges and requirements across the entire supply chain. The semiconductor industry is experiencing unprecedented times, with growth being seen across all the different market segments, with new applications such as autonomous driving, 5G, internet of things (IoT), artificial intelligence, machine learning, etc. growing and bringing new opportunities for innovation. [1]

New applications are also bringing new market requirements. If performance has been primarily important for computing, especially for mainframe and enterprise computing; this changed as technologies moved into the consumer space, a more price-sensitive market, and cost has become a critical requirement for applications such as personal computers, tablets, smartphones and now wearables. The need for more compact form factors and thinner devices has driven the industry to develop packaging technologies that can further drive miniaturization while enabling improved performance and cost. With the wide range of applications emerging, such as, 5G, IoT, artificial intelligence, processing at the edge, autonomous driving, infotainment, etc., more functionalities will be required, from computing to networking, connectivity, storage, sensing, and power management. With Moore's Law slowing down, it is becoming more difficult and costly, with a longer time to develop and integrate multiple functions at the device level using traditional transistor scaling and monolithic integration of system-on-chip (SoC) with front-end-of-line (FEOL) processing. Not all the functionalities require the most advanced technology nodes – therefore, optimizing the manufacturing process, technology and infrastructure, as shown in Figure 1, and then heterogeneously integrating the different functionalities within a system, have shown that market requirements can be achieved and further improved; this approach brings more flexibility, faster time to market, lower cost and optimized signal integrity and power. Integration through mid- and back-end-of-line (MEOL/BEOL) processing are bringing several benefits, reasons why System-in-Package (SiP) and Heterogeneous Integration (HI) are gaining a lot of interest in the industry with many players expanding their activities in this area and new ones entering this space. [2-3]

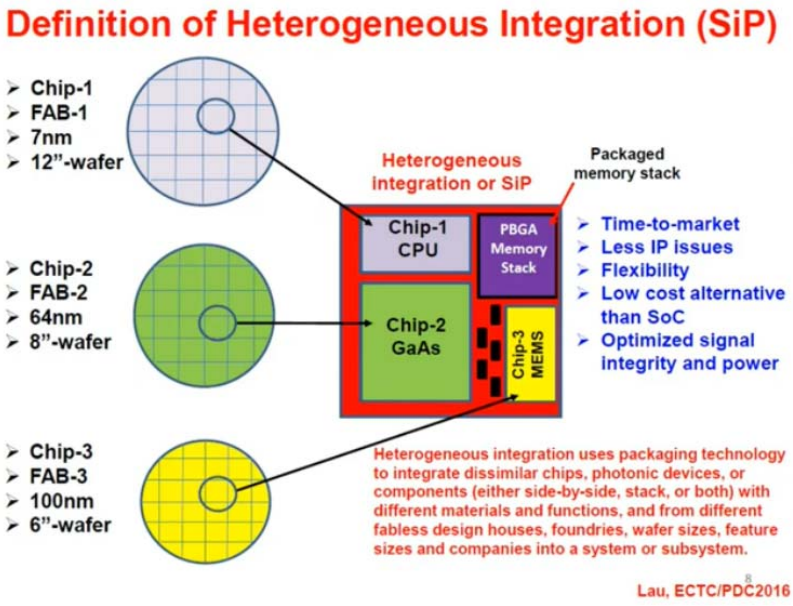

Figure 1. Definition of Heterogeneous Integration [3]

Integrating multiple functionalities within the same system is not something new. The industry saw the introduction of multi-chip modules (MCM) by IBM for data centers and enterprise applications in the 1980's. At that time, the volumes and yields were low and the supply chain customized for this application, while the modules were using flip-chip type packages and co-fired ceramic substrates [4]. Over the years, as shown in Figure 2, system-in-package (SiP) integration has significantly evolved, with a strong boost at the beginning of this century, driven by the introduction and evolution of smartphones. Cost, form factor and performance have become key criteria driving development and further adoption of SiPs and heterogeneous integration. At first it targeted radio-frequency (RF) modules, and then expanded to other devices as well, such as optical and non-optical sensing, processors and memory integration. Various packaging technologies have been developed and adopted in the smartphone, from wire-bonded and flip-chip packages to wafer level packaging, and even 3D integration using the through-silicon-via (TSV) for miniaturization of micro-electromechanical systems (MEMS) and image sensors integrated with the processors in more compact packages.

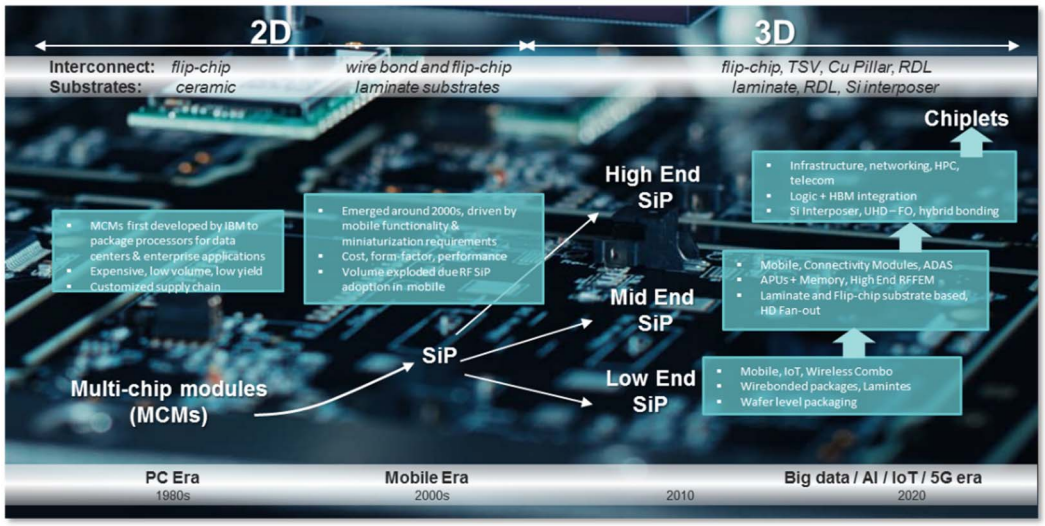

Figure 2. Heterogeneous Integration Benefits and Applications. [5]

A wide range of system-in-package solutions are currently available to address the needs of various applications:

- *Lower end SiPs* using wire-bonded type modules and wafer level packaging technologies. Such SiPs can be found in mobile, internet of things (IoT) and wireless combo applications. While mobile is the main driver behind WLP adoption, there are new applications in the automotive, industrial and consumer market segments adopting such packages.

- *Mid-end SiPs* are dominated by application processors integrated with memories and connectivity modules, using flip-chip packages and high-density fan-out technologies. With the automotive market moving to more advanced technology for its computing needs, packaging platforms are also evolving from traditional to advanced packaging and system-integrated solutions.
- *High-end SiPs* are primarily found in networking, high-performance computing, servers and cloud computing, infrastructure and telecom applications. Targeting high-end processors, they require closer integration with high bandwidth memories, processor customization and chiplets for specific loads by integrating artificial intelligence and networking functions in new structures or partitioning accelerators for increased yield. A wide range of packages and interconnect technologies have been adopted and considered, from multi-chip modules supported by flip-chip ball grid array (FCBGA) to 2.xD integration using organic interposer and finer structuring at the substrate level, to the well-known 2.5D silicon interposer, 3D integration with through-silicon-vias (TSVs) and hybrid bonding, ultra-high density fan-out (UHD FO) integration, and more recently, bridge technologies as alternatives to the expensive 2.5D silicon interposer, with various flavors (bridge embedded in the substrate or fan-out).

When it comes to wafer level packaging, while fan-in was introduced and continues to target lower-end applications, as shown in Figure 2, given its limitation of I/Os to the periphery of the dies, fan-out has started to get more adoption and expand from lower-end mobile to mid-end and more recently also to high-end networking applications.

WLP technology includes wafer-level chip-size packages (WLCSPs), fan-out wafer-level packages, wafer capping and thin film capping on MEMS devices, wafer-level packages with TSVs, wafer-level packages with Integrated Passive Devices (IPD), and wafer-level substrates featuring fine traces and embedded integrated passives. There are wafer-to-wafer stacking technologies and die-to-wafer bonding that will support stacked die WLP for future products to reduce size and cost. While many of these technologies are still in the developmental stage, they represent solutions to cost and power-level reductions, and performance and/or size challenges for a wide range of products in the future.

2. Market Drivers and Applications for WLP

We live in an information age, in a digital-driven society. While digital transformation is not new, and the interest has been increasing for the past 5 years, digitalization has been significantly accelerated by the pandemic. With home schooling and remote working, more options for entertainment from gaming to on-line video streaming, and social media, reports document increases of up to 70% in internet usage, 12% in streaming, and 10% in personal computers, laptops, tablets. [6]

The market drivers are changing, as shown in Figure 3. Data growth will drive the need for more computing, storage, connectivity, sensing and networking, with many of these functionalities being combined and brought within the same package. While packaging technologies in the past were playing more of a protective role within the supply chain of integrated circuits (IC), currently various packaging and heterogeneous integration technologies are being proposed to address the critical performance, power, and area (PPA) values required for micro-systems. [7]

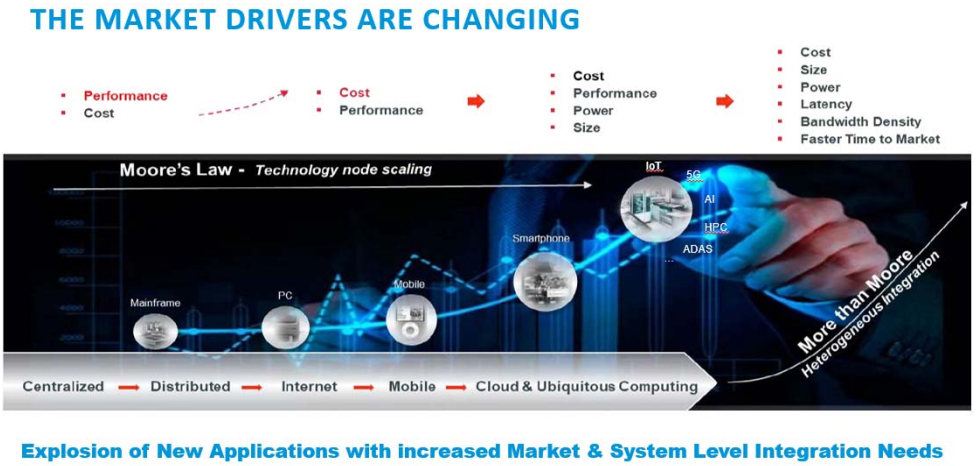

Figure 3: Computing Market Requirements Evolution. [2]

While higher performance, lower power consumption, miniaturization and lower cost continue to be the main requirements for several applications, new requirements, such as low-latency high-bandwidth memory, will bring

new challenges and opportunities for further innovation. To address current and future market requirements, new technologies are being developed: chiplets, bridges, and other innovative and disruptive heterogeneous integration technologies. As Dr. Douglas Yu mentioned in his keynote presentation at the 70th IEEE Electronic Components and Technology Conference (ECTC) in 2020, these new technology disruptions and changes initiate "an exciting new semiconductor era and create a new industry landscape". [7]

Heterogeneous integration, which enables integration of two or more active dies with disparate technologies as well as passive components within one package, can bring several benefits to address the industry's limitations with Moore's Law, as shown in Figure 4. By manufacturing the devices separately and then packaging them together using various advanced packaging platforms, the cost reduction trend can be maintained, while more flexibility can be brought into product development and the manufacturing process. Devices can be manufactured at their optimum technology node and infrastructure, reducing the development costs and time required to develop and bring new systems and products to market. This integration technology provides a pathway to higher performance, smaller form factor, and lower power consumption compared with traditional single-die packages.

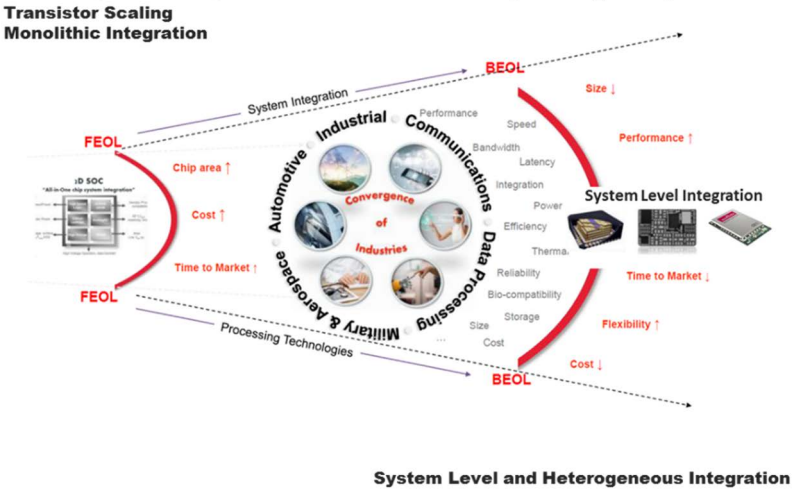

Figure 4. From 2D SoC Integration to Heterogeneous Integration. [8]

Due to its benefits, application of heterogeneous integration has expanded into various applications across several market segments, as shown in Figure 5. Heterogeneous integration today can be found in sensing and MEMS modules, for logic and memory integration, in RF and FEM modules, in wireless connectivity packages and power management modules, with applications across all the market segments, from mobile, IoT, automotive, healthcare, high performance computing and data centers to aerospace and defense.

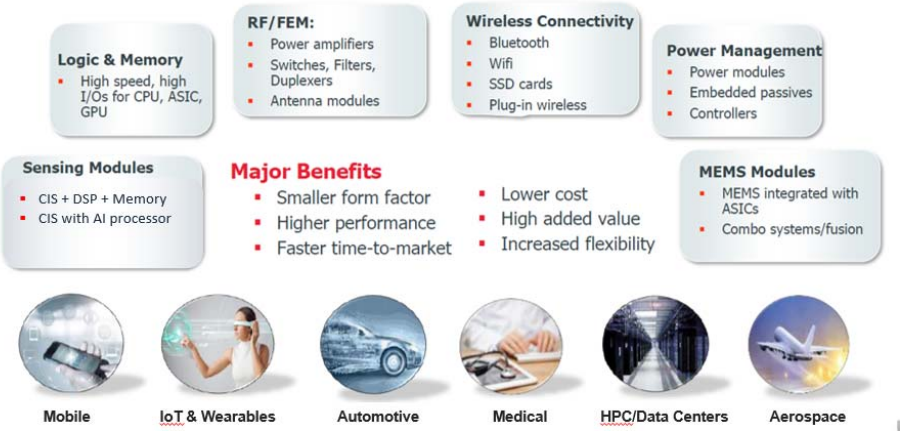

Figure 5. Heterogeneous Integration Benefits and Applications. [2]

As shown in Figure 1 and Figure 6, there are a wide range of advanced packaging platforms that can be used for individual die packaging and system-in-package: from embedded technologies to wafer level packaging, flip chip, 2.xD and 3D integrated packages with through-silicon-vias. Most of the packaging platforms have started with single-die packaging and two-dimensional (2D) integration. With the increase in input/output (I/Os) and the need to address increased functionalities, these platforms have further evolved to address multi-die and three-dimensional

(3D) integration. This chapter will focus on wafer level packaging, which includes fan-in and fan-out packaging platforms, as shown in Figure 6.

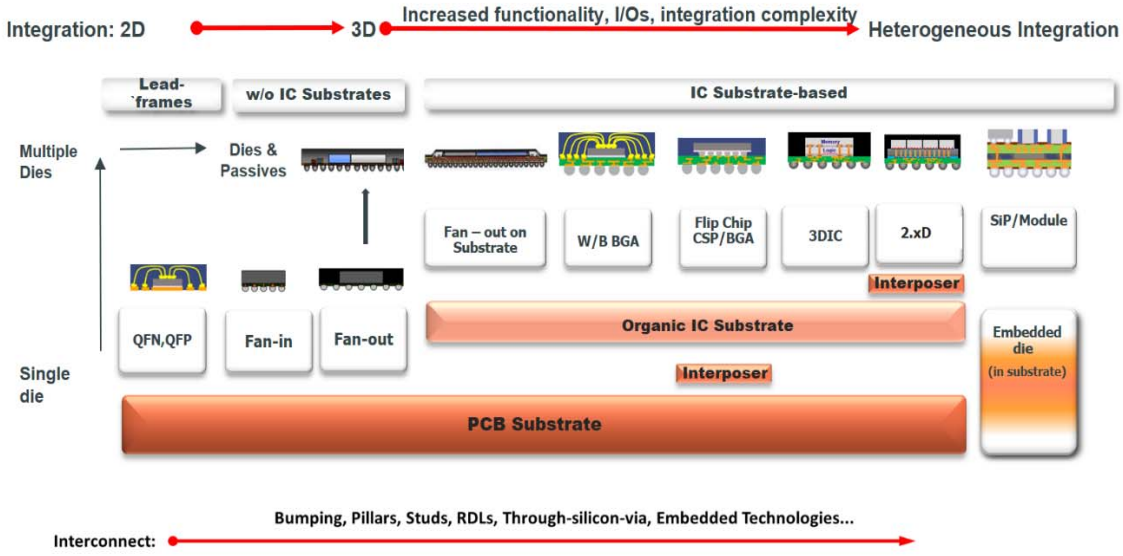

Figure 6. Advanced packaging platforms. WLP, comprising fan-in and fan-out approaches, is the focus of this chapter. [9]

There are several market drivers, as shown in Figure 7, that can be successfully addressed by WLP (both fan-in and fan-out) such as:

- lower packaging and test costs;
- reduced form-factor, since WLP enables thinner packages and smaller footprint;
- better electrical performance due to shorter interconnects that also provide lower parasitics and enable higher system speeds and frequencies;
- better thermal performance and lower power consumption;
- densification and reduction in form factor and integration capabilities of IPDs; or
- enabling system-in-package and 3D packaging.

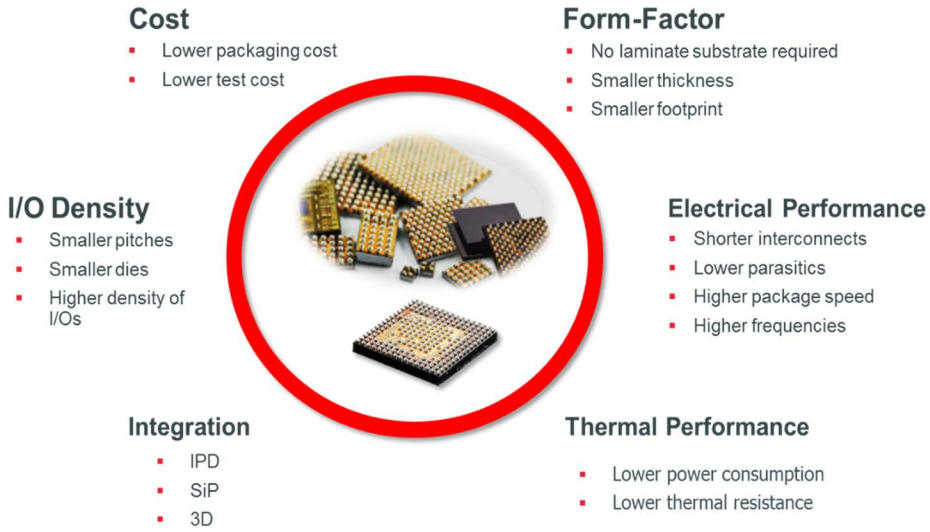

Figure 7. WLP Market Drivers. [9]

If many similarities can be found between fan-in and fan-out, fan-out has some specific market drivers and increased capabilities, while also requiring additional processing steps, as highlighted in Figure 8. Fan-in is mainly applied to the mobile market (90% of its market) while fan-out has also found applications in the high-end market (networking and computing applications) due to its ability to address higher I/O count and larger package form factors; it also has improved reliability performance and increased capability for 3D integration using PoP and SiP heterogeneous integration.

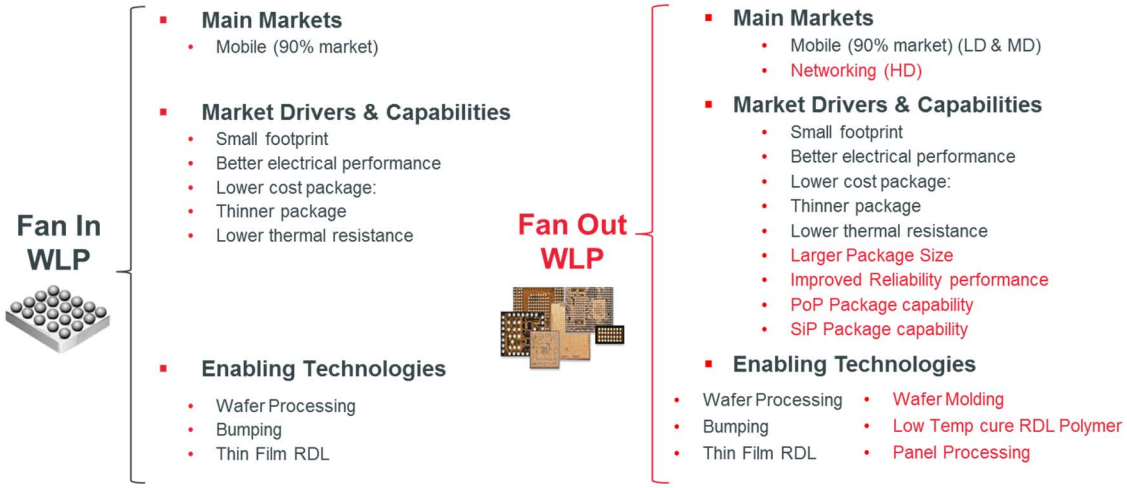

Figure 8. Main Markets, Drivers, Capabilities and Enabling Technologies for WLP. [9]

While wafer processing, bumping and thin-film redistribution lines (RDL) are the primary technologies used for both packaging platforms, molding, low temperature cure RDL polymers and panel processing are additional processing steps required, and alternative infrastructure must be built to support fan-out packages.

Wafer Level CSP (fan-in) was the first generation of wafer-level package product to be introduced into the marketplace. Due to its simplicity, it provides the quickest path to market for new devices and applications with certain package requirements in terms of performance, size, power, thermal and reliability requirements. The key differentiators for this packaging platform are: lowest cost and simplest design, smaller footprint and thinner packages, shorter interconnections, no need for laminate, and a simplified supply chain and manufacturing infrastructure, with shorter lead times. Fan-in has become one of the most mature technologies and continues to be used for a large variety of products, becoming the workhorse for mobile and IoT applications, where high adoption is being seen in SiPs. The high-end smartphones (especially Apple and Samsung) have become the key drivers for fan-in growth in the last five years. Although only 4 such packages were adopted in iPhone 4, the numbers of such packages have significantly increased to 44 in iPhone 7 and 55-60 packages in iPhone X/11. The latest iPhones have fan-in packages ranging from 1x1mm to 7x7mm in size. [10]

The leading applications, as shown in Figure 9, can be found in a wide range of applications in the analog/mixed signal/digital domain such as wireless combo and RF transceivers, MEMS, CMOS image sensors, driver ICs and power-management ICs.

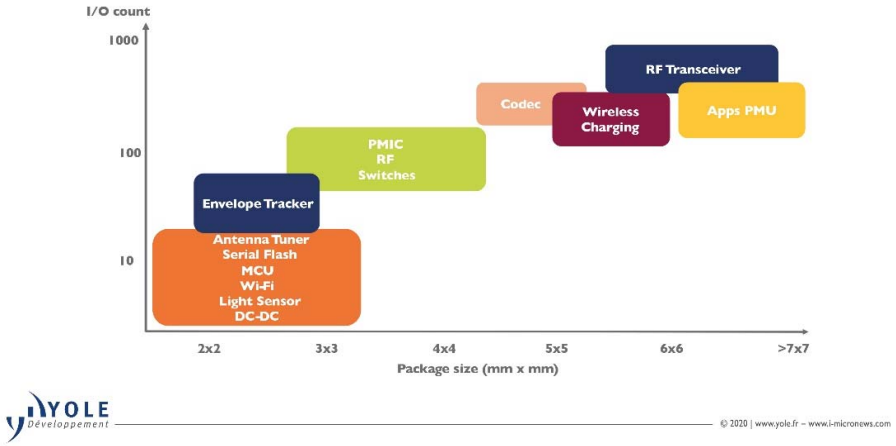

Figure 9. Applications of WLCSP Packaging Platform. [9]

Fan-in continues to offer very attractive solutions at low cost, low form factor, thin and compact packages, ease of assembly, with the needed scalability and reliability, reasons why they will continue to see adoption in smartphones, tablets, and the wearable and consumer segments, such as earbuds and smartwatches. iWatch 5 already included 25-30 fan-in components while the latest Apple Airpods had 14-18 fan-in packages. With 5G deployment and transition from 4G to 5G, the RF content is significantly increasing, driving the switch from wire-bonding to WLCSP packages. While adoption of 3D WLCSP for image sensors is already well known in the mobile and

consumer market, in recent years it has also expanded to other markets such as industrial and automotive. Some of the main applications driving further adoption of 3D WLCSP technologies and image sensors are multi-cameras in smartphones, automotive applications, security and surveillance markets, and automation. Among all these markets, the automotive segment has the highest growth and adoption of image sensors, driven by the need for smaller package sizes and high reliability performance.

Functional integration of fan-in packaged dies into SiP has a considerable impact on the fan-in supply chain, cost and overall demand. While the number of total WLCSP packages continues to grow and is forecast to reach close to 30B units in 2021, adoption in SiP packages was estimated to reach 12.2B units by the end of this year. [10-12]

While WLCSP is a great and simple package and is already well established in the industry, fan-out has a much wider coverage in terms of capabilities and applications. Fan-out technology is one of the most active packaging platforms today, with various flavors and integration schemes being developed and brought to the market. It is a platform forecasted to have one of the highest growth rates in advanced packaging, attracting new players to this market. In terms of applications, fan-out is segmented, as shown in Figure 10, into core, high density and ultra-high density fan-out.

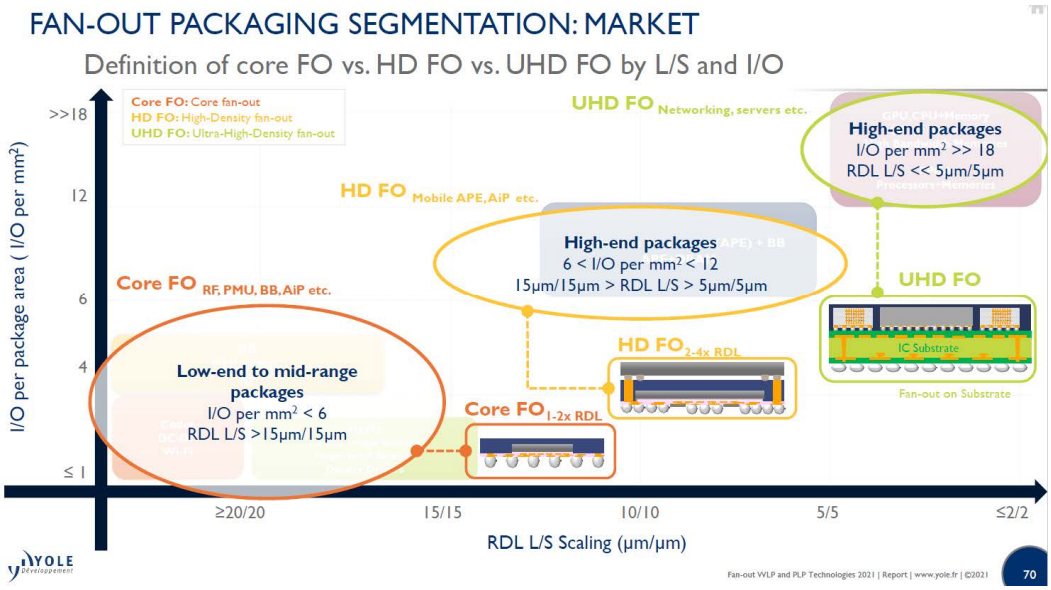

Figure 10. Applications of Fan-out Packaging Platform. [13]

- *Core fan-out:* mainly addressing the lower-end applications such as RF, Power Management, Baseband, Antenna-in-Package
- *High Density (HD) fan-out:* with applicability to application processors in mobile applications
- *Ultra-High Density (UHD) fan-out:* targeting networking and servers, for packaging the higher end processors, integrating them with high-bandwidth memories, enabling partitioning of high-end accelerators or chiplet integration, as an alternative to silicon interposer technologies

Fan-out technology has several advantages that makes it a very attractive packaging platform for many applications, from low- to mid- and high-end, across multiple market segments. It is also more suitable for heterogeneous integration than WLCSP. Some of the key differentiators for fan-out are smaller footprint and thinner packages than Flip-Chip Ball Grid Array (FCBGA) since no laminate is required, and a simplified supply chain and manufacturing infrastructure. Fan-out is not limited to the periphery of the die as would be the case with WLCSP; on the contrary, it has flexibility in adapting the fan-out zone to various needs, therefore providing design flexibility. It also provides shorter connections between dies and bumps, enables scalability in bump pitch, and it was found to have better board-level reliability than WLCSP and better electrical and thermal performance than flip chip and wire-bonded BGA packages [13]

3. Wafer Level Packaging Overview: Technologies, Integration, Evolution and Key Players

WLP has been defined as a technology in which all the IC packaging process steps are performed while the devices are still in a wafer structure before singulation. The original WLP designs required that all package I/O terminals be continuously located within the chip outline (fan-in design), producing a true chip-size package. This structure constituted a fan-in Wafer Level Package with the sequential processing of a complete silicon wafer. From a systems perspective, using this structure, the limitation on complexity of a WLP was how many I/O could be placed under

the chip and still have a board design that can be routed. WLPs can provide a solution when requirements for a continued decrease in size, increase in IC operating frequency and demand for cost reduction are not met by traditional packaging, e.g. wire-bonding or flip-chip bonding.

There are also products that have come to market that are not practical to manufacture using this standard WLP structure. These new packages have been described as "Fan-out" WLP. They are processed by placing individual sawn die into a polymer or other matrix material that has the same form factor as a typical silicon wafer. These "reconstituted" artificial wafers are then processed through all the same processes that are used for "real" silicon wafers, and finally sawn into separate packages. The die are spaced in the matrix such that there is a perimeter of matrix material surrounding each placed die. These embedded devices can have redistribution traces (RDL) designed to "fan-out" to an area larger than the original die. This allows a standard WLP solder ball pitch to be used for die that are too small in area to allow this I/O pattern without "growing" the die to a larger size. With the implementation of this technology, it is no longer only intact silicon wafers that can be processed as a "WLP", but hybrid silicon/other material matrices in wafer form that also can now be loosely classified as WLP products.

With the introduction of TSVs, IPDs, chip-first and chip-last fan-out technologies, MEMs and sensor packaging technologies and heterogeneous integration of processors with memory, a wide range of WLP technologies, with various integration schemes, have been developed and brought to the market. Figure 11 shows some of the integration schemes that have already been adopted over a broader range of applications, from lower I/O counts and applications of WLCSP to higher I/O counts and greater functional complexity, using fan-out technology. These packaging technologies open new opportunities for WLPs in the packaging field. [14-15]

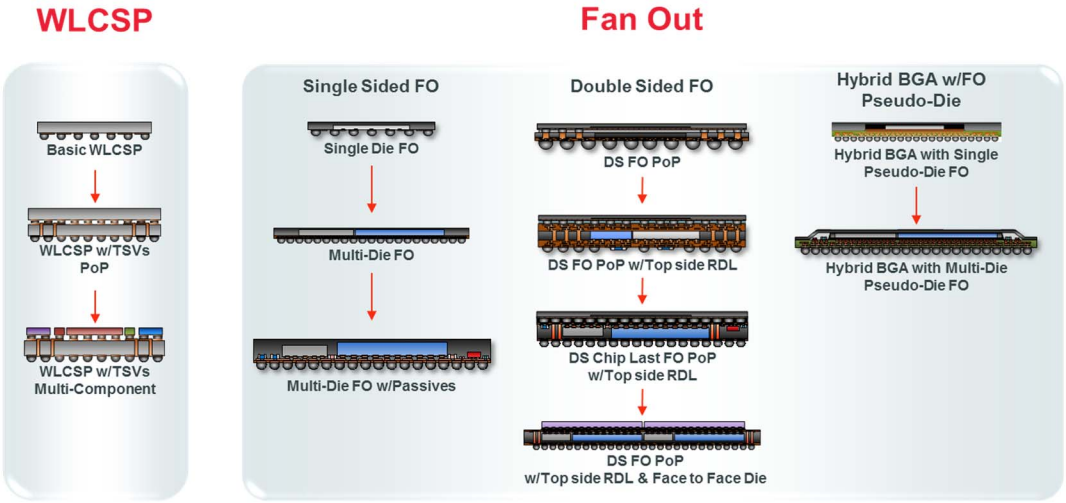

Figure 11. Heterogeneous Integration using WLP. [15]

Various flavors of WLP have been developed. The evolution of WLCSP and Fan-Out will be described below.

3.1 Wafer Level Chip Scale Packaging (WLCSP)

Wafer Level Chip Scale Packaging (WLCSP) came into its own around the year 2000 and has been mostly limited to single-die packaging. By the nature of the package, WLCSPs have limited capability for integration of multiple components. Figure 12 shows a simplistic image of a basic single die WLCSP.

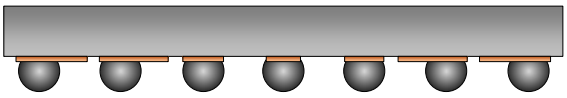

Figure 12. Basic Single Die [14]

Prior to that time, the majority of packaging processes were mechanical, such as grinding, sawing, wire-bonding, etc. The packaging process steps were performed predominantly after die singulation, as illustrated by the simplified process flow of Figure 13.

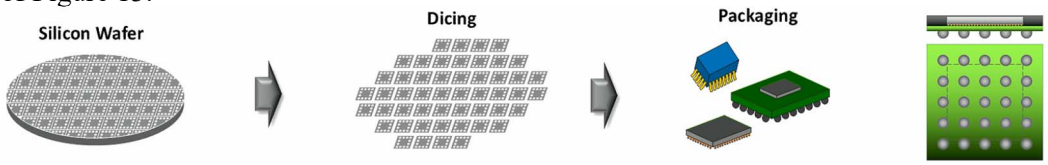

Figure 13. Traditional packaging process flow. [14]

WLCSP was a natural extension of wafer bumping, which had been used since the 1960's by IBM. The primary difference was the use of large solder balls at a coarser pitch than used for traditional bumped die. Unlike previous packaging, nearly all the WLCSP packaging process steps are done in parallel while still in wafer form, as opposed to in a series of steps as in Figure 13. A simplified illustration is shown in Figure 14.

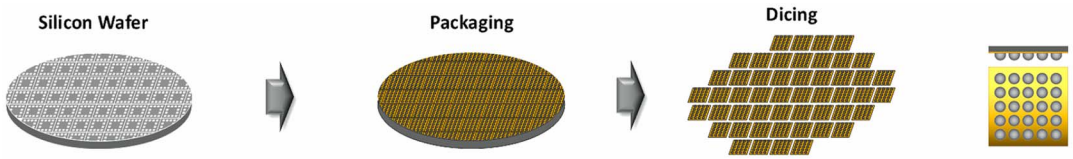

Figure 14. Wafer Level Chip Scale Packaging (WLCSP) process flow. [14]

With WLCSP, since the die itself becomes the package, it is the smallest package that can be manufactured. Because of the size reduction capability, it has become widely used for small mobile applications. The earliest versions were simply solder balls placed on special "Under Bump Metallization" (UBM) that renders the die pad solderable. However, as the complexity of the devices increased, it became necessary to add metal redistribution trace layers in order to route the solder balls away from their respective pads. These redistribution layers (RDL) became the norm, with WLPs increasing in size and complexity. The WLPs were still single-die solutions, and new processes, materials, and structures were developed that allowed at least one additional thinned die to be mounted "opossum" style on the underside of the die, between the existing solder balls. This die was thin enough that the WLCSP could be mounted and still have clearance for the under-mounted die. This became one of the first "heterogeneous" WLPs, as shown in Figure 15.

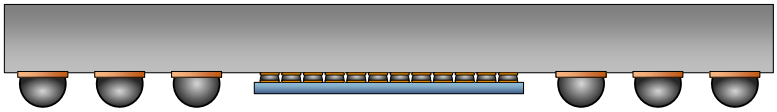

Figure 15. WLCSP with Second Die mounted on Underside. [14]

With the development of Through-Silicon-Via (TSV) technology for 3D applications, TSVs could be formed in the WLCSP, providing double-sided connectivity. While TSV integration uses "via first" and "via last" processes, in the case of WLCSP the "via last" approach was adopted. Such integration enabled mounting of a second die on the top of the primary WLCSP die, or other components, such as passives. This process has been used by the MEMS industry to mount a logic or analog die on top of a MEMS die, or vice versa, as shown in Figure 16. This became another level of WLCSP heterogeneous integration complexity.

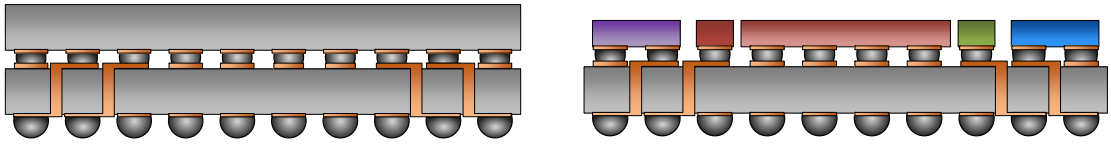

Figure 16. WLCSP Through Silicon Vias for Double Sided Mounting. [14]

Similar integration has been adopted for CMOS image sensors in mobile, and more recently, for sensing applications in automotive [16-18]. Due to shorter electrical interconnects, smaller form factor and low-cost packaging solutions, 3D WLCSP with TSVs is replacing the traditional packaging that used chip-on-board (COB) technology. Main challenges, as for most of the automotive applications, is meeting reliability requirements. For example, the package shown in Figure 17, developed for backside illumination (BSI) of the CMOS image sensor for automotive application, is using a 5.82mm x 5.22mm package with a thickness of 850um, 3:1 TSV aspect ratio and a silicon-to-package ratio of 99.27%. [16-17]

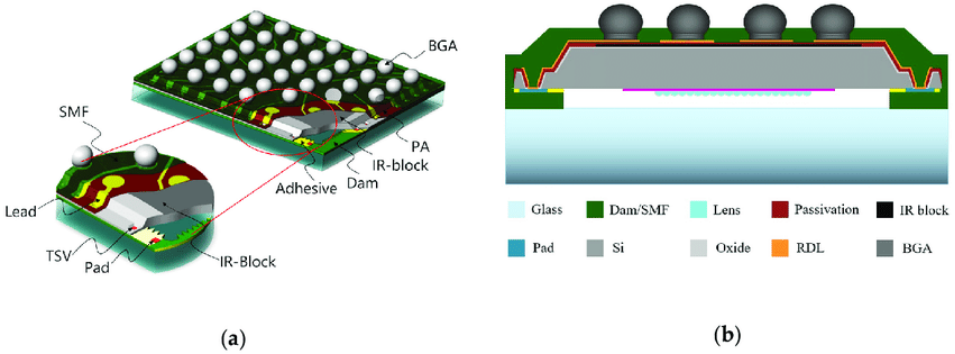

Figure 17: (a) 3D view of the CIS-WLCSP structure; (b) Cross-section of the CIS-WLCSP. [16]

As the technology nodes advance, there are more challenges for reliability and chip-package interaction (CPI) as the size of WLCSP packages increases. It is not only about the reliability performance; it is the adverse effects that can occur during subsequent processes after WLCSP manufacture, from shipping and handling to final assembly onto the board. New disruptive technologies were introduced, such as the fan-in M-Series products (based on a Deca license), to address the increasing interest in side wall protection. 6-sided (6S) protection now has become an industry gold standard, ensuring a high degree of board-level reliability.

The overall market continues to be dominated by the large OSATs (ASE/SPIL, Amkor and JCET); however, in recent years, foundries such as TSMC and Samsung have also entered the market and are offering full turnkey solutions. Integrated device manufacturers (IDMs) such as TI, NXP and ST Micro continue to be a critical part of the WLCSP supply chain. [10]

3.2 Fan-out Wafer and Panel Level Packaging

While advances have been made in WLCSP and high adoption of such packages for SiPs have been observed in the industry, the packages are still limited in physical area to the actual die size. As wafer nodes advance, with geometries shrinking, the die themselves could be reduced in size, bringing more challenges to this type of package. This created a dilemma – where to place the solder balls at the coarser pitch required for WLCSPs. To address this requirement, the "Fan-Out" packaging platform had been conceptualized and was in various stages of development as early as 1983 [19]; however, it went into volume production much later, in May of 2009 with the Infineon eWLB product. This was a simple single-die package, as shown in Figure 18.

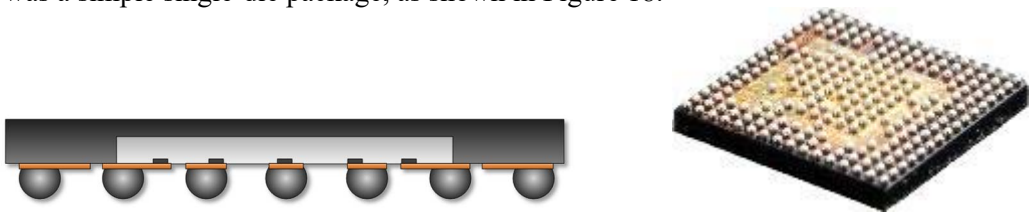

Figure 18. eWLB Fan-Out. [14]

The first two primary commercial contenders were Motorola/Freescale with their Reconstituted Chip Package (RCP) technology, and the above-mentioned embedded Wafer Level BGA (eWLB) technology developed by Infineon, both shown in Figure 19. These two resulted in similar Fan-Out package structures, with processing variations.

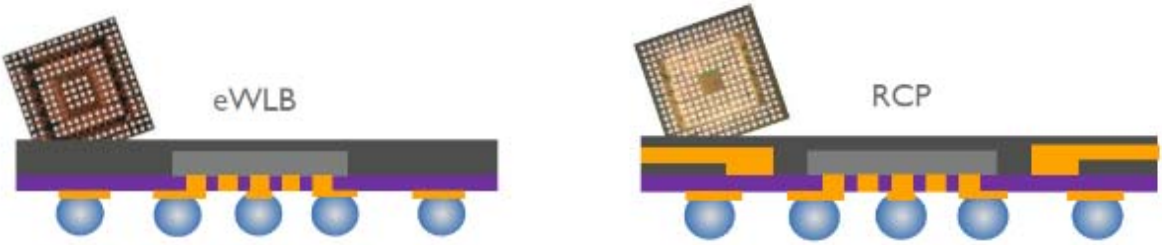

Figure 19. The First Two Fan-Out Concepts: eWLB and RCP. [13]

- *RCP technology:* The RCP included a copper frame layer embedded in the mold compound along with the die, acting as a potential ground plane and/or a stabilizer for alleviating TCE mismatch to the printed circuit board after final assembly, as shown in Figure 20.

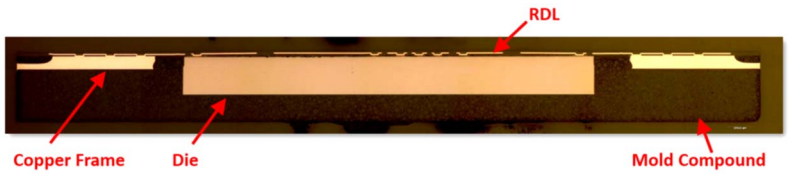

Figure 20. WL Fan-Out packaging with copper frame. [20]

While RCP seemed to have better performance in preventing die shift, it had limited adoption in the market. It was a technology that was developed earlier; however, it took until 2011 to make it into manufacturing. NXP was the first one to bring RCP to manufacturing for its 77GHz radar applications, in collaboration with Nepes.

- *eWLB technology:* Infineon took eWLB into volume production in early 2009. Both processes were an extension of standard Wafer Level Chip Scale Packaging (WLCSP) processing technology, with the "wafer level" processing performed on a plastic molded reconstituted wafer instead of the standard silicon wafer. The die were first singulated and embedded on five sides in a mold compound, leaving the die pad side exposed. The molded "reconstituted" wafer was then processed in a similar manner to WLCSPs, with modifications to the materials, equipment and processes to accommodate the variability of the molded wafer. eWLB had the advantage of having a simpler process flow due to fewer process steps, a reason why it had a higher adoption in the market

The similarity to the WLP process flow can be seen in Figure 21. With the additional molded area, the final packages become larger than the die sized WLCSP.

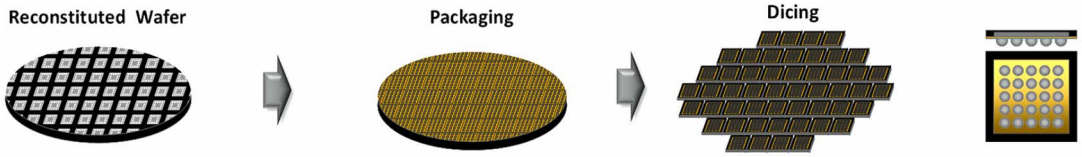

Figure 21. WL Fan-Out packaging process flow. [14]

Although Casio described their EWLP package as a "Fan In/Out Package" in 2006 [21], the term was not generally used to describe a reconstituted wafer package until Infineon began describing their eWLB package as "Fan-Out WLB". The term "Fan Out" has since gained widespread acceptance in the packaging industry. Infineon, in their patent for the eWLB technology, described all previous packaging in this way: "Conventional packages or casings for circuit units are therefore constructed using a so-called 'fan-out design". The only package that does not "fan-out" the interconnections is a WLCSP, as it is die-sized, and can only fan inward.

The eWLB and RCP fan-out packages were processed as a chip-first structure, in that the die was molded into a reconstituted "plastic wafer" initially, before the interconnecting trace redistribution layer (RDL) was added to the device. eWLB and RCP can be classified as "die down" chip-first processes, as the die is placed on the temporary carrier before overmolding in a die-face-down position. Figures 22 and 23 give a simplistic process flow of the chip-first and die-down eWLB and RCP structures.

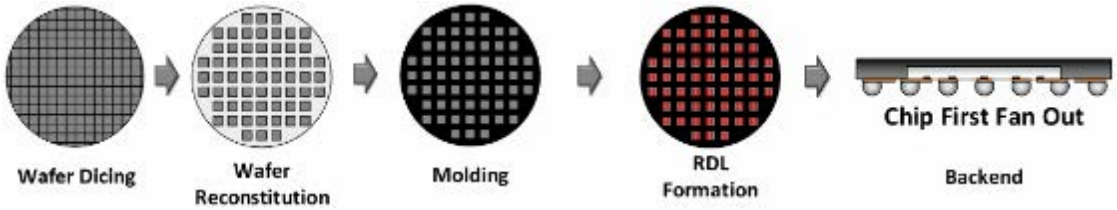

Figure 22. Chip-First process flow. [14]

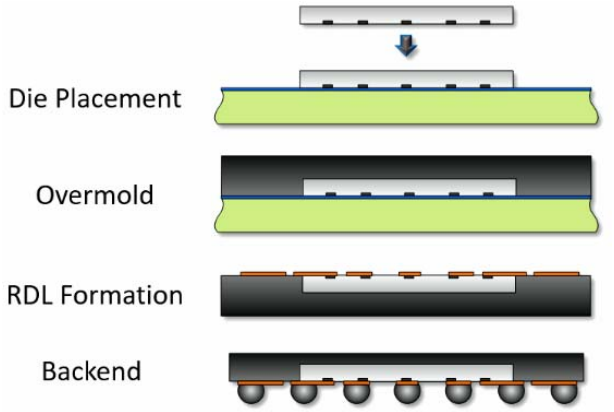

Figure 23. eWLB Fan-Out Die-Down process flow. [14]

This same process can be extended to allow for the inclusion of multiple die and/or passive devices within the mold compound of the reconstituted fan-out wafer as shown in Figures 24 and 25. Multiple suppliers began producing what we today call fan-out packages with multiple die and passive components integrated into heterogeneous packages in engineering or limited production, as early as the mid 1990's.

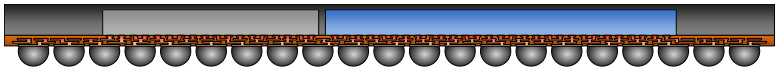

Figure 24. Multi-Die eWLB type Fan-Out. [14]

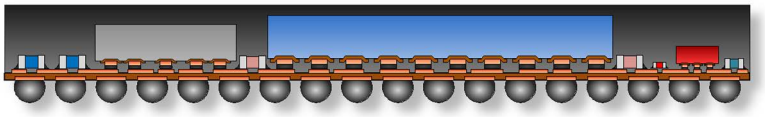

Figure 25. Multi-Die eWLB Fan-Out with passive components. [14]

The chip-first reconstitution fan-out process can also be used with a die-up process and structure. A simplified image of a fan-out package with this structure can be seen in Figure 26.

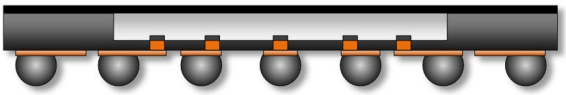

Figure 26. Chip-First Die-Up Fan-Out Package. [14]

A simplified process flow for the die-up fan-out is illustrated in Figure 27.

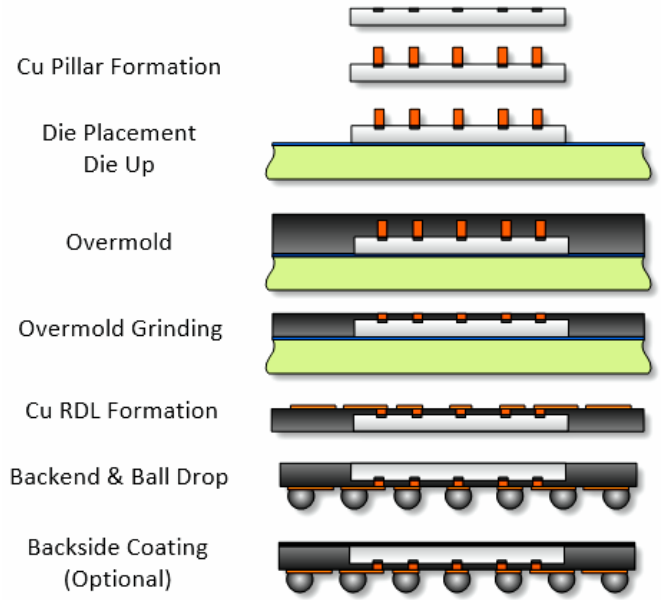

Figure 27. Fan-Out Die-Up process flow. [14]

A third variation of fan-out is chip-last fan-out. In this version the trace RDL pattern is produced on a temporary carrier using the same thin film RDL manufacturing processes as chip-first fan-out. The die are first bumped, typically with copper pillar bumps while still in wafer form, singulated, flip-chip assembled onto the RDL pattern, and then overmolded with mold compound. This fan-out structure and process are similar to a standard flip-chip BGA; the only difference is that the interconnecting trace pattern is formed using a thin-film RDL process on a temporary carrier. A simplified process flow is illustrated in Figure 28.

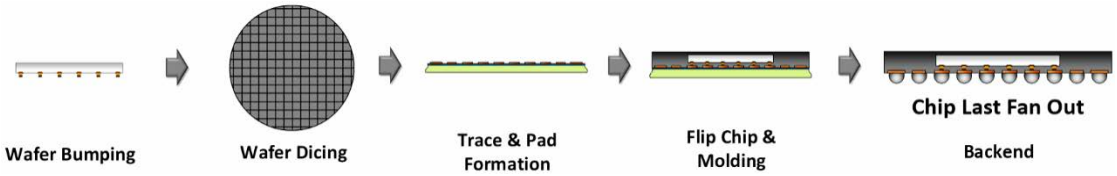

Figure 28. Chip-Last process. [14]

The first eWLB in volume production was a single-die package, combining baseband with PMIC and RF features. The dies were about 5x5mm in 8x8mm fan-out packages that varied between 183 and 217 solder balls. Fan-out has evolved over the years, leading to even smaller and thinner packages, as would be the case of the ultra-small and ultra-thin (0.8mm x 0.53mm x 0.14mm) Encapsulated Chip Package (ECP), shown in Figure 29, a fan-out technology developed by JCET. ECP technology allows for not only fan-out single-chip and multi-chip packaging, but also ensures five-sided package protection. It has a small size and large fan-out ratio, and can effectively overcome the wafer warpage problem, which is one of the main fan-out challenges.

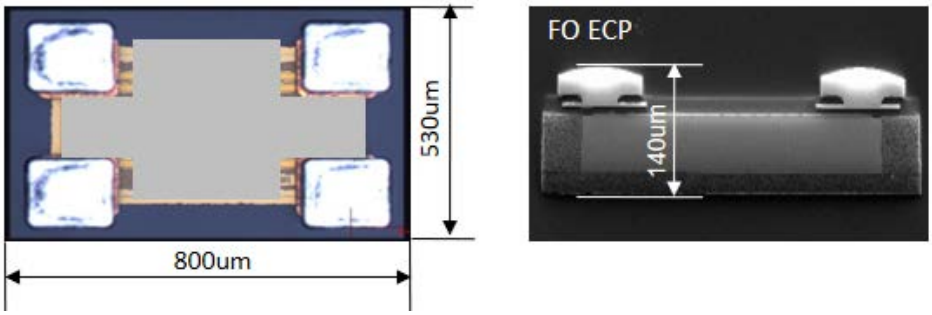

Figure 29. Ultra-Small/Thin Fan-Out (FO) ECP by JCET. [22]

This ECP technology uses a chip-first and face-down packaging process, and a laminating molding film instead of the conventional liquid and powder molding compounds. This lamination process replaces the wafer molding process for ultra-thin packaging of the chips and allows the reconstituted wafers to achieve a high degree of flatness while avoiding cavities in the wafers. Additionally, it can effectively reduce the chip offset problem and achieve small-size, large fan-out ratio packaging. Due to low modulus characteristics of the lamination film combined with the silicon support on the backside of the reconstituted wafer, the ECP process can effectively reduce and overcome warpage related problems. [23]

Mass production of fan-out started with smaller, low density, core fan-out packages. Although the majority of early fan-out activity was focused on 2D single-sided applications, 3D and double-sided fan-out structures were also explored by 2004. Various methods of interconnecting the bottom and top dies of a fan-out package have been evaluated, with at least seven structures tried in engineering modes. Two different processes/structures are currently being used either in engineering and qualification or volume production. One is the use of some form of preformed structure with vias formed through either an organic or inorganic carrier material, such as circuit board or silicon die. These are embedded simultaneously with the die, usually in a chip first, die-down type of fan-out. This is illustrated with a simplified cross-section graphic in Figure 30, with expanded illustrations of the 3D interconnects for fan-out in Figure 31.

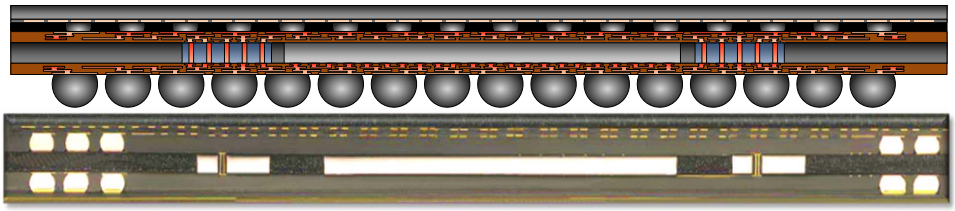

Figure 30. 3D Fan-Out with Embedded TSV Die. [14]

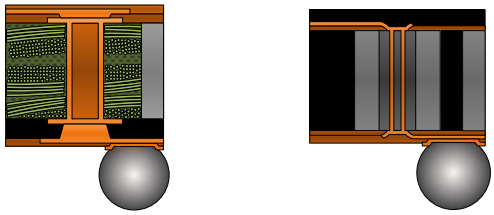

Figure 31. 3D Fan-Out Interconnect using circuit board via or TSV in silicon. [14]

One possible process flow for this type of 3D structure is shown in Figure 32.

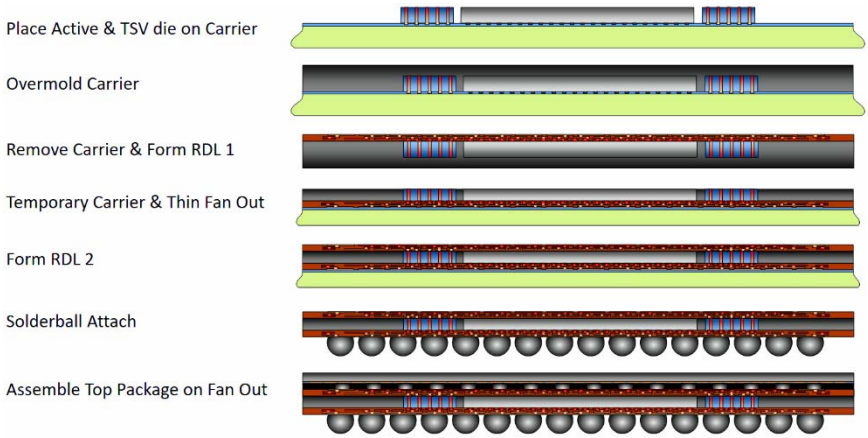

Figure 32. 3D Fan-Out with Embedded TSV Die process flow [14]

A second method involves plating a copper post onto a temporary carrier prior to die attach, and grinding the mold material to expose the copper posts on the top of the fan-out package. A second RDL layer is then formed on the top side of the fan-out, making connections to the exposed copper posts, with a separate device mounted onto this top RDL trace layer, as shown in Figure 33 with the expanded illustration of the 3D interconnect for fan-out in Figure 34.

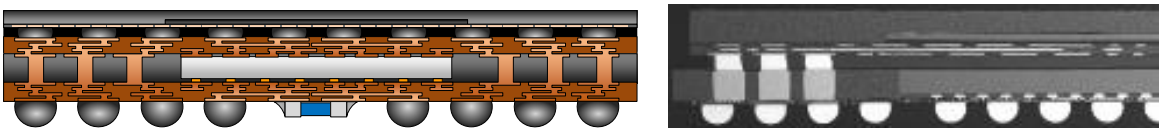

Figure 33. 3D Fan-Out with plated copper through posts [14]

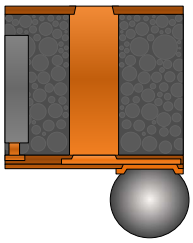

Figure 34. 3D Fan-Out Interconnect with plated copper through post. [14]

A process flow for this type of 3D structure is illustrated in Figure 35.

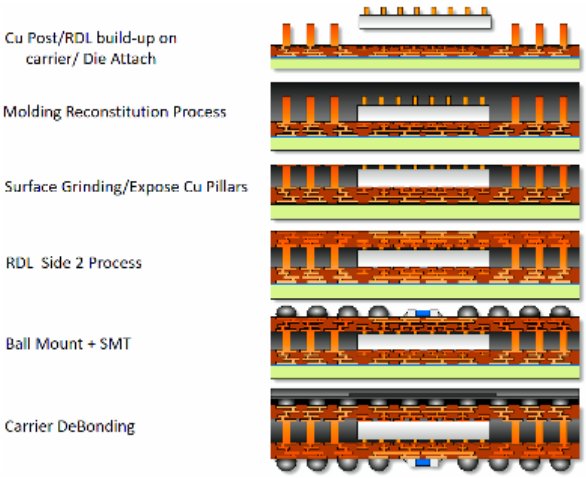

Figure 35. 3D Fan-Out with plated copper through posts process flow. [14]

The best known example of this technology is also the highest-volume fan-out product that has been in production since 2016 when Apple introduced its iPhone 7. Other than the routinely upgraded advanced CMOS front-end process node, iPhone 7's Application Processor (AP) was packaged using a revolutionary wafer-level-packaging technology called "Integrated Fan-out" (InFO), innovated and developed by TSMC.

At first glance, InFO is just another fan-out technology with mold compound embedding silicon chips and Cu interconnects. However, an insightful look reveals its genuine innovation and engineering prowess. Technologically, InFO created several "industry firsts in wafer level package":

- Packaging size larger than 10 x 10mm, Chip size larger than 8 x 8mm with very high BGA pin count.
- First high-volume manufactured multi-die fan-out that included logics and passives
- First 3D PoP integration of logic and memory package with high-density multi-row backside BGA arrays
- First chip-first face-up embedded-die fan-out technology
- First mobile AP PoP package with final thickness below 1mm
- First foundry-delivered advanced fan-out technology in mass production

InFO demonstrated to the semiconductor industry how an old technology can be re-innovated and transformed into a leading-edge packaging platform. The success has triggered an avalanche of renewed research interest in fan-out and chip embedding technologies. This happened worldwide and included foundries, OSATs, research institutes and academia. Subsequently, a cascade of similar wafer-level and panel-level packaging technologies were announced. InFO triggered a renaissance in the packaging industry, helped to maintain Moore's Law and enable More Than Moore. It helped to bring the packaging industry to the center stage of the semiconductor industry.

Mass production of fan-out started with smaller, low density, core fan-out packages. Then, in 2016, two high-density, more-complex fan-out products hit the market. The higher volume fan-out product was the Integrated Fan-Out (InFO) package from TSMC, first used as a ~15mm x 15mm Fan-Out Package-on-Package (FOPoP) product with over 1300 solder balls. It was used in implementing the Apple A10 processor with a standard DRAM memory package assembled on top of it, as shown in Figure 36 [24].

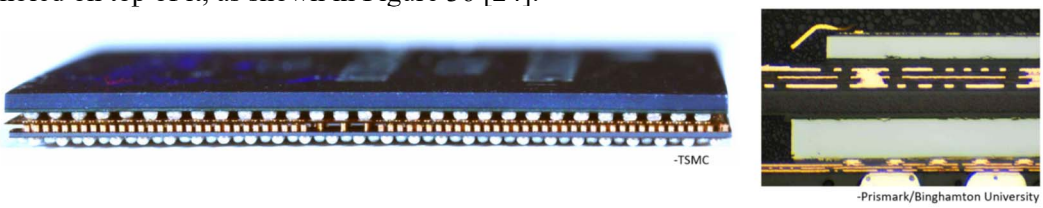

Figure 36. TSMC's Integrated Fan-Out package (InFO). [24]

A second high-density fan-out product was released by ASE, called Fan-Out Chip-on-Substrate (FOCoS), which was a hybrid combination of a fan-out composite die mounted on a BGA substrate, targeting high-end networking and server applications. For such high-end applications, typically packages had multiple high-density die mounted on a silicon interposer with TSVs to interconnect the die and to fan-out the connections to a larger pitch, with this subassembly mounted onto a BGA substrate. Although effective, it has been an expensive package, encouraging the search for a lower-cost alternative. This has led to the use of fan-out technology in a hybrid form of package for low to medium density multi-die applications. The FOCoS product combined two large dies into a 32mm x 25mm fan-

out bumped die, rather than a stand-alone package. The dies are embedded in a fan-out assembly using lines and spaces (L/S) down to 2 microns, with multiple layers of RDL for interconnectivity. The fan-out assembly is not treated as a standalone package, but has fine pitch solder or copper pillar bumps instead of the large solder balls that a fan-out package would normally have. This fan-out assembly is treated as if it were a composite- or pseudo-die, and is then flip-chip assembled onto a BGA substrate in the same manner as the interposer package described above. An example of this hybrid package on substrate which ASE calls FOCoS (Fan-Out Chip-on-Substrate) is shown in Figure 37.

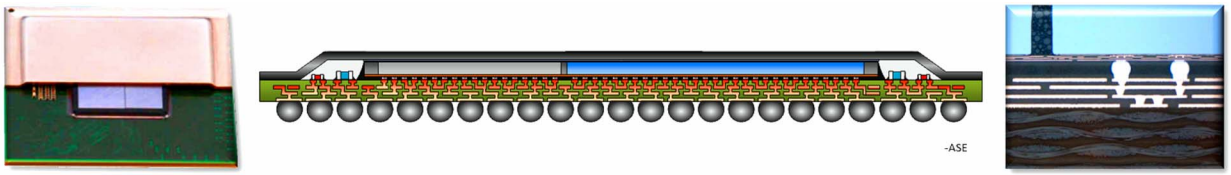

Figure 37. ASE's Fan-Out Chip-on-Substrate (FOCoS). [14,25]

Both the InFO and the FOCoS are multi-die packaging solutions; therefore, both are new forms of heterogeneously integrated packages with high I/O counts and fine-trace geometries. The InFO combines a fan-out bottom package with a standard DRAM mounted on the top side through connections of the InFO fan-out, making it a 3D-connected package. The InFO subsequently went into very high-volume production for smart phones, and in effect kick-started renewed interest in fan-out technology. It showed the industry that high-density fan-out was a viable and successful alternative to traditional packaging. FOCoS fan-out was the first attempt at replacing an interposer solution for the interconnection of the two large high-I/O count die with a less expensive solution. The FOCoS was a specialized product for networking and server applications, and it was produced in relatively lower volumes, but usually at very high RDL densities. Both of these fan-out packages incorporated 3 RDL layers, with much higher complexity than previously produced fan-out packages. Since 2016, fan-out has evolved into more complex structures, with multiple process variations and new players entering the market.

InFO, and wafer-level system integration (WLSI) as a whole, has triggered a new wave of re-thinking about chip-package-system co-design, as the boundaries across the silicon-package-substrate-system are becoming more blurred. The impact has propagated through the semiconductor supply chain, upward to EDA, fabless, layout, and downward to testing, reliability, and system houses. Figure 38 shows a top view of the copper through-via post layouts for the TSMC InFO package. [26-27]

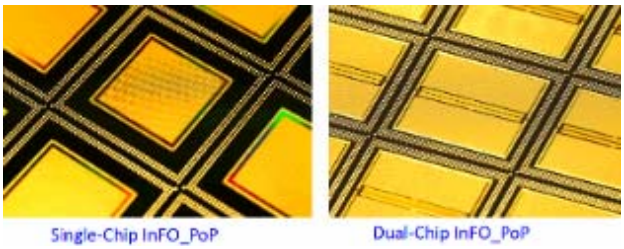

Figure 38. Top view of InFO Packages showing 3D Through-Via Pads. [26]

New processes and structures will continue to evolve as suppliers strive to develop alternate methods to interconnect die in 3 dimensions, such as the technique reported by Nepes of using a deep photoimageable via for the layer-to-layer connection, as shown in Figure 39.

Figure 39. Nepes 3D Fan-Out Package with Deep Photo Defined Vias. [28]

The state-of-the-art solution (as of 2021) that in effect synergizes the InFO and the FOCoS technologies is called InFO-based "InFO_oS (InFO on Substrate)" which features multiple tiers of high-density 2/2µm L/S RDL to integrate multiple advanced-node chiplets for optimizing cost and performance. With InFO_oS, TSMC has demonstrated the industry's first 2.5x reticle size of fan-out (2100 mm^2); assembled with a 110mm × 110 mm package substrate. [26]

TSMC has taken the fan-out concept one step further, incorporating multiple SoCs and memories into a multilayer stack-up using an advanced structure called 3D Multi-stack (MUST) system-integration technology; the 3D MUST-

in-MUST (3D-MiM) fan-out package [29-30] addresses the growing need for high bandwidth with low latency for high-speed data communications. An example of this is shown in Figure 40.

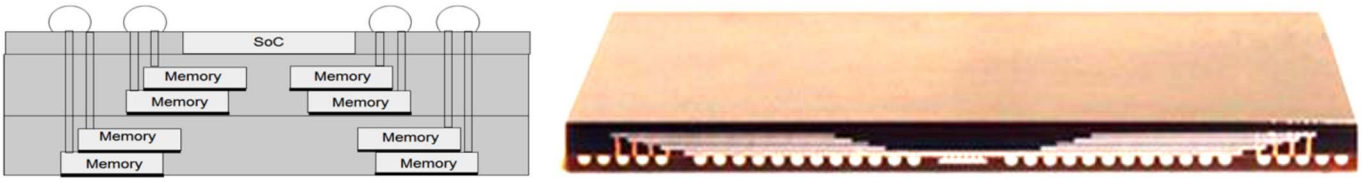

Figure 40. TSMC 3D-MiM Package. [29]

By leveraging current infrastructure for yield and capacity, TSMC was able to provide a logic/memory integration at a competitive cost. 16 memories are embedded in two fan-out tiers, with the system-on-chip (SoC) integrated into the 3rd tier using inFO technology. The package is an alternative to flip-chip and fan-out package-on-package (POP) needed for mobile or computing devices with a thinner profile but high memory capacity and bandwidth. [29]

TSMC has developed a comprehensive portfolio of 3D packaging solutions with integration technologies both in the front end and back-end, as highlighted in Figure 41.

- *Front End:* using the above-mentioned System on Integrated Chips (SoIC) with the option of chip-on-wafer and wafer-on-wafer processing
- *Back End:* using two very well established platforms, Chip on Wafer on Substrate or CoWoS (with silicon interposer) and Integrated Fan-out or InFO (using RDL interconnect), both with the option of integrating local silicon interconnect (LSI) - silicon bridges, as shown in Figure 42.

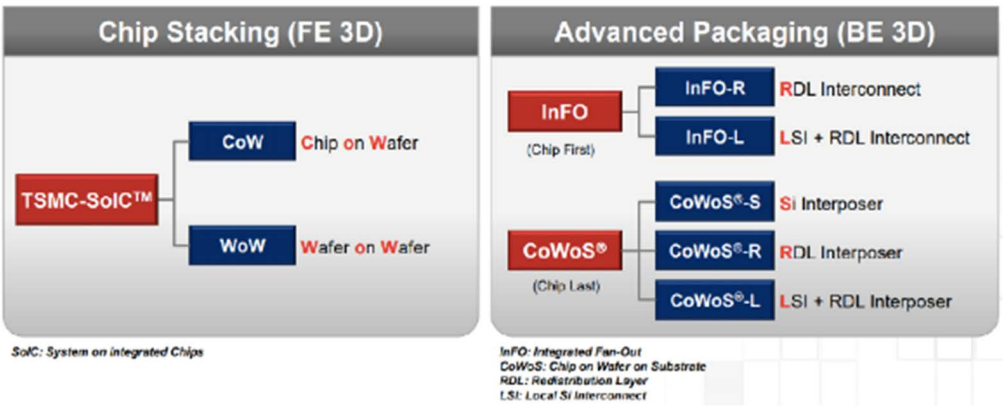

Figure 41. TSMC 3D FabricTM – The Ultimate System Integration Composition. [31]

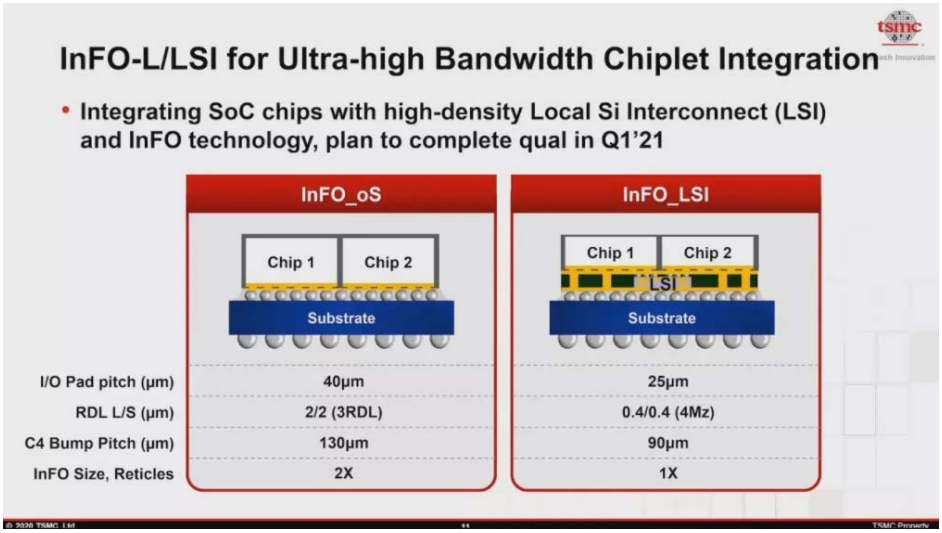

Figure 42. TSMC's InFO-L/LSI for Ultra-High Bandwidth Chiplet Integration. [31]

InFO_LSI, introduced in 2020, is the embedded-bridge technology that TSMC has developed to provide local interconnect solutions (similar to silicon bridge) for chiplets with even more advanced, submicron L/S and 90um C4 bump. [32]

Various flavors of embedded local interconnect technologies have been developed in recent years. Intel was the first one to develop and bring a bridge technology to the market with Embedded Multi-Interconnect Bridge (EMIB). It was designed to meet the needs of high-end communications systems and introduced in production for the first time in the Altera Stratix 10 FPGA family. This technology introduced for the first time a thin piece of silicon (< 75um), with very fine-pitch interconnects (~ 2um L/S), embedded in an organic substrate, as a bridge to connect the dies, as shown in Figure 43. [33]

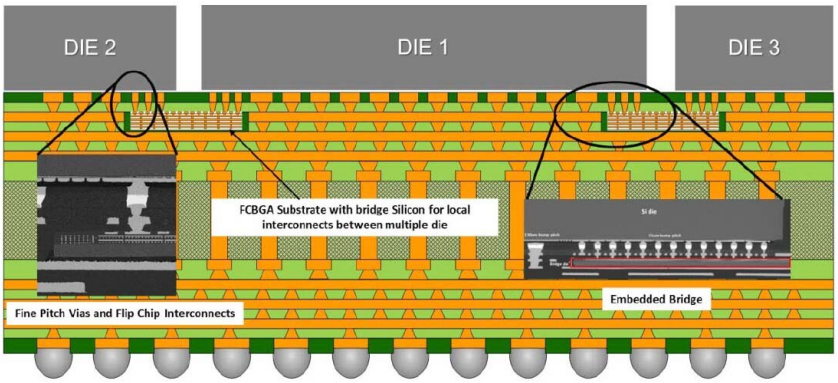

Figure 43. Intel's Embedded Multi-Interconnect Bridge (EMIB) technology. [33]

While Intel has used the approach of embedding the bridge in the substrate, TSMC has positioned the bridge above the substrate, in the fan-out package, maintaining the traditional assembly processes. Besides Intel and TSMC, SPIL and IBM have also developed their own bridge embedding solutions as alternative technologies to the expensive 2.5D Silicon Interposer [34-36], and recently AMD has announced its first product adopting a bridge technology, called Elevated Fan-out Bridge (EFB), for the MI200 processor. [37-39]

The technology developed by Siliconware Precision Industries Co., Ltd. (SPIL) pertains to a scalable chiplet package technology, namely Fan-Out Embedded Bridge (FOEB), which is shown in Figure 44.

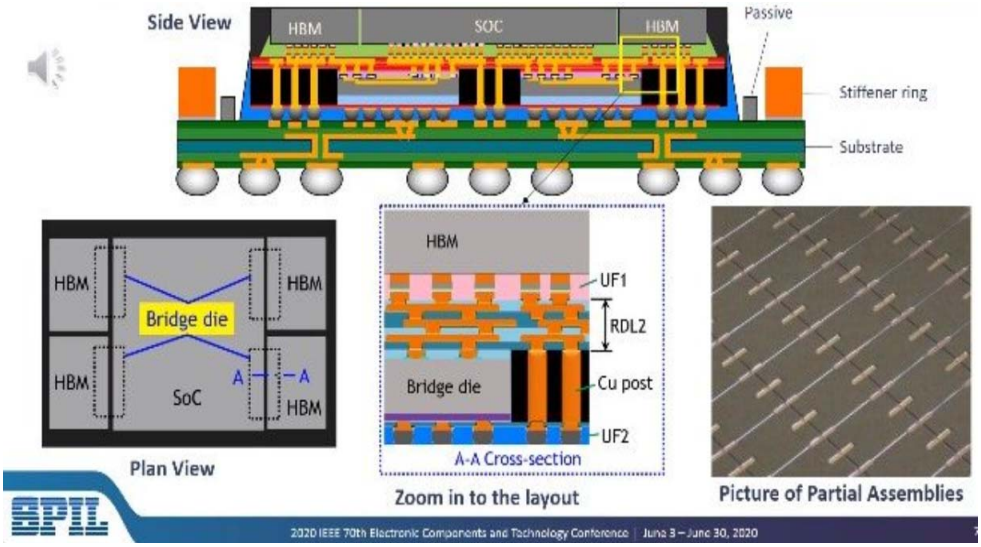

Figure 44. SPIL's Fan-Out Embedded Bridge (FOEB). [35]

By means of Bridge (Dummy) Dies, the FOEB solution enables "near-monolithic" BEOL connections between multiple dies. It can achieve multiple RDL layers and silicon bridges that offer much finer L/S than prior arts, such as the 2.5D TSV-based Silicon Interposer, 3D Die Stacking (e.g., Die-to-Wafer), and Fan-Out Multi-Chip Module (FO-MCM). [34-35] The process flow is shown in Figure 45 and main characteristics comparison with EMIB technology in Table 1:

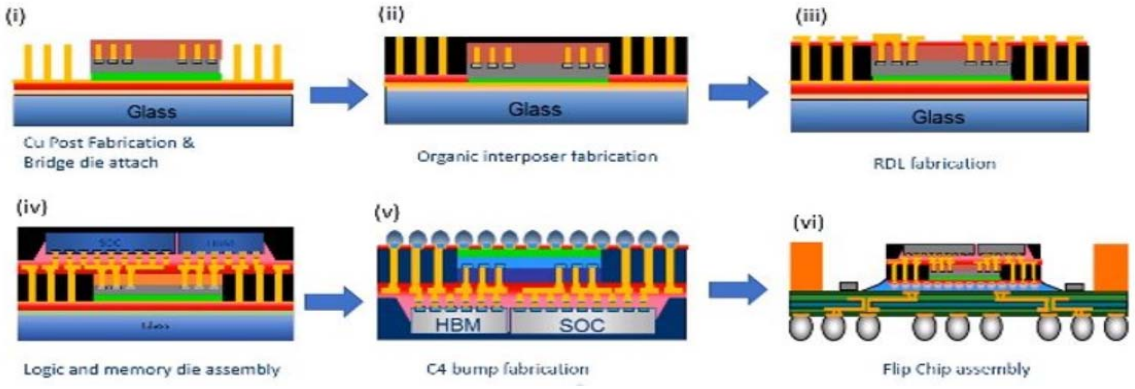

Figure 45. SPIL's Fan-Out Embedded Bridge (FOEB). [35]

Table 1. Platform Comparison: SPIL's Fan-Out Embedded Bridge (FOEB) vs EMIB. [35]

Platform (Ref. Flip Chip MCM)		FOEB	EMIB
Manufacturing values	*Supply Chain*	Simple	Complex
	Assembly Cycle time	High	Std Flip chip process (Exclude substrate fabrication time)
	Yield (%)	> 99	80 ~ 90% (Include substrate embedded bridge)
Chip	*Die QTY*	3 ~ 30	3 ~ 16
	Interposer	Organic/RDL	Organic Substrate
	I/O density	Very high	Std. Flip chip
	µ-Joint (pitch, µm)	25~40	-
	Cu Line (L/S, µm)	0.8/0.8 ~ 10/10 (Scalable)	> 5/5
	Chiplets integration	Excellent	Excellent
	Design Scalable	Good	Good
	Warpage	Low	Low
Package	*Size (mm^2)*	> 55*55	> 55*55
	Coplanarity	Comparable	Comparable
	C4 Stress	Low	Depend on die size

IBM's version of embedded bridge technology, called Direct Bonded Heterogeneous Integration (DBHI) Silicon Bridge Technology, uses direct Copper Pillar connections between chips from a silicon bridge to form a subassembly which is then joined to a laminate chip carrier with a standard C4 pitch, as shown in figure 46.

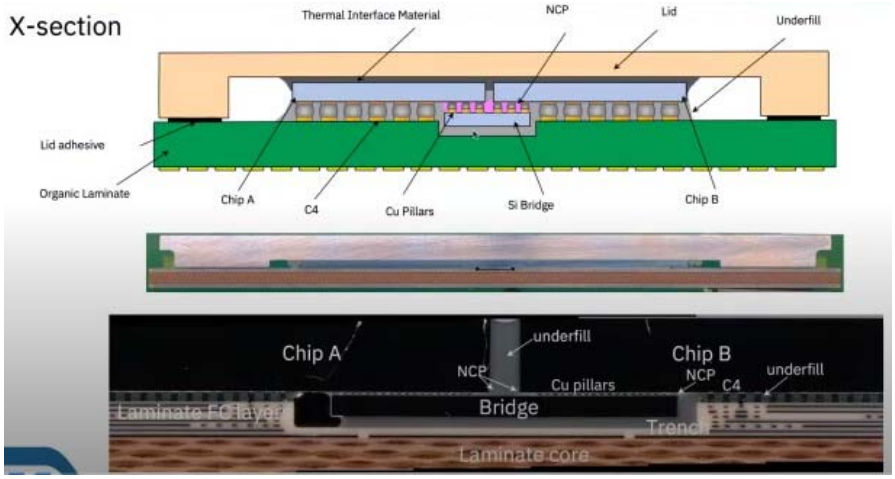

Figure 46. IBM's Direct Bonded Heterogeneous Integration (DBHI) Silicon Bridge Technology. [36]

The major difference of IBM's bridge technology versus Intel's EMIB technology is that the chips are attached to the laminate using standard C4 bumps, traditional assembly processing, and connected with the bridge that has the microbumps. There is no need for multiple-pitch bumps on the same device (die or the bridge) as would be the case for EMIB technology [36].

The latest announcement and bridge technology came this year from AMD. AMD's Elevated Fan-out Bridge (EFB) technology, shown in Figure 47, is built above the substrate, similar to SPIL's FOEB or TSMC's InFO_LSI. The graphic processor unit (GPU) and high-bandwidth memory (HBM) stack are placed on top of a mold using copper pillar interconnects. These copper pillars allow the coarse-pitched contacts on the chips to make contact with the substrate below using traditional assembly techniques, while using a silicon bridge for the finer L/S and small pitch microbumps needed for HBM. Raising the HBM and GPU has created room for the silicon bridge without the need to embed the bridge in the substrate. [37-39]

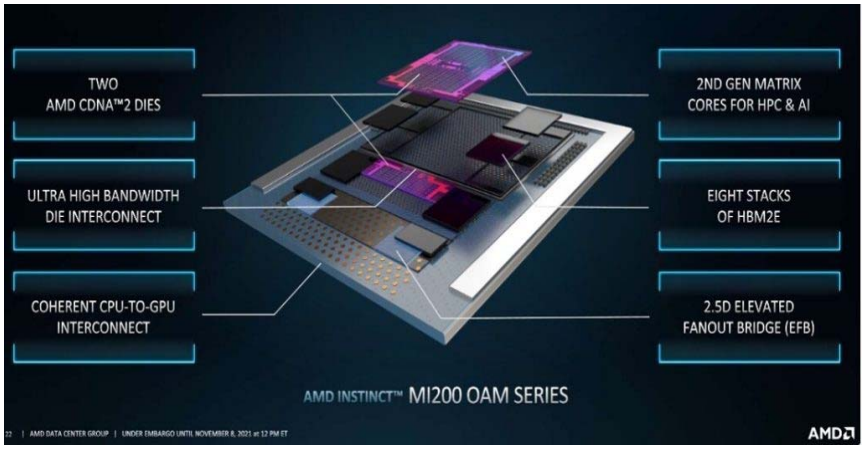

Figure 47. AMD Elevated Fan-out Bridge Technology. [37]

When compared to traditional 2.5D interposer integration, which was used on AMD's previous generation MI100 processors, there are several benefits highlighted in Figure 48: there is no need to use a large and complex silicon interposer, therefore the cost will be lower; and since the bridge is above the substrate, there is no need for special organic substrates, therefore the assembly process is closer to traditional flip-chip packaging. AMD believes that EFB is both cheaper and less complex than Intel's EMIB technology, providing also a more scalable solution based on a lithographic process rather than substrate based, considering current IC substrate shortages in the industry.

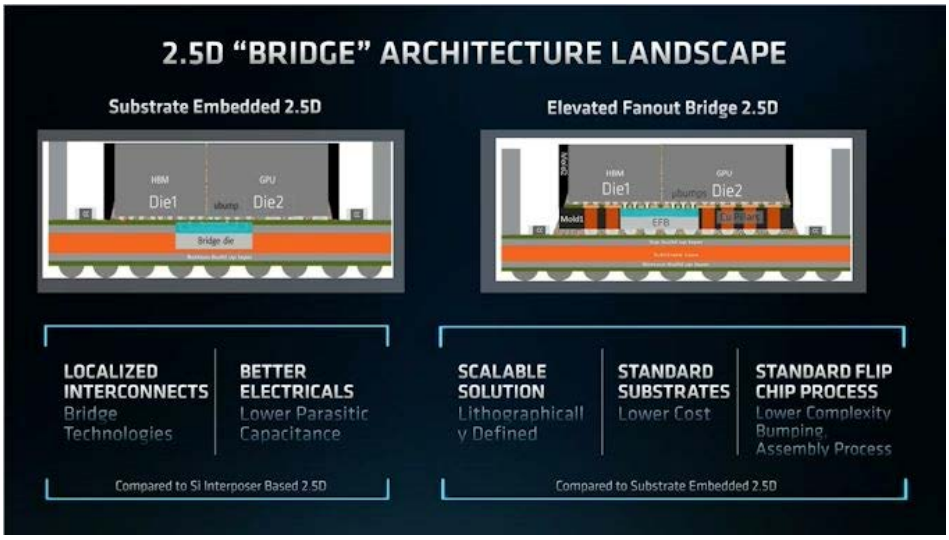

Figure 48. AMD Elevated Fan-out Bridge Technology. [37]

In all the bridge-embedded (in-substrate or fan-out) technologies brought to the market, the bridge has the same functionality: to provide the necessary local interconnections required to connect the dies through a much smaller and high-density silicon bridge, rather than a massive and complex silicon interposer.

For high-density applications, while the more popular solution continues to be the silicon interposers (with submicron line and space capability), alternative solutions are being brought to the market by various industry players, in the form of hybrid technologies. For those applications where the fan-out hybrid solution is useable, the electrical and thermal performance have been shown to be better than a silicon interposer, because a fan-out usually offers better dielectric properties compared to that of an interposer, which is composed of silicon oxynitride (SiOxNy) material. The fan-out solution is also usually thinner than the interposer, which has typically been in the 75 micron

range. The 2.5D interposer and the new, hybrid fan-out solutions will co-exist, at least in the near future, with each filling its particular niche markets. Table 2 shows the industry fan-out solutions targeting 5um L/S and below:

Table 2. Industry Activities for Ultra-High Density Fan-out, ≤ 5um L/S [40]

Company (Package)	*Process*	*Line/Space (µm)*	*Status (as of 2021)*
TSMC (InFO)	Face-up	2/2	Production
ASE (FOCoS)	Face-down	2/2	Production
SPIL (FOEB)	Face-down	2/2	Qualification
JCET (XDFOI)	Face-down	2/2	Qualification
Amkor (SWIFT)	Face-down	5/5	Qualification
TFME (FOPoS)	Face-down	5/5	Development

Besides the technologies described in this chapter, additional development activities are on-going in the industry to develop organic interposers and fanning out interconnects from the substrate level with capabilities of 10/10µm L/S and below, down to 2/2 µm L/S. Such developments are using similar redistribution lines on substrates rather than coming from the die, and are being developed as alternative, lower cost technologies to the 2.5D Interposer. They can be manufactured by substrate makers and integrated in the final package by OSATs, foundries or IDMs. There are various technologies (2.1, 2.3 and others); therefore they will be categorized for now under 2.xD and described in further detail in the next edition of this chapter.

As described, a significant amount of activity went into bringing fan-out technologies to the market. While first developments started around 1983, production did not happen until 27 years later when Infineon brought the first products to the market using their eWLB technology. It was a long journey, and, even though today we have a wide range of technologies available, it still feels that we are just at the beginning of what this journey can be. Currently there are a wide range of fan-out technologies with market adoption across various applications and markets, from mobile to consumer, computing, industrial, medical and automotive. Fan-out is probably one of the most versatile packaging platforms brought to the market, a platform suitable for single and multiple die, 2D to 3D heterogeneous integration, and with solutions for low- to mid- and high-end applications. Multiple suppliers are either in production already or are in various stages of development. Currently the state-of-the-art line and space (L/S) capability of the fan-out versions is 1 micron or greater. Figure 49 shows an overview of industry activities and evolution of fan-out based technologies.

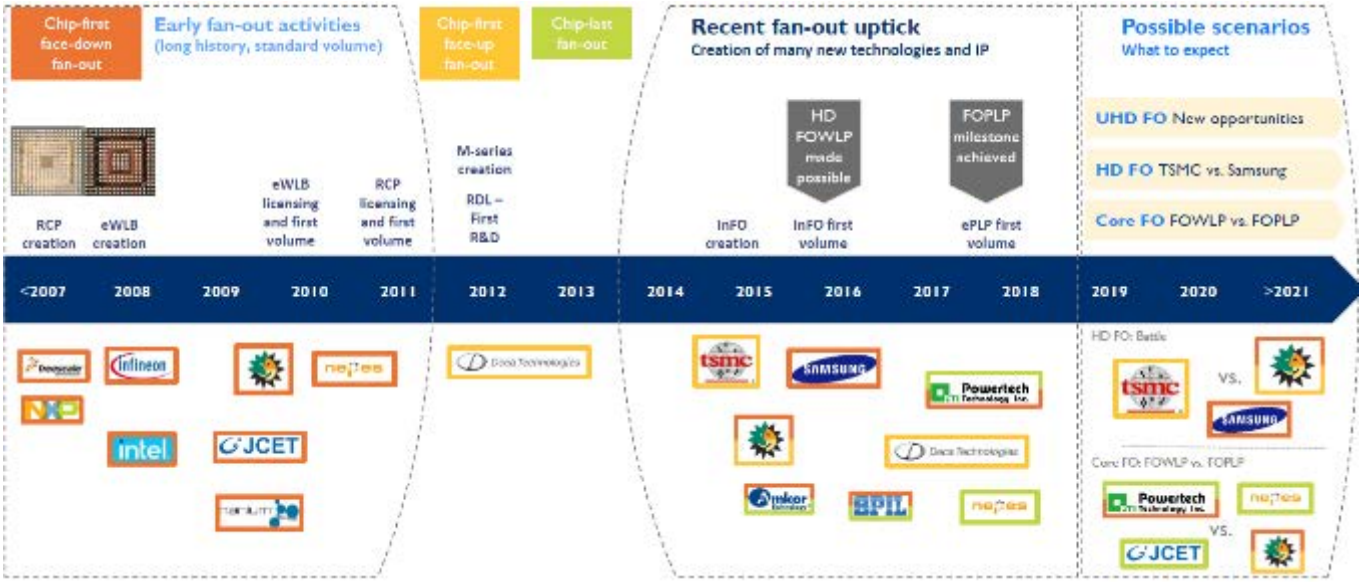

Figure 49. Fan-out Evolution, Key Industry Players and Technologies. [13]

4. Technical Challenges

The various packaging technologies included in the category "WLP" face a number of challenges in the industry, some generic and some specific to the individual WLP processes and structures. The challenges vary from logistics;

marketing; technical in the form of processing, materials and equipment; and implementation into the semiconductor ecosystem. The technical challenges will be addressed in separate sections later in this chapter.

For WLCSP, there are several challenges in the near term. As the silicon technology nodes advance, there are more challenges for reliability and chip-package interaction (CPI) as the size of the WLCSPs increases. It is not only the reliability performance; it is the adverse effects that can occur during subsequent processes after WLCSP manufacture. This includes shipping and handling, and final assembly onto circuit boards. There has been an increased interest in adding five- or six-sided protection in the form of mold-type compound around the WLCSP to provide additional protection for processes after manufacture.

Another challenge for WLCSPs, as the technology nodes advance, is singulation. The most common methods of die singulation for WLCSPs are mechanical saw and laser skiving. However, the mechanical saw process creates some level of mechanical damage in the form of small cracks in the sidewall that can propagate further into the die structure and cause device failures. The challenge is to develop a cost-effective, high-volume method of die singulation that causes minimal damage during processing. Even with fan-out packaging, any damage induced by singulation is still present, but hidden in the mold compound. The mold compound can provide support, and may help to prevent further crack propagation, but it is still present. Singulation, therefore, is important for all forms of WLP structures.

When it comes to fan-out, there have been three basic structures/processes for what is commonly called "Fan Out": chip-first die face-down, chip-first die face-up, and chip-last. These basic structures have been expanded to include many variations, some major and some minor. With all the variations that are now available, it is becoming more and more challenging for the end user to understand the differences between them, as well as the advantages and disadvantages of each. Each supplier of fan-out has its own set of structures, with potentially different material sets and process flows, and nomenclature to differentiate themselves. This makes it a significant challenge for the end user to not only choose a package structure, but also to be able to have a second source for any defined structure. This can have a negative impact on high-volume implementation.

Historically, there have been several primary challenges in the manufacture of chip-first fan-out packages. These have been:

1. Die-shift during reconstitution molding
2. Low temperature polymer for RDL isolation
3. Reconstituted wafer warpage during processing

There have been other challenges specific to individual fan-out processing flows, but these have been challenges for all of these various flows. Each challenge has been met with processing materials, equipment, or modified processes to mitigate the risks imposed by them.

In addition to these historical challenges, the increasing complexity and variations of fan-out structures and processes has introduced, and will continue to introduce, new and increasingly complex challenges.

4.1 Challenge of Die Shift

For chip-first fan-out, the dies are typically bonded, either face down or face up, onto a temporary carrier using some form of temporary adhesive system. The carrier is shaped similar to but slightly larger than the resulting molded reconstituted wafer or panel. During the bonding process, the elevated bonding temperatures result in mismatches of thermal coefficients of expansion (TCE) of the various components involved, as well as shrinkage of the mold compound itself. The thermal coefficients of expansion include the die themselves, the adhesive system, the carrier, and the mold compound. The end result of both these factors is that the die do not end up in the same relative positions that they were placed in by the pick-and-place system prior to molding. This excludes any issues of the die actually physically moving as a result of poor bonding to the carrier, commonly described as "flying die".

The normal corrective action is to mold dummy die of similar physical parameters as the actual die, placing them in the relative positions that are desired in the final product. After molding, the actual shifted positions of all the die are measured. The database for die placement is then modified to offset the die placement by the inverse of the measured shift, with the intent that the die will end up in the positions desired. In reality, there is some degree of predictability overlaid with a lesser degree of random shift. Therefore, this method does not eliminate die shift, but merely minimizes it. The other factor that this technique cannot correct for is die rotation, or theta movement. That is completely random, and the larger the die, the more the theta shift can affect further processing. Uncorrected and corrected die shift are illustrated in Figure 41.

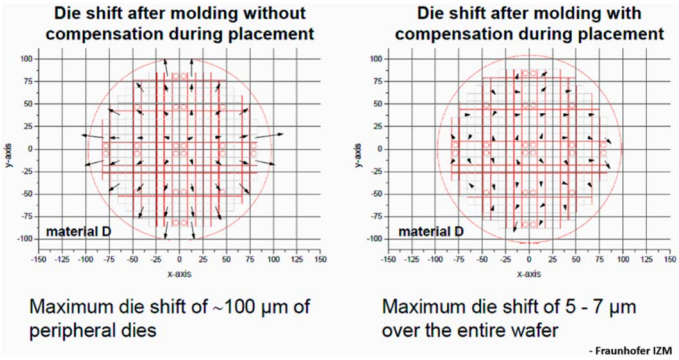

Figure 50. Die Shift without and with Compensation. [41]

The impact of this die shift becomes apparent when the first layer of polymer isolation and RDL is formed on the reconstituted wafer. Excessive die shift results in the first layer via, connecting the die pad to the RDL capture pad, misaligning to the die pad, with an open electrical connection.

To accommodate as much die shift as possible, the normal practice is to use a die placement system that is as accurate as practicable and choose the components of the molding process to minimize actual die shift. The pick-and-place accuracy is a tradeoff, as generally the more accurate the placement, the lower the placement throughput, and hence productivity. Typical desired placement accuracies are currently less than or equal to 5 microns at this time for standard fan-out processes.

There are other techniques currently in use to accommodate larger die shifts. These include direct imaging to compensate individually or in selected regions for die shift, and the use of additional copper pads that are enlarged to accommodate larger die shifts, while still allowing via connectivity.

A compensation technique that is currently in production was developed by Deca Technologies, and goes by the tradename "Adaptive Patterning™" [42]. With this process flow, a high speed, relatively lower accuracy placement system is used for enhanced throughput. The die are first bumped with copper pillars for every die pad. The die are then placed on the temporary carrier in a die-pad-up orientation, with lower accuracy, but at a significantly higher placement speed. This is different than the typical eWLB type of placement, which is a die-pad-down orientation. After molding, the reconstituted wafers are ground to expose the copper pillars. The positions of every die in the molded wafer are then measured in X, Y, and Theta. This measurement data is used to recalculate a new RDL pattern which best fits the measured data. The new RDL pattern design is imaged onto the molded wafer using a direct imaging system. This technique has benefits with the migration to panel fan-out, as the die shift gets progressively worse with the greater distance from the panel's center.

The typical die shift experienced in manufacturing is a radial shift, either inward or outward to/from the center of the wafer/panel, depending on the various materials selections. The shift is usually proportional to the distance from the geometric center of the wafer/panel. This means that as the panel size increases, so does the die shift towards the outer perimeter of the panel. This radial shift also incorporates some degree of random shift superimposed onto the radial shift. For large-panel processing using chip-first technology, techniques must be developed to either minimize the shift or to compensate for it in order to achieve high-yielding manufacturable fan-out panels.

4.2 Challenge of Low-Temperature Polymer for RDL Isolation

The mold compound currently used for chip-first fan-out reconstitution typically has a glass transition temperature (Tg) of between 175 and 185 C. Historically, the Polyimides (PI) and Polybenzoxazoles (PBO) which have been used for wafer-level processing in wafer-level chip-scale packaging (WLCSP) have curing temperatures in excess of 300 C. If these polymers were used for RDL formation after fan-out wafer reconstitution, the mold compound would soften excessively, causing deformation of the fan-out wafer.

Early polymers with cure temperatures of less than 250 C existed with the release of fan-out products using eWLB and RCP technologies, but their materials properties were not ideal for the reliability requirements of fan-out packaging. Around 2012, new polymers became available that had improved materials properties and performed significantly better through both package-level and board-level reliability testing. This enabled the expanded use of fan-out to a larger customer base.

New and improved polymers with cure temperatures below 200 C are being developed and should become available from multiple suppliers. These will allow the processing of RDL layers at temperatures at 170 C or below in the future.

For single-sided chip-last fan-out, this is not an issue, as the RDL and RDL polymer layers are formed prior to molding, and therefore the cure temperature of the polymers does not impact processability. However, if the fan-out packages require RDL layers on the second side, this process occurs after molding, and therefore suffers the same limitations for cure temperature as chip-first fan-out.

4.3 Challenge of Reconstituted Wafer Warpage during Processing

When chip-first fan-out is used, whether die-down or die-up, and if RDL and backend processing occurs on an unsupported molded wafer, the typical structure is unbalanced between the upper and lower materials structures. This results in unbalanced TCEs between the top and bottom structures, resulting in warpage of the reconstituted molded wafer. The warpage can be relatively simple concave or convex, or more complex and asymmetrical as shown in Figure 51.

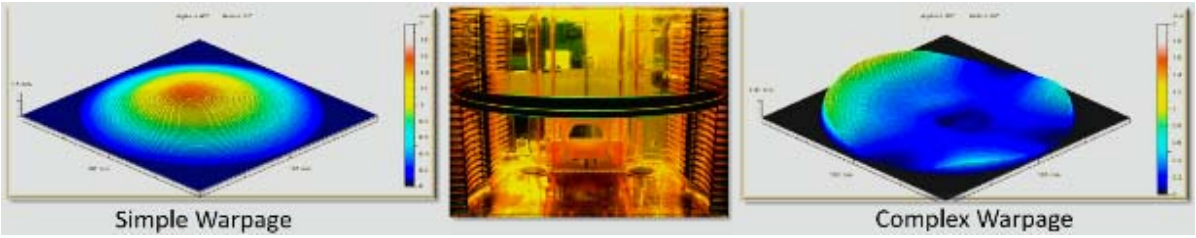

Figure 51. Examples of Molded Wafer Warpage. [43]

The warpage can create several processing issues during fan-out processing. The first is simply that handling systems may have difficulty in reliably picking up and transferring the molded wafers, as the vacuum pickup heads may not have proper seals. Also, the warped wafers may not fit properly into carriers, or into the processing equipment, resulting in chipped or broken molded wafers. And specific processes that rely on a planar processing surface may not give reliable processing results, such as imaging, inspection, laser marking, and test. To make matters worse, as new materials and thermal exposures are encountered at each stage of the process, the wafer warpage changes, often dramatically, as shown in Figure 52, which shows warpage at various process steps in an eWLB process flow.

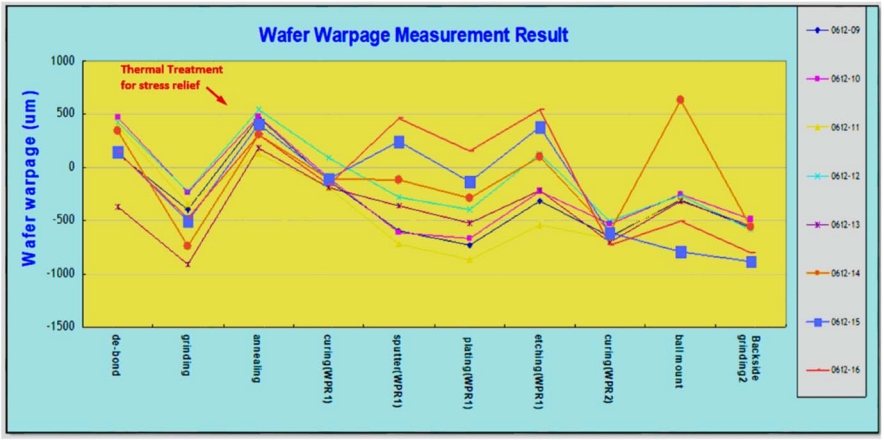

Figure 52. Warpage Measurements for 8 Reconstituted Fan-Out Wafers during processing. (ASE) [43]

Different methods have been used to reduce the amount of warpage, including lowering process temperatures, using lower-stress materials, and better balancing of the top and bottom structures. A more effective solution is to use a rigid temporary carrier during RDL processing, test and as much of the back-end process as possible. This technique is used by necessity for chip-last fan-out, as the RDL structure must be formed onto a temporary carrier through all of the RDL formation process and is also in place during the flip chip of the die onto the RDL structure.

These are all challenges that were encountered early in the development of fan-out packaging. However, as the complexity of fan-out packages increases, newer challenges are being encountered. These include:

- Finer line and space features in RDL formation
- 3D through-mold connectivity
- Incorporation of passive components into the fan-out structure
- Die placement speed and accuracy
- Mold compound development
- Thinner fan out packages

4.4 Challenge of Finer Lines and Spaces in RDL Formation

Early fan-out packages typically had line and space geometries in the 10 to 15 micron range. This has evolved to high volume packages with a 5 to 10 micron range, and selected fan-out package technologies in lower volume production with 2 micron line and space capabilities for the hybrid Fan Out on Substrate packages. These include ASE's FOCoS and TSMC's InFO_OS hybrid packages where a multi-die fan-out with bumps is flip-chip assembled onto a BGA substrate. Examples of fine lines and spaces on these latest high density fan-out products are shown in Figure 53.

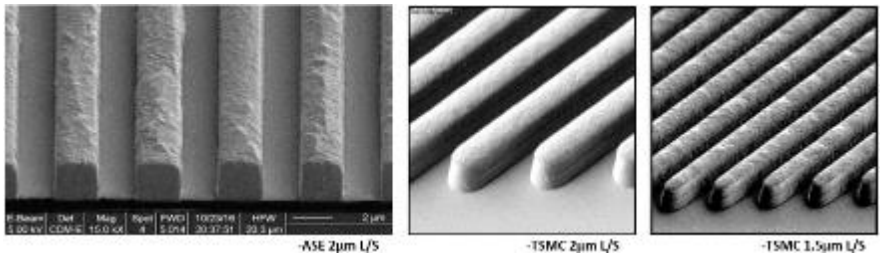

Figure 53. Fine Line and Space Geometries on High-Density Fan-Out [26]

These finer lines and spaces allow fan-out technologies to compete with interposer solutions for very high-density packaging applications. But to be more comparable with these interposer solutions, future fan-out packages will need submicron lines and spaces. Multiple companies are working on various solutions which would allow these densities in production. In order to achieve these finer lines and spaces, techniques and processes other than those currently in use may be necessary. This may include Damascene types of processes and structures in order to achieve quality submicron lines and spaces. [26]

Another challenge as the RDL lines and spaces go below 1µm is line-to-line leakage and long-term electromigration concerns. The Cu redistribution line (RDL) in advanced fan-out (FO) packages is fast approaching a submicron-scale feature size (i.e., below 1µm), for achieving high-density (I/O number > 1000) packaging requirements. This ever-downsizing trend of the Cu RDL gives rise to an increasing current density, which can readily exceed 10^5 A/cm^2 in a, say, 2µm/2µm line/space (L/S) trace structure. Such electromigration reliability concerns caused by the electrical-thermal coupling effects are to become more prominent accordingly and warrant future work. [44]

4.5 Challenges of 3D Through-Mold Connectivity

Initial fan-out manufacturing capabilities, such as the eWLB, were limited to 2D structures, with all connections and components on a single side of the fan-out assembly. With the expanding use of fan-out for System in Package (SiP) and Package on Package (PoP) applications, there is a need for through-mold connections to allow double sided connectivity. A wide variety of techniques have been developed for connecting the front and back side of the fan-out package, examples of which can be seen in Figure 54.

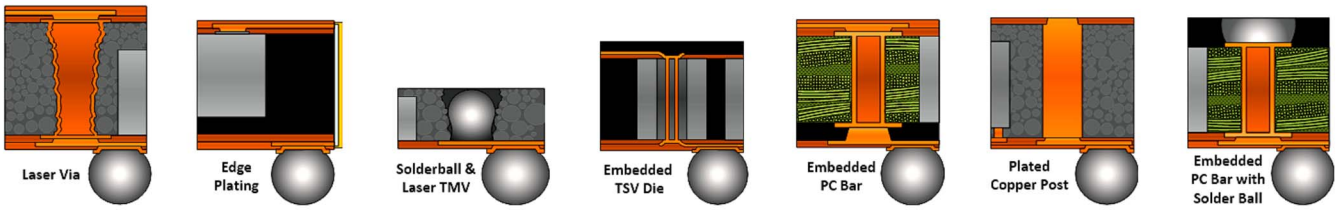

Figure 54. Examples of Through-Mold Connectivity for Fan-Out. [43]

The one structure that is in high volume production with TSMC's InFO fan-out is the Plated Copper Post, Figure 55, including examples from TSMC's InFO and ASE's FOPoP structures.

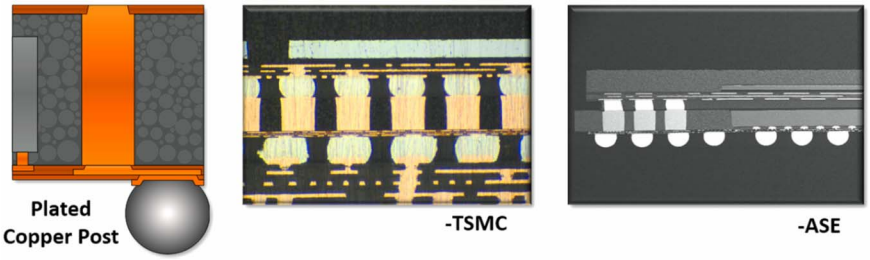

Figure 55. Through-Mold Connectivity Structure in Volume Production. [43]

This through-mold configuration can be implemented with either a chip-first die-up fan-out process, or a chip-last fan-out structure, but because of the process flow, it is not practical for chip-first die-down such as eWLB. For an eWLB process flow, all of the other structures in Figure 45 are able to be used; however, they each have their pros and cons.

The challenge going forward is to reduce the cost of implementing through-mold via connectivity and increase the density of the through vias. Is there a more cost-effective technique for electrically connecting the bottom and top sides of a fan-out package, or a method that is simpler from a manufacturing perspective?

4.6 Challenges of Incorporation of Passive Components into Fan-Out Structures

In addition to including one or more active die within a fan-out package structure, if we are going to use fan-out for heterogeneous integration, we want to also include passive component structures within the package. There are currently three basic techniques for including passive components with a fan-out package structure. The most obvious would be to physically mount a discrete passive component within the molded structure. Currently the two fan-out structures that lend themselves to this technique are chip-first die-down and chip-last processes. Examples of these two structures are shown in Figure 56.

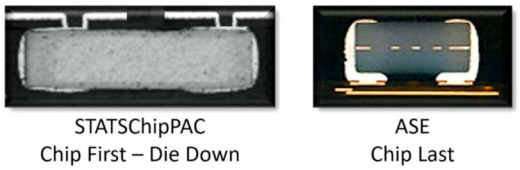

Figure 56. Discrete Passive Components Embedded in Fan-Out Packages. [43]

Technical issues currently limit discrete components embedded into chip-first die-down fan-out structures to those with copper terminations, not solder or tin. For chip-last fan-out, any component that can be surface-mount assembled can be included in the fan-out package, limited only by device thickness.

A second technique that currently can be used is the inclusion of an integrated passive device or devices within a die-like structure, which can be assembled into a fan-out in the same manner as the active silicon die. Examples of these types of integrated passive devices can be seen in Figure 57.

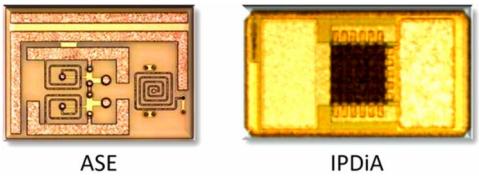

Figure 57. Integrated Passive Die. [43]

The third method of integrating passives into the fan-out package is to incorporate them into the copper RDL structure itself, usually in the form of planar inductors, as shown in Figure 58. These can be extended over the mold compound with favorable Q values. The technique can be extended to include antenna structures in specific applications.

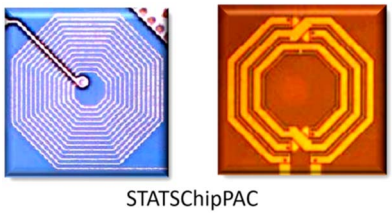

Figure 58. Integrated Passive Die [43]

Each of these current methods has its advantages and disadvantages for the various fan-out structures and processing flows. But often the disadvantages create limitations on implementation in high-volume production at acceptable cost and process complexity levels. The challenge is to develop structures and methods that will allow for more complex heterogeneous integration of multiple forms of passives into each of the fan-out technologies.

4.7 Challenges of Die Placement Speed and Accuracy

As discussed in section 4.1 on die shift, with all current chip-first processing technologies, without some form of post mold compensation technique, very accurate die placement equipment is typically used during the reconstitution process. The original eWLB process used flip-chip bonders with placement accuracies of around 10µm; today's

systems have accuracies of 5 µm or better. But, at these levels of placement accuracy, the die placement throughput is typically less than 10 thousand die per hour. This number varies with die size. The alternative is to use some form of die shift compensation after molding and post-mold cure of the reconstituted wafer. This may allow the use of less accurate "Chip Shooter" type of surface mount placement systems with considerably higher throughputs.

The majority of production fan-out today is manufactured using a 300mm round wafer format. The industry is working toward the implementation of larger processing formats using rectangular panels instead of round. The larger format presents several challenges for die placement. As stated earlier, a large panel can hold as much as five times the number of die as a 300mm round wafer format. With this increase in die counts, it is desirable to have a placement system that is significantly faster than today's systems. And accuracy cannot be sacrificed, as die shift typically increases with distance from the geometric center of the molded wafer/panel. This means that we need systems as least as accurate as today's systems, if not more so. The challenge then is for faster, highly accurate die placement systems that can place a wide variety of die sizes, and also passive components, as necessary for tomorrow's System in Package (SiP) applications.

4.8 Challenges of Mold Compound Development

Early volume manufacturing of fan-out packaging used liquid mold compounds to encapsulate the die in chip-first structures. [45-46] The die were encapsulated in these liquid compounds using either compression molding [45] or using dam and fill [46]. The liquids were initially simpler to use, and are still in use today; however, they have been supplemented by lower-priced granular mold compounds, also in production. As companies worked to develop panel fan-out, mold compounds in sheet form have also been developed. Future challenges are to produce better mold compounds with less shrinkage, improved adhesion, lower warpage after molding, smaller filler material, easier distribution over the larger panel sizes, and lower costs. As higher-frequency packaging becomes more common with 5G mm-wave applications and automotive radar, mold compounds need to be developed with lower dielectric constants and dissipation factors. Also, thermally conductive mold compounds are needed for high-power applications.

4.9 Challenges of Thinner Fan-Out Packages

As fan-out packaging becomes more prevalent in the Mobile market of cell phones and wearables, it is imperative that the overall package thickness is reduced below that achievable in today's fan-out packaging. Even in the advancing automotive market, as the sheer quantity of electronics in each automobile increases, it is becoming more important to reduce each package in size and thickness. Several factors have been limiting the thinness of current fan-out packages, including warpage and handling. Today's fan-out can be manufactured with an unsupported molded wafer/panel, or using a temporary carrier. The carrier approach adds cost for products designed for the low-end market, so much of this market is utilizing unsupported reconstituted wafers through the process steps. However, this technique can suffer from excessive warpage during manufacturing. The challenge is to make the required improvements in materials, processes, and structures in order to achieve lower packaging thicknesses.

Current fan-out production is predominantly in 300mm wafer format, but as we move to the larger panel sizes, the large number of die per panel can limit who benefits from the lower cost structure inherent in panel processing. The difference in die per panel can easily be as large as 5:1 between rectangular panels and round wafers. If all die on a panel are uniform, then the main beneficiaries of panel fan-out technology are those customers with high enough volume requirements to fully support running their product on these larger panels. A challenge, then, is how to utilize larger panels to benefit even the smaller customer. One way would be to partition the panel and combine different dies on the same panel. This would require all of the processes to accommodate different die sizes within one panel. The more problematic processes would be those such as imaging and plating, with additional problems created at test and backend processing.

As the complexity of the products which are processed using WLP technology increases, more challenges will emerge, and creative engineering will be required to address them as they appear.

5. Supply Chain Activities and Considerations

WLCSP has seen strong market adoption in mobile, and with the growth of IoT, consumer and automotive applications, the industry will continue to adopt this packaging platform. OEMs will continue to drive the adoption, given that many of the packages are found on their smartphones, tablets, and now more and more adopted in airpods, smartwatches and other wearables. In terms of manufacturing, the WLSCP supply chain is already a well-established market, as shown in Figure 59, with OSAT, IDMs and foundries active in this space.

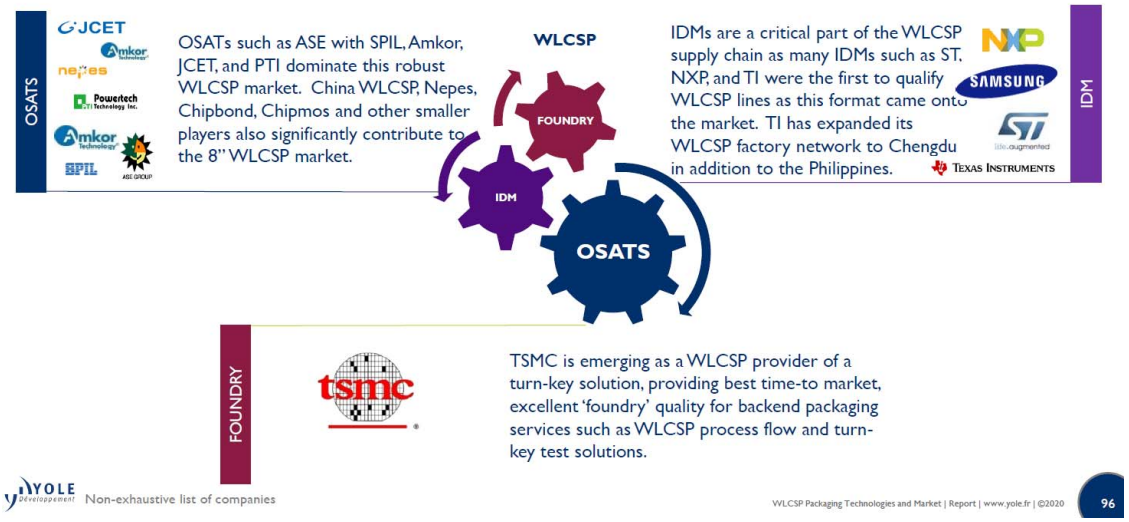

Figure 59. WLCSP Supply Chain and Players. [10]

WLCSP is already a mature packaging platform manufactured mostly on wafer infrastructure, using both 200 and 300mm wafers. Adoption of such packages for new applications will require future development and considerations for the specific requirements of the applications. Main materials and processes continue to be the Tin-Silver-Copper (SAC) alloy as a ball drop, redistribution lines of 5-10um in thickness using polyimide (PI) or poly benzooxazole (PBO) materials. Adoption of TSVs for 3D WLCSP use typically via-last integration with smaller aspect ratios and larger via sizes that are easier to manufacture with no significant challenges. New emerging trends for WLCSP are low temperature dielectrics, increasingly thinner packages and six-sided (6S) side mold protection of packages for increased reliability. [10,47]

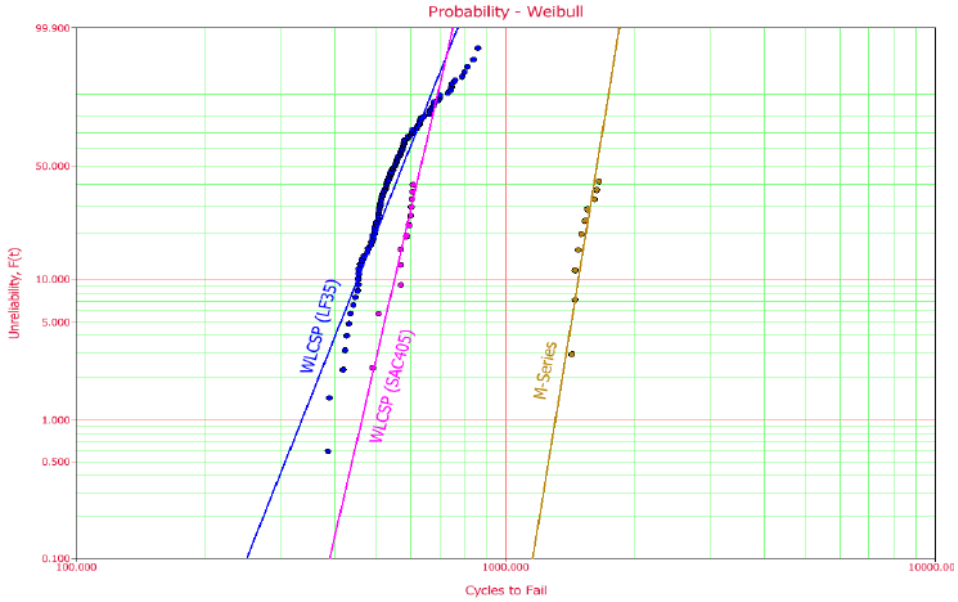

Figure 60. Board Level Reliability Comparison between M-Series and WLCSP. [10, 42]

Figure 60 shows superior performance in thermal cycling for the packages using M-Series protection compared with traditional WLCSP-based platforms. However, this side wall protection comes with added cost compared to the simpler WLCSP process flow. [10]

Fan-out Wafer Level Packaging, and also more recently panel packaging, has attracted many industry players, from OSATs, foundries, and IDMs as well as OEMs. Figure 61 from Yole Developpement shows the map of the main players, both manufacturing as well as end-users.

Figure 61. Fan-out Supply Chain: Manufacturers and End-Users. [13]

Fan-out Wafer Level Packaging (FOWLP) has historically utilized multiple geometric configurations for processing, from round wafer formats to small and large rectangular panels. However, the predominant configuration for high-volume manufacturing since 2009 has been either 200mm or 300mm round wafers. These geometries have allowed fan-out suppliers to make use of the large manufacturing and equipment infrastructure that exists for wafer processing. As the reconstituted fan-out wafers do not have the flatness and rigidity of silicon wafers, some modifications have had to be made to the equipment and processes to accommodate the limitations of the polymer matrix used in the reconstitution process. These limitations include the low Tg of the matrix material, die shift during molding, and the tendency for warpage during processing.

In spite of these limitations, fan-out packages are being produced in very high volumes by multiple suppliers in wafer format. However, as die and the resulting packages are rectangular in shape, round wafers do not provide the most effective areal density for maximum processing efficiency. Currently, 300mm wafers are the largest wafer format in volume production. For manufacturing cost considerations, the larger the panel that can be processed, the more cost-effective the pricing can be per device. For this reason, development is ongoing for transitioning fan-out production from round wafers to larger rectangular panels.

There have been no standards developed for rectangular panel production, with different manufacturers choosing various sizes of panels based on their proposed manufacturing methods and technologies. Since many of the processing technologies that work for round wafers are not suitable for larger rectangular panels (e.g. spin-on coatings), suppliers are required to develop alternative materials, processes and equipment.

Fan-out volumes have been slow to ramp up, but as new applications and manufacturing capabilities are evolving, the volumes are beginning to increase. The somewhat limited historical volumes have not supported multiple large-capacity panel production lines, but these will become cost effective as more customers begin using fan-out.

Within recent years, multiple suppliers have begun offering fan-out packaging solutions manufactured using larger rectangular panel formats. Each supplier has developed its solution optimized for a specific structure, processing flow, and materials set. Some have adapted technologies from the LCD display panel industry, some from the circuit board manufacturing industry, and others are modifying wafer processing technologies to enable rectangular panel processing. This is significantly different than the situation for wafer processing of fan-out, where there was more standardization of structure and materials sets. This standardization made it relatively easy for the end customers to source product from multiple suppliers, while expecting products to be equivalent in form and function.

Instead of following the wafer-level roadmaps to 450mm, panel-level packaging might be the next big step. First companies, including SAMSUNG SEMCO, Nepes, Powertech and Deca with ASE, have already announced that they are preparing for panel-level packaging in volume manufacturing. Sizes considered for the panel range from 300mm x 300mm to 457mm x 610mm or 510mm x 515mm up to 650mm x 650mm. The main challenge at the moment is the missing standardization on panel formats. SEMI has started an initiative on the standardization topic.

With the wider variety of structures and materials sets evidenced by the various developers of larger format panel fan-out, this may become more challenging for the end customers. Table 3 summarizes a recent roster of leading FOPLP suppliers.

Table 3: Production Status of Fan-out Panel Level Packaging (FOPLP). [48]

FOPLP Supplier	*Chip First/Last*	*Process*	*Panel Size (mm)*	*Line/Space (µm)*	*Status (as of 2021)*
Amkor	Chip-Last	Face-down	650 x 650	5/5	Qualified for Production
Eswin	Chip-First	Face-up & Face-down	510 x 515	8/8	In Qualification
Nepes	Chip-First	Face-up	600 x 600	15/15, 10/10	M-Series in Production
PTI	Chip-First	Face-down	510 x 515	15/15	In Production
Samsung	Chip-First	Face-down	405 x 510	7/8	In Production
Unimicron	Chip-First	Face-down	510 x 515	35/35	Qualified for Production

An easy upscaling of technology when moving from wafer to panel level is not possible. Materials, equipment and processes have to be further developed or at least adapted. A view along the process chain offers lots of possibilities but also challenges. Carrier material selection for a chip-first approach should be considered, addressing not only the thermo-mechanical behavior but also properties such as weight and stability. Pick-and-place assembly on carrier is independent from wafer or panel formats and might be a bottleneck. Here new equipment or even new approaches for high-speed but also high-accuracy assembly are required. Compression molding is typically used for chip embedding and to form the reconfigured wafer or panel. Liquid, granular and sheet-type molding compounds are available. All allow chip embedding with pros and cons in cost, processability and cleanroom compatibility. For RDL formation, a large variety of lithography tools and dielectric materials options exist. As dielectrics, photosensitive as well as non-photosensitive or liquid versus dry-film materials can be considered. Mask-based lithography such as stepper technology, and maskless-based tools such as laser direct imaging (LDI), are available for panel sizes. Both offer differing capabilities and strategies to overcome challenges due to die placement accuracy and die shift after molding. Finally, solutions for grinding, balling and singulation are needed.

Equipment used in the processing of WLCSP and fan-out have quite a few similarities, with key differences being capability of handling warped wafers, and low-temperature processing. There are also additional process steps used by the different fan-out variations. These include wafer or panel molding, flip chip assembly for chip last, and surface mount of passive components. Even the conventional WLCSP processing equipment often requires special considerations to properly process fan-out packages. Here are a few examples regarding the plasma vapor deposition (PVD) process, one of the most common equipment choices used in both WLCSP and fan-out [49]:

- Fan-out has some more-stringent capability requirements than systems used solely for WLCSP.
- In the case of fan-out, the mold layer is low-temperature tolerant, so needs to be treated carefully in the PVD system. It also shrinks as it cures and distorts the wafer. Warpage of 6mm is not unusual for FO wafers, an order of 20x more distorted than wafers in the front-end of the fab.
- In mobile applications particularly, contact resistance affects battery life. Packaging companies want low contact resistance, while at the same time that they are introducing new materials onto the substrate. These materials have the potential to contaminate the metal-to-metal contact, so PVD vendors need to develop new solutions to keep bond surfaces clean in the presence of gaseous contaminants.
- The most common application for PVD in packaging is as a seed layer for Cu or solder plating. After plating through resist, the bulk of the seed layer needs to be removed by wet etch. As lines get narrower, the anisotropic wet chemistry can undercut the plated lines and cause weaknesses. If the PVD process could deliver improved step coverage, the seed layer would be thinner and be quicker to remove. Techniques for improving PVD step coverage are well understood from the front-end, but not always transferrable to the new packaging materials.

Handling, especially automated handling of molded large panels including storage and transport, is still an open topic since until now only custom-made solutions exist. The next update and edition of this chapter will expand the supply chain analysis to include additional materials and equipment used for WLP that have not been covered in this chapter.

There are many process flow options with regard to different applications. But still to be answered is the question of "Where is the sweet spot?", taking performance, yield, cost and panel size into account. Table 4 shows an overview of the different packages and their characteristics and forecasted roadmap.

Table 4

Year of Production	*2019*	*2020*	*2021*	*2022*	*2023*	*2024*	*2027*	*2030*	*2033*
Package size (mm2) Including fan-out and multi-die packages (Min/Max)									
a. WLCSP- Memory	20/250	20/250	20/250	20/250	20/250	20/250	20/250	20/250	20/250
b. WLCSP-Standard Logic and Analog/Linear	0.49/42	0.49/49	0.42/64	0.42/64	0.42/64	0.42/64	0.42/64	0.42/64	0.42/64
c. WLCSP- Wireless: Bluetooth, FM, GPS, WIFI	0.49/42	0.49/49	0.42/64	0.42/64	0.42/64	0.42/64	0.42/64	0.42/64	0.42/64
d. Fan out	1.5/256	1.2/324	1/400	1/484	1/576	1/576	1/576	1/576	1/576
Hybrid Fan Out [Fan Out on Substrate - Max (mm²)]									
a. Fan Out "Die"	324	432	858	1156	1225	1225	1225	1225	1225
b. Package	1800	2700	3600	4500	4900	4900	4900	4900	4900
Number of RDL Layers per side									
a. WLCSP	3	3	3	3	3	3	3	3	3
b. Fan Out	4	5	6	6	6	6	6	6	6
UBM Thickness (µm)									
a. Standard Logic and Analog/Linear (low power)	1.5–20 µm	1.5–25 µm	1.5–30 µm	1.5–30 µm	1.5–30 µm	1.5–30 µm	1.5–30 µm	1.5–30 µm	1.5–30 µm
b. Standard Logic and Analog/Linear (high power)	1.5–20 µm	1.5–25 µm	1.5–30 µm	1.5–30 µm	1.5–30 µm	1.5–30 µm	1.5–30 µm	1.5–30 µm	1.5–30 µm
UBM Metallurgy (see footnote)									
RDL Conductor Thickness									
a. Standard Logic and Analog/Linear (low power)	2–15 µm	2–20 µm	2–20 µm	2–20 µm	2–20 µm	2–20 µm	2–20 µm	2–20 µm	2–20 µm
b. Standard Logic and Analog/Linear (high power)	2–15 µm	2–20 µm	2–20 µm	2–20 µm	2–20 µm	2–20 µm	2–20 µm	2–20 µm	2–20 µm
RDL Metallurgy (see footnote)									
Wafer Saw Street Width minimum - WLCSP (µm)									
a. All saw based singulation for WLCSP	40 µm	35 µm	35 µm	35 µm	33 µm	33 µm	31 µm	31 µm	31 µm
b. Advanced singulation (non saw techniques)	15 µm	15 µm	15 µm	15 µm	12 µm	12 µm	11 µm	11 µm	11 µm
Package Pincount Maximµm									
a. WLCSP	289	289	289	300	300	300	300	300	300
b. Fan Out WLP	800	1000	2000	2000	2000	2000	2000	2000	2000
Package Ball Pitch Minimum (Note 6)									
a. All WLP	200 µm	200 µm	200 µm	200 µm	175 µm	175 µm	150 µm	150 µm	150 µm
Package Preformed Solderball Max Diameter for Min Ball Pitch (Note 6)									
All categories	75 µm	75 µm	75 µm	75 µm	65 µm	65 µm	55 µm	55 µm	55 µm
Package Minimum Backgrind Thickness (Note 6)									
a. WLCSP	90 µm	90 µm	80 µm	80 µm	80 µm	80 µm	75 µm	75 µm	75 µm
b. Fan Out WLP	180 µm	180 µm	175 µm	175 µm	160 µm	160 µm	150 µm	150 µm	150 µm
Type of WLP structure and metallurgy (bump, ball, column, solder, Cu, other) (see footnote)									
Multiple Die Fan Out (Max. dies)									
a. Fan Out (2D Side by Side Die, each package)	7	7	7	7	7	7	7	7	7
b. Fan Out (2D Side by Side Discrete Components, each package)	8	12	16	16	16	16	16	16	16
c. Fan Out (3D Total Die, each package)	3	4	5	6	6	6	6	6	6
d. Hybrid Fan Out (Fan Out on substrate - Total Die, each package)	5	12	16	20	20	20	20	20	20
Fan Out WLP Technology	Note 8	Note 8	Note 8	Note 8	Note 8	Note 8	Note 8	Note 8	Note 8

Notes

1. Entries defining the metallurgy that do not show changes over the next 15 years have been removed from the table. Any changes that might occur will be a result of the development of new materials.

2. the definition of WL-CSP is limited to 1.2 times die dimension or 1.4 times die area. Otherwise the fan-out product would be just fan-out WLP vs. CSP

3. Ball Mettallurgy is projected to be SAC for the next 15 years

4. UBM Metallurgy will have a number of variations depending on the company and the specific application. The metallurgy is not projected to change over the 15 years of the Roadmap.

5. RLD Metallurgy will have the same metallurgy for all device types and it is not forecast to change over the 15 years of the Roadmap

6. Type of WLP structure and metallurgy (bump, ball, column, solder, Cu, other). This metallurgy is not projected to change over the 15 years of the Roadmap. The metallurgy will be 2ML/2P/ Plated Cu/Solder Bump/Ball/ Copper Pillar where ML= metal layer and P=polymer.

7. These parameters are driven by PCB manufacturing and cost issues and do not represent a limitation of the technology.

8. Single Die, 2D Multi Die, Passives, Through Vias, Doubled-sided, 3D Modules, Hybrid Fan Out on Substrate

6. Summary and Final Conclusions

Higher performance, miniaturization, lower cost and the need to integrate more functionalities within a system are driving the industry for increased advanced packaging adoption across several applications and markets. The importance of advanced packaging and especially heterogeneous integration is becoming greater, considering that scaling and cost reduction will not be able to continue, at the same rate, by only following Moore's law. Advanced nodes do not bring the desired cost-benefit any more, and R&D investments in new lithography solutions and devices

below 10nm nodes are rising substantially. In order to respond to market demands, the industry seeks further performance and functionality boosts in integration. Packages are now requested to bridge this gap and revive the cost/performance curve while at the same time adding more functionality through integration. They become enablers for new designs, new performance levels and new applications.

WLPs with fan-in design historically were applied for lower I/O counts and small die sizes, with low power requirements. Advances in fan-in resulted in packages with I/O counts greater than 400 that passed JEDEC reliability criteria, packages being used at current high enough to allow their usage in Power Management ICs. They are primarily being used in portable consumer markets where small size, thickness, weight, and electrical performance are additional advantages to cost. Major trends include work for cost-efficient rerouting with multi-layer RDL and improved design and simulation tools for WLP technologies.

Fan-in wafer level packages remain an important packaging platform. The mobile market continues to be the main driver for fan-in WLP adoption, with more than 90% of all fan-in packages found in handsets, especially smartphones and tablets, while new markets such as automotive, consumer and industrial are starting to adopt such packages. With the momentum of 5G adoption in communication and consumer applications, increased number of connected devices, growth of the internet of things (IoT), and the need for smaller and better-performing devices in automotive, the adoption of fan-in packages is expected to grow and expand across several markets.

While WLCSPs started high-volume manufacturing around 2000, the first commercially available high-volume fan-out packaging technology, Embedded Wafer Level BGA (eWLB), was brought to production in 2009. The early volume production of fan-out packages were single-die products with relatively low I/O counts, and lines and spaces of 15 microns or larger. During those early years, fan-out was thought to be a packaging technology for relatively small packages and low I/O counts. If fan-out historically were known as stand-alone packages, within the last decade it has evolved towards using more complex integration schemes, with some products even going into production using interposers with Through Silicon Vias (TSVs) for the high-density interconnection of multiple die.

While 2.5D interposers are providing the necessary interconnects for high end applications, they are very expensive and more complex, considering also that through-silicon-via technology integration is needed. This has led to new developments in the industry to reduce the cost and complexity. Various flavors of embedding smaller silicon bridges, either in the substrate or on top of the substrate, in the fan-out layers, were developed and brought to the market. Companies such as Intel with EMIB, TSMC with InFO_LSI, SPIL with FEOL, IBM with DBHI and most recently AMD with EFB have shown such alternative embedded bridge solutions for 2.5D integration.

Fan-out is one of the latest packaging trends in microelectronics with high interest across the entire supply chain. Besides technology developments towards heterogeneous integration (including multiple-die packaging, passive component integration into the package, and redistribution layers or package-on-package approaches), larger substrate formats are also brought into the supply chain of this platform. Before, the substrate market was significantly impacted by fan-out introduction and replacement of FCBGA; today, with new developments especially targeting high-end applications, organic substrates are becoming a very important component of ultra-high-end fan-out packages.

The manufacturing of WLP has also evolved from round wafer forms to rectangular panel formats, with multiple suppliers coming to the market in an attempt to reduce the manufacturing costs.

With the introduction of TSVs, IPDs, chip-last fan-out, MEMs packaging technologies, chiplets, silicon bridges, and embedded technologies, WLP products can be used in a much broader range of applications, with higher I/O counts and greater functional complexity. These packaging technologies open new opportunities for system-in-package and heterogeneous integration.

Acknowledgments

The chair of the WLP technical working group would like to acknowledge the contribution of David Butler from SPTS (UK), Doug Yu from TSMC (TW), Hiroyuki Shida from Shin-Etsu (JP), John Hunt from ASE (US), Michael Liu from JCET (US), and Tanja Braun from Fraunhofer Institute of Technology (GE), for their contribution to this work.

The WLP Technical Working Group is looking for new volunteers. If you have experience in WLP and are interested in participating and working together with a great team of industry experts to shape the industry roadmap for these exciting and growing packaging platforms, please reach out to Rozalia Beica, chair of this technical group, at r.beica@ats.net

References

[1] Beica, R. “Heterogeneous Integration: The New Driver of Innovation and Growth”, IMPACT Taiwan, 2021
[2] Beica, R. “The Growth of Heterogeneous Integration & The Role of Substrate Technologies”, TPCA Shenzhen, China 2021
[3] Lau, J “Heterogeneous Integration” PDC Course IEEE Electronic Components & Technology Conference 2016
[4] Smithsonian The Chip Collection: http://smithsonianchips.si.edu/ice/cd/PKG_BK/CHAPT_12.PDF
[5] Beica, R. “Enabling Information Age through Advanced Packaging Technologies and Electronic Materials”, 24th Pan Pacific Microelectronics Symposium, Hapuna Beach, Hawaii, 2018, pp. 4
[6] Beech, M. “COVID-19 Pushes Up Internet Use 70% And Streaming More Than 12%”, Forbes, 2020
[7] Yu, D. “Innovative Heterogeneous Integration Technologies Initiate a New Semiconductor Era”, Keynote IEEE 70th Electronic Components and Technology Conference (ECTC), 2021
[8] Beica, R. “The Growth of Heterogeneous Integration and The Importance of Electronic Materials”, IMAPS Device Packaging Conference, Fountain Hills, AZ, 2018
[9] Beica, R., Hunt, J. “Heterogeneous Integration Roadmap; WLP TWG”, IEEE 68th Electronic Components and Technology Conference (ECTC), 2018
[10] Vaibhav Trivedi, “WLCSP/Fan-In Packaging Market and Technology Report”, Yole Developpement, 2020
[11] Kumar, S.; Chiotraga, S.; Shoo, F., “Status of the Advanced Packaging Industry”, Yole Developpement, 2021
[12] Ivankovic, A., “Fan-in Packaging Market and Technology Report”, Yole Developpement, 2015
[13] Chitoraga, S.; Shoo, F. “Fan-out Packaging Market and Technology Report”, Yole Developpement, 2021
[14] Hunt, J. “Fan-out Variations: Structures and Processes for Low and High Density”, IMAPS Device Packaging, 2021
[15] Beica, R., Hunt, J. “Hetergoeneous Integration Roadmap; WLP TWG”, HIR Annual Symposium, 2020
[16]Xiao, Z.; Yao, M.; Yu, D.; Zhou, M.; Dai, F.; Fan, J.; Zhang, W. “Development and reliability study of 3D WLCSP for CMOS image sensor using vertical TSVs with 3: 1 aspect ratio”, Microsyst. Technol. 2017, 23, 4879–4889.
[17] Zhou, T et al., “Development of Reliable, High Performance WLCSP for BSI CMOS Image Sensor for Automotive Application”, MDPI, 2020
[18] Charbonnier, J. et al., “Wafer level packaging technology development for CMOS image sensors using Through Silicon Vias”, Proceedings of the 2008 2nd Electronics System-Integration Technology Conference, Greenwich, UK, 1–4 September 2008; pp. 141–148.
[19] Prutec Limited, Patent Application GB 2 130 794A, United Kingdom
[20] Kang, L, Nepes Corporation, “FOWLP Technology as Wafer Level as System in Package (SiP) Solution”, IMAPS 13th International Conference and Exhibition on Device Packaging, 2017
[21] Casio, “Advanced Substrate Technologies for SiP/SoP”, CPMT Seminar, 2006
[22] Michael Liu, JCET Group https://www.jcetglobal.com/en/site/TechInfo/582
[23] Hsiao, K. “Ultra-Small Fan-out Packaging Solutions”, SemiEngineering 2021 https://semiengineering.com/ultra-small-fan-out-packaging-solution/
[24] Tseng, CF et al., TSMC, “InFO (Wafer Level Integrated Fan-Out) Technology”, IEEE 66th Electronic Components and Technology Conference (ECTC), 2016
[25] Chang, Keng Tuan et al., ASE, “Ultra High Density IO Fan-Out Design Optimization with Signal Integrity and Power Integrity”, IEEE 69th Electronic Components and Technology Conference (ECTC), 2019
[26] Yu, C.H. et al., TSMC, “High Performance, High Density RDL for Advanced Packaging”, IEEE 68th Electronic Components and Technology Conference (ECTC), 2018
[27] Y. P. Chiang, S. P. Tai, W. C. Wu, J. Yeh, C. T. Wang and D. C. H. Yu, TSMC, “InFO_oS (Integrated Fan-Out on Substrate) Technology for Advanced Chiplet Integration”, IEEE 71st Electronic Components and Technology Conference (ECTC), 2021
[28] Lee, Jun Kyu et al., Nepes, “Three-Dimensional Integrated Circuit (3D-IC) Package Using Fan-out Technology”, IEEE 69th Electronic Components and Technology Conference (ECTC), 2019
[29] Su, An-Jhih et al., TSMC, “3D-MiM (MUST-in-MUST) Technology for Advanced System Integration”, IEEE 69th Electronic Components and Technology Conference (ECTC), 2019
[30] Garrou, P. “Passing the Advacned Packaging Baton to TSMC and Its 3D MIM”, 2019 https://www.3dincites.com/2019/06/iftle-417-passing-the-advanced-packaging-baton-to-tsmc-and-its-3d-mim/
[31] Mills, M. “TSMC LSI Technology Will Replace Interposer”, 2020, https://itigic.com/tsmc-lsi-technology-will-replace-interposer/
[32] Reiter, H. “The TSMC 2020 Virtual Technology Symposium and OIP Ecosystem Forum”, https://www.3dincites.com/2020/09/the-tsmc-2020-virtual-technology-symposium-and-oip-ecosystem-forum/
[33] Garrou, P. “Intel EMIB Implementation in the Stratix MX”, IFTLE 324, https://electroiq.com/iftle-324-intel-emib-implementation-in-the-stratix-mx/
[34] J. Lin et al., "Scalable Chiplet package using Fan-Out Embedded Bridge" 2020 IEEE 70th Electronic Components and Technology Conference (ECTC), 2020
[35] Garrou, P “SPIL Fan-out Embedded Bridget (FOEB) Technology”, IFTLE 456, 2020 https://www.3dincites.com/2020/07/iftle-456-spil-fan-out-embedded-bridge-foeb-technology/

[36] Garrou, P. “IBM simplifies Si Bridge Technology”, IFTLE 491, 3DinCites, 2021 https://www.3dincites.com/2021/07/iftle-491-ibm-simplifies-si-bridge-technology/
[37] Smith, R. “AMD Announces Instinct MI200 Accelerator Family: Taking Servers to Exascale and Beyond”, 2021, https://www.anandtech.com/show/17054/amd-announces-instinct-mi200-accelerator-family-cdna2-exacale-servers/2
[38] Swaminathan R., “Advanced Packaging: Enabling Moore’s Law’s Next Frontier through Heterogeneous Integration (HI)”, IEEE 3D IC Conference, 2021
[39] Garrou, P. “Milan-X: TSV, Hybrid Bonding & the Elevated Fanout Bridge”, IFTLE 507, 3DinCites, 2021 https://www.3dincites.com/2021/12/iftle-507-amd-milan-x-tsv-hybrid-bonding-the-elevated-fanout-bridge/
[41] Braun, T. et al., Fraunhofer IZM, “Through Mold Vias for Stacking of Mold Embedded Packages”, IEEE 61st Electronic Components and Technology Conference (ECTC), 2011
[42] Scanlan, C et al., Deca Technologies, Inc, “Adaptive Patterning for Panelized Packaging”, IWLPC, 2012
[43] Hunt, J. ASE contribution
[44] Liang, C. L. et al., National Cheng Kung University, “Electromigration Failure Study of a Fine-Pitch 2?m/2?m L/S Cu Redistribution Line Embedded in Polyimide for Advanced High-Density Fan-Out Packaging”, IEEE 70th Electronic Components and Technology Conference (ECTC), 2020
[45] Brunnbauer, M. et al., Infineon, “An Embedded Device Technology Based on a Molded Reconfigured Wafer”, IEEE 56th Electronic Components and Technology Conference (ECTC), San Diego, 2006
[46] Leung, John, Freescale Semiconductor, “Packaging Technology for Mobile platforms”, Packaging Conference, Shanghai, 2006
[47] SystemPlus Consulting, https://www.systemplus.fr/wlcsp-fan-in-has-been-selected-by-apple-for-its-latest-iphones/
[48] Liu Michael, JCET contribution
[49] David Butler, SPTS contribution

Edited by Paul Wesling