HETEROGENEOUS INTEGRATION ROADMAP

2019 Edition

Chapter 23: Wafer Level Packaging

http://eps.ieee.org/hir

Table of Contents

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

Chapter 23: Wafer-Level Packaging (WLP)

Scope

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 for 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 for WLCSPs and Fan Out technology in various articles and books published on the subjects. This chapter is an attempt to look at the WLP technology as it has developed to date, and project forward to future needs and challenges.

Executive Summary

Wafer level packaging (WLP) came into its own around the year 2000. 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 1.

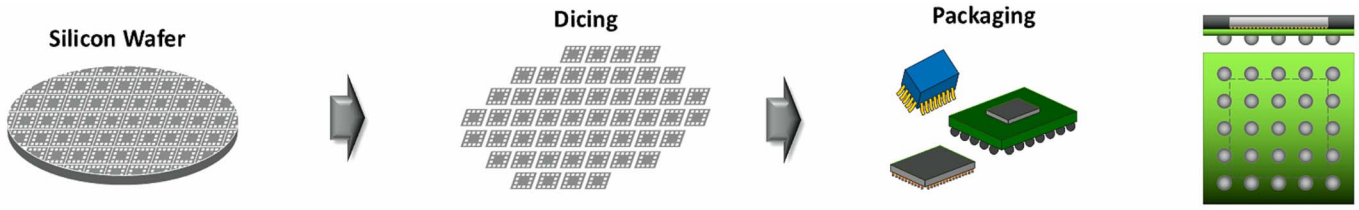


Figure 1. Traditional packaging process flow [17]

WLP 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 solderballs at a coarser pitch than used for traditional bumped die. Unlike previous packaging, nearly all of the packaging process steps are done in parallel while still in wafer form, as opposed to in a series of steps as in figure 1. A simplified illustration is shown in figure 2.

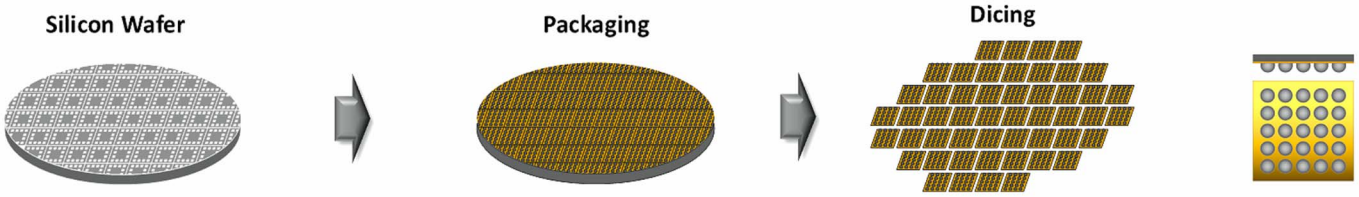


Figure 2. Wafer level packaging process flow [17]

With WLP, 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 solderballs 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 solderballs 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 solderballs. This became one of the first "heterogeneous" WLPs, as shown in figure 3.

Figure 3. WLP with Second Die mounted on Underside [17]

With the development of Through Silicon Via (TSV) technology for 3D applications, what has been called "via last" processes were used to make top-side die connections to the pads normally located on the die underside. 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 4. This became another level of WLP heterogeneous integration complexity.

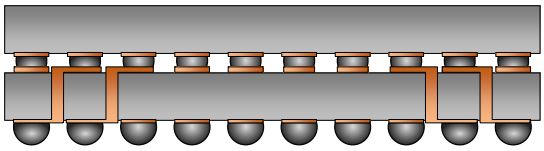


Figure 4. WLP with TSVs for double sided connectivity [17]

However, the packages were still limited in physical area to the actual die size, and as wafer nodes advanced, with geometries shrinking, the die themselves could be reduced in size. This created a dilemma – where to place the solderballs at the coarser pitch required for WLPs. Although what we now call "Fan Out" packaging had been conceptualized and was in various stages of development as early as 1983[1], There were multiple companies developing various versions of what we now call "Fan Out", with the two primary commercial contenders being Motorola/Freescale with their Reconstituted Chip Package (RCP) and Infineon with their Embedded Wafer Level BGA (eWLB). These two resulted in similar package structures, although there were variations in the processing. The RCP also included a Copper frame layer embedded into the mold compound along with the die, acting as both a potential ground plane and stabilizer for TCE mismatch to a circuit board after final user assembly as shown in figure 5.

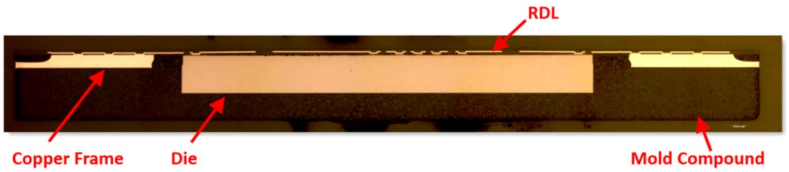


Figure 5. WL fan-out packaging process flow [18]

Infineon took the embedded wafer level BGA (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 molded wafers. The similarity to the WLP process flow can be seen in figure 6. With the additional molded area, the final packages become larger than the die sized WLCSP.

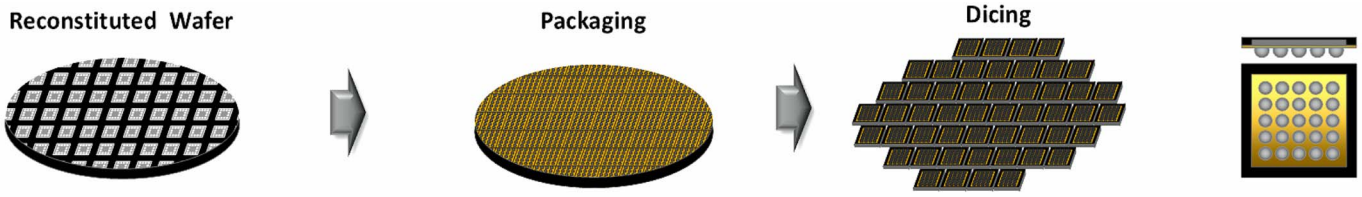


Figure 6. WL fan-out packaging process flow [17]

Although Casio described their EWLP package as a "Fan In/Out Package" in 2006 [2], 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. It was interesting that Infineon chose to use this term. Previously, as in their patent for the eWLB technology, they described all previous packaging in this way: 'Conventional packages or casings for circuit units are therefore constructed using a so-called "fan-out design"'. They were technically correct. In reality, the only package that does not "fan out" the interconnections is a WLCSP, as it is die sized, and can only fan inward.

The first eWLB in volume production was a single-die package, combining baseband with PMIC and RF features. The die were about 5x5mm in 8x8mm fan out packages that varied between 183 and 217 solderballs. An example is shown in figure 7.

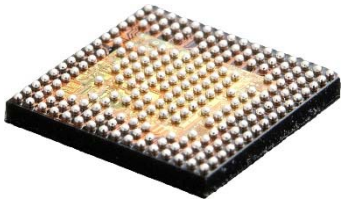


Figure 7. Early Infineon eWLB baseband fan out [17]

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. 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 PoP with more than 1300 solderballs. It used the Apple A10 processor with a standard DRAM memory package assembled on top of it, as shown in Figure 8 [19].

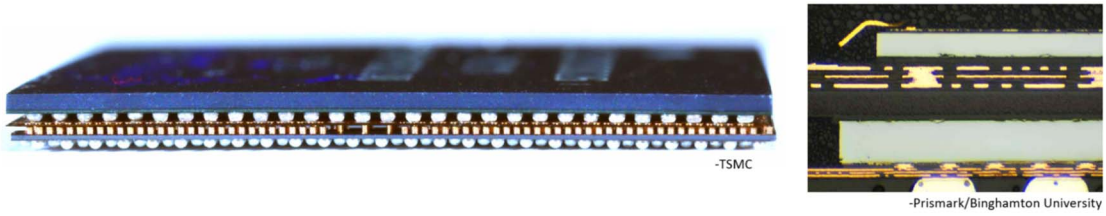


Figure 8. TSMC's Integrated Fan Out package (InFO) [19]

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, figure 9.

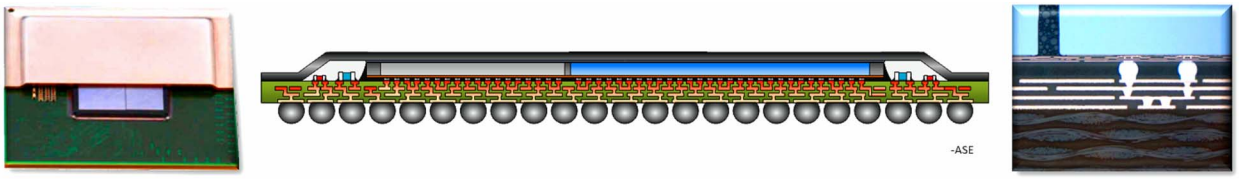


Figure 9. ASE's Fan Out Chip on Substrate (FOCoS) [17]

Both the InFO and the FOCoS were multi-die packages; therefore, both were new forms of heterogeneous integrated packages with high I/O counts and fine trace geometries.

The InFO combined a fan out bottom package with a standard top DRAM mounted on the top side through connections of the InFO fan out, making it a 3D-connected package. The InFO 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 heterogeneous packaging.

The FOCoS product combined two large die into a 32mm x 25mm fan out bumped die, rather than a stand-alone package. This fan-out replaces an interposer solution for the interconnection of the two large high I/O count die, for a less expensive solution. The FOCoS was a specialized product for network and server applications and was produced in lower volumes, but very high RDL densities. Both of these fan-out packages incorporated 3 RDL layers, with much higher complexity than previous production fan-out packages. Since 2016 fan-out has evolved into more complex structures, with multiple process variations. Although the initial volume production fan-out was manufactured in round wafer form, rectangular panel formats have evolved with multiple suppliers in an attempt to reduce the manufacturing costs.

1. Introduction

Our industry is on the cusp of profound and exhilarating change, where digital transformation, mobility, and connectivity are further expanding into new markets while technologies are converging to bring more and more applications into smaller and smarter devices with increased performance, smaller form factor and lower power consumption. The semiconductor industry is breaking records; therefore, the expectations are high for its future, given the explosion of new applications driven by global mega trends, such as environmental consciousness, growth of emerging markets and urbanization, health-care and well-being, connectivity, mobility, enhanced user experience and artificial intelligence, as shown in figure 10.

Figure 10. Global mega trends with impact on semiconductors [3]

Over the past 6 decades, computing and packaging trends have significantly evolved, from the first transistor and through-hole packaging to the advanced systems we have today using 10nm-technology-node manufacturing and a wide range of advanced packaging platforms, addressing single- and multi-die packaging.

Major technology advances in semiconductors were mainly driven by Moore's Law and scaling the technology node. Performance was initially the primary requirement, with the development of personal computers, followed by adoption of internet and then mobile applications. Now the needs of miniaturization and cost reduction, along with improving performance, have become critical drivers. With the growth of connectivity, the internet of things, artificial intelligence, and a wide range of new applications across several markets, the complexity of requirements will continue to increase. Using the same traditional Moore's Law, driving and developing next-generation devices has already started to face various challenges – from larger chip sizes to increased developmental times and costs – reasons why the industry is looking into alternative architectures and advanced integration technologies.

While Moore's Law progress continues, mostly with respect to technology scaling though not cost, the industry's attention has started to expand into More than Moore and Heterogeneous Integration to address growing requirements, as shown in figure 11.

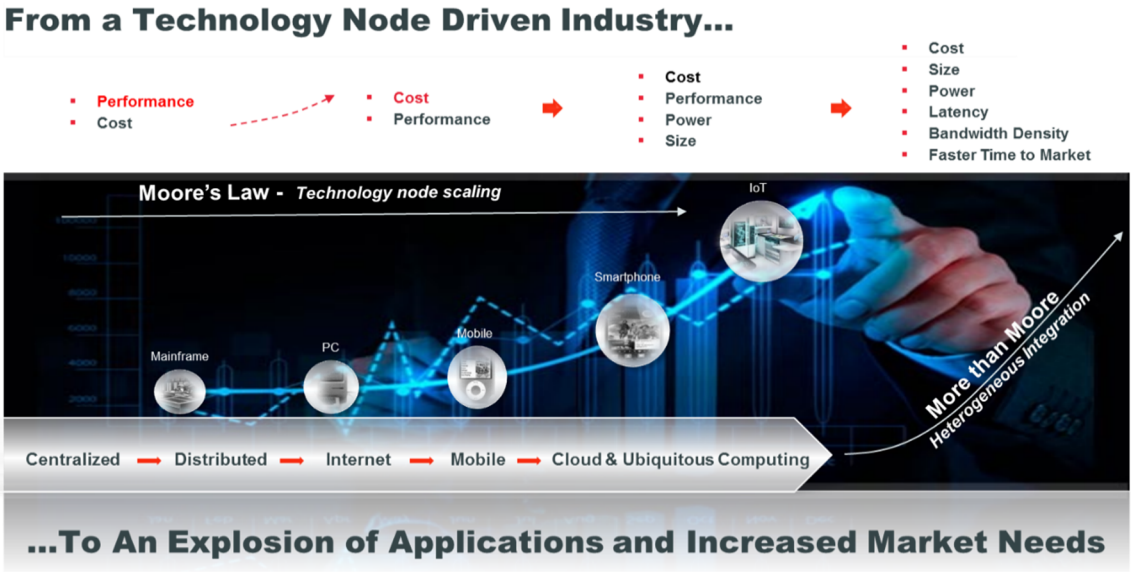


Figure 11. Computing and market requirements evolution [3]

There are several major benefits and considerable added value that heterogeneous integration can bring, from smaller form factors, higher performance, and faster time to markets, to lower cost and increased flexibility. Due to these benefits, application of heterogeneous integration has expanded into various applications across several market

segments, as shown in figure 12. 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 systems, with applications in mobile, IoT, automotive, healthcare, high performance computing and data centers as well as aerospace and defense market segments.

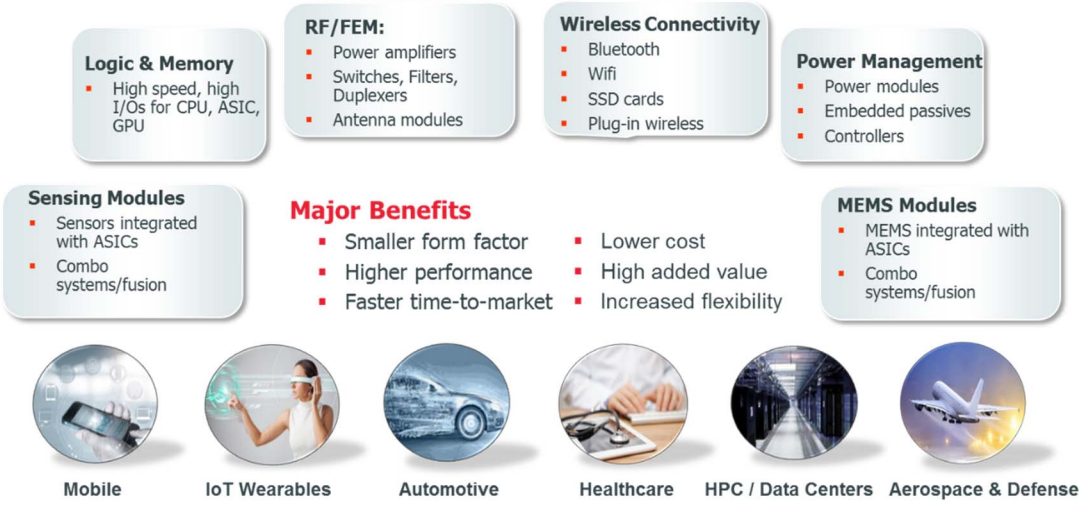


Figure 12. Heterogeneous Integration Benefits and Applications [3]

2. Scope

Heterogeneous integration (HI), 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 13. 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.

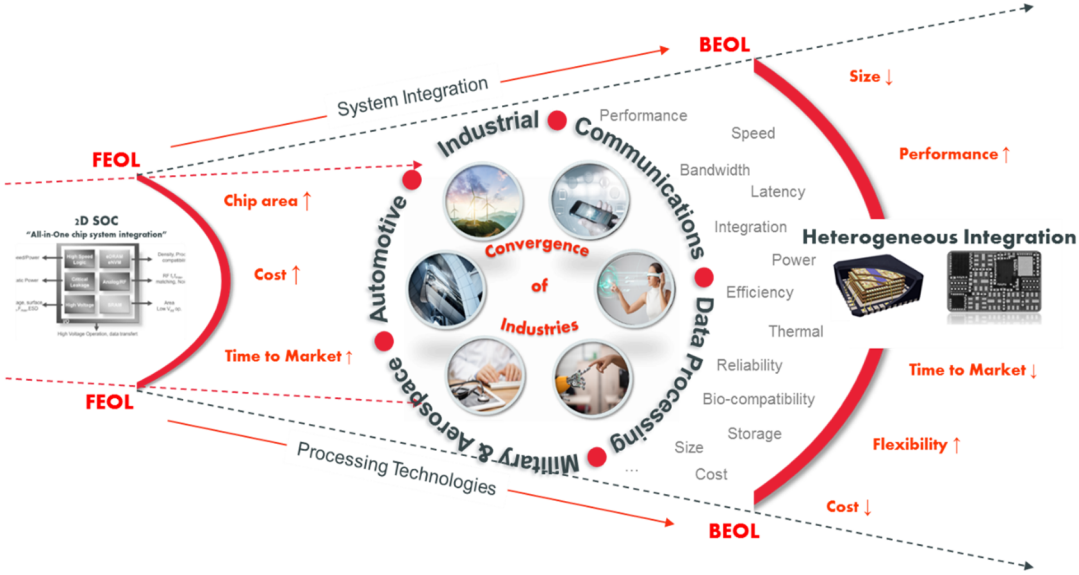


Figure 13. From 2D SoC integration to heterogeneous integration [4]

There are various packaging platforms that can be used for heterogeneous integration, from embedded technologies to wafer level packaging, flip chip and 3D IC based technologies using device stacking, and through-silicon via interconnects.

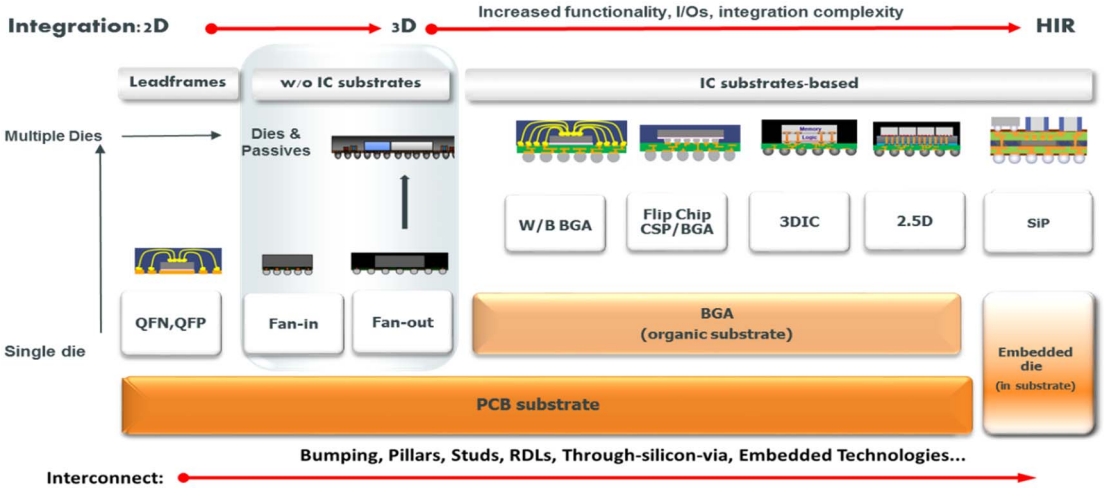


Figure 14. Advanced packaging platforms. WLP, comprising fan-in and fan-out approaches, is the focus of this chapter [5]

Most of the packaging platforms in figure 14 have started with single-die packaging and two-dimensional (2D) integration. With the increase in 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 14.

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 MEMS or 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).

This chapter is organized into 8 sections:

- Executive summary
- Introduction
- Scope
- Difficult challenges
- Discussions on key technical issues
- Equipment considerations
- Summary and final remarks
- References

3. Difficult 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 & equipment; and implementation into the semiconductor ecosystem. The technical challenges will be addressed in separate sections later in this chapter.

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 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.

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.

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 are those customers with high enough volume requirements that can 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 like 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.

4. Technical Overview

Wafer level packaging (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 IO 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. However, there are 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.

WLP technology includes wafer-level chip-size packages (WL CSPs), Fan-out wafer-level packages, wafer capping and thin film capping on MEMS devices, wafer-level packages with Through Silicon Vias (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 consumer products in the future.

4.1 Market Drivers and WLP Capabilities

There are several market drivers, as shown in figure 15 below, that can be successfully addressed by WLP (both fan-in and fan-out) and which result in increased adoption of both packaging platforms, for individual packages as well as for further evolution and adoption in multi-die and heterogeneous integration. These are:

- 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.

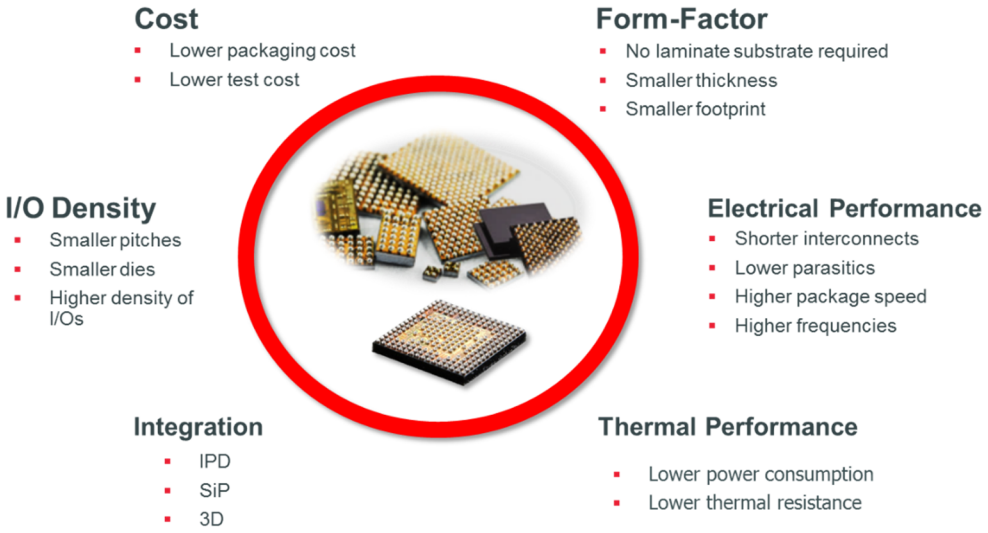


Figure 15. WLP market drivers [5]

If there are many similarities within fan-in and fan-out packaging market drivers, main markets, and technologies and processes applied, there are also specific ones when it comes to fan-out, as shown in figure 16. Fan-in is mainly applied to the mobile market (90% of its market); fan-out, besides representing 90% of the mobile market, has also found applications in the high-end market (networking) due to its ability to address higher I/O count and larger package form factors; it has also improved reliability performance and increased capability for 3D integration using PoP and SiP heterogeneous integration.

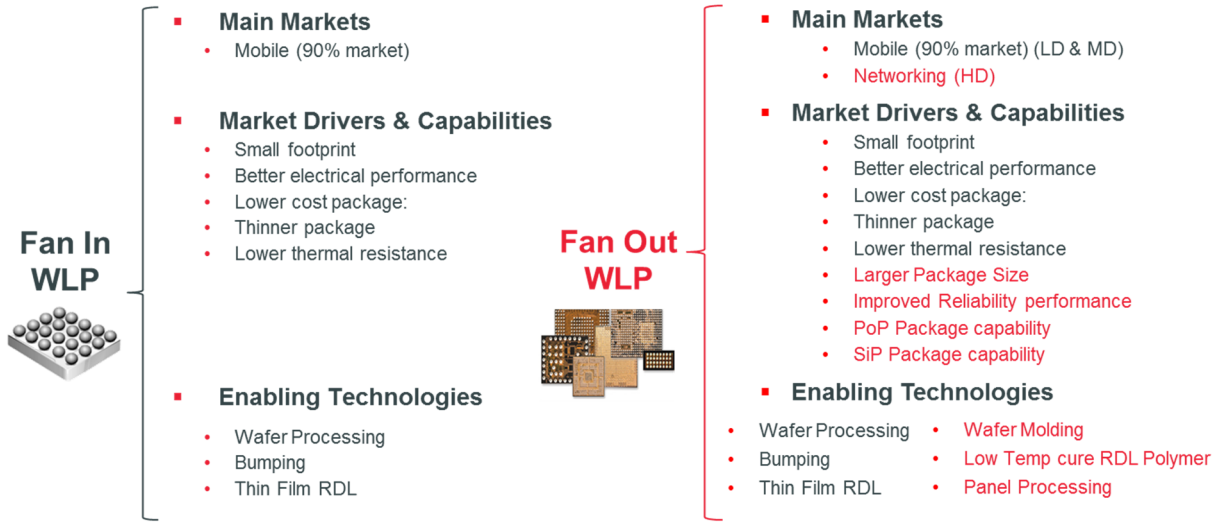


Figure 16. Main Markets, Drivers, Capabilities and Enabling Technologies for WLP [5]

4.2 WLP variations, integration schemes, applications and market growth

Wafer Level CSP was the first generation of wafer-level package product to be introduced into the marketplace. Today WLP technology (fan-in WLP), with and without redistribution layers (RDL), has become one of the most mature technologies and is used for a large variety of products.

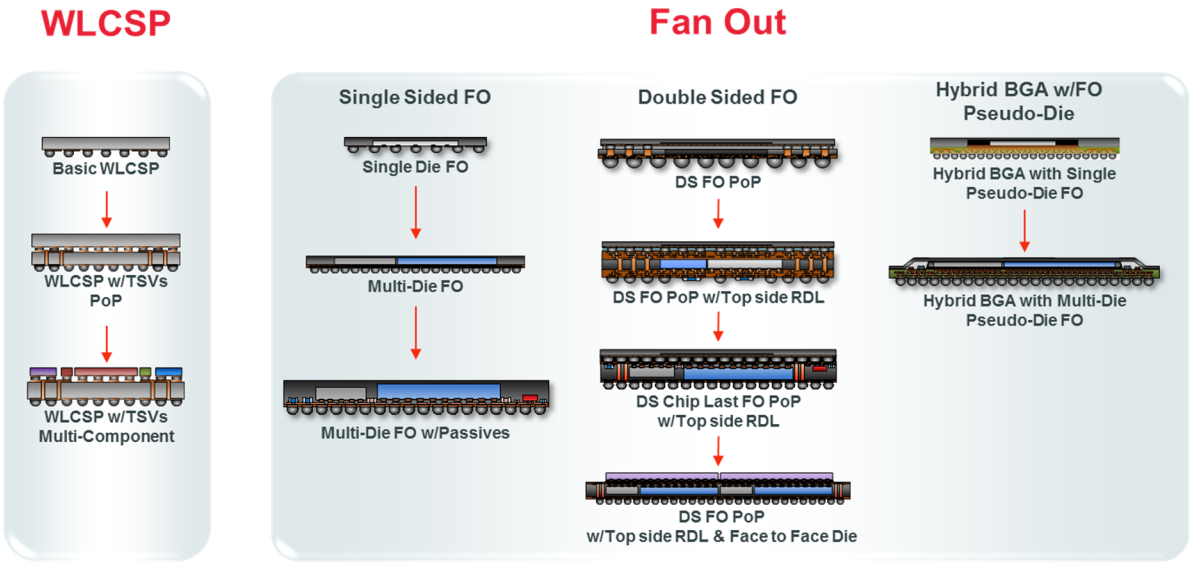


Figure 17. Heterogeneous integration using WLP [5][17]

WLPs with fan-in design historically had been for lower I/O counts and small die sizes, with low power requirements. Recently, WLPs have been made which pass JEDEC reliability criteria with I/O counts greater than 400 at currents 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.

With the introduction of TSVs, IPDs, chip-last fan-out, and MEMs packaging technologies, WLP products using integration schemes such as the ones presented in figure 17 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 WLPs in the packaging field.

WLCSPs started high-volume manufacturing around 2000 and have been mostly limited to single die packages. The first commercially available high-volume fan-out packaging technology, Embedded Wafer Level BGA (eWLB), developed by Infineon and introduced into production in 2009, also was limited to single die packages.

By the nature of the package, WLCSPs have limited capability for integration of multiple components. Figure 18 shows a simplistic image of a basic single die WLCSP.

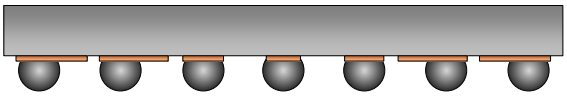


Figure 18. Basic Single Die [17]

Probably the first use of WLCSP for integration was the mounting of a thin flip-chip die face up on the underside of the WLCSP. This die was thin enough that the WLCSP could be mounted and still have clearance to the under-mounted die, as shown in Figure 19.

Figure 19. WLCSP with thin die mounted to underside [17]

With the development of TSV via-last technology, TSVs could be formed in the WLCSP, providing double-sided connectivity. This allows for the mounting of a second die on the top of the primary WLCSP die, or other components such as MEMS devices or passives, as shown in Figure 20.

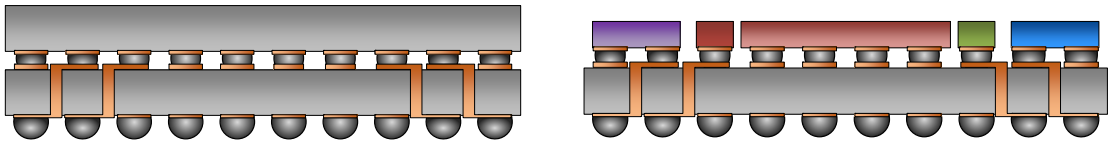


Figure 20. WLCSP Through silicon vias for double sided mounting [17]

Fan-in wafer level packages remain a highly important and constant platform. Fan-in wafer level packages, as shown in figure 21, represents 16% of the total number of packages while serving 4.4% of the total wafer market at 1.5% of the total semiconductor revenue.

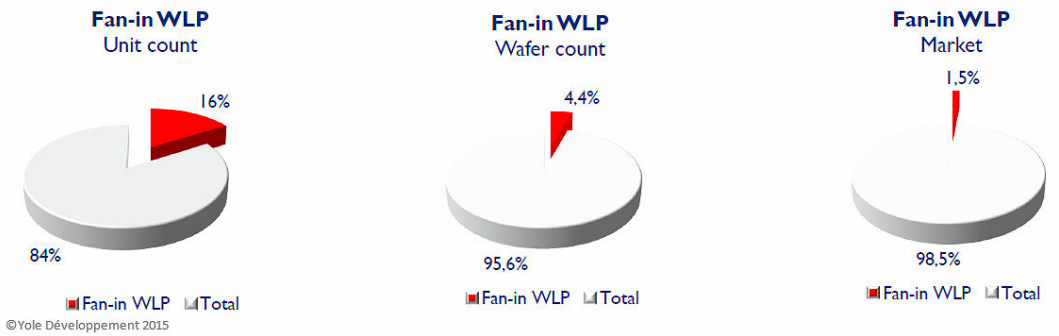
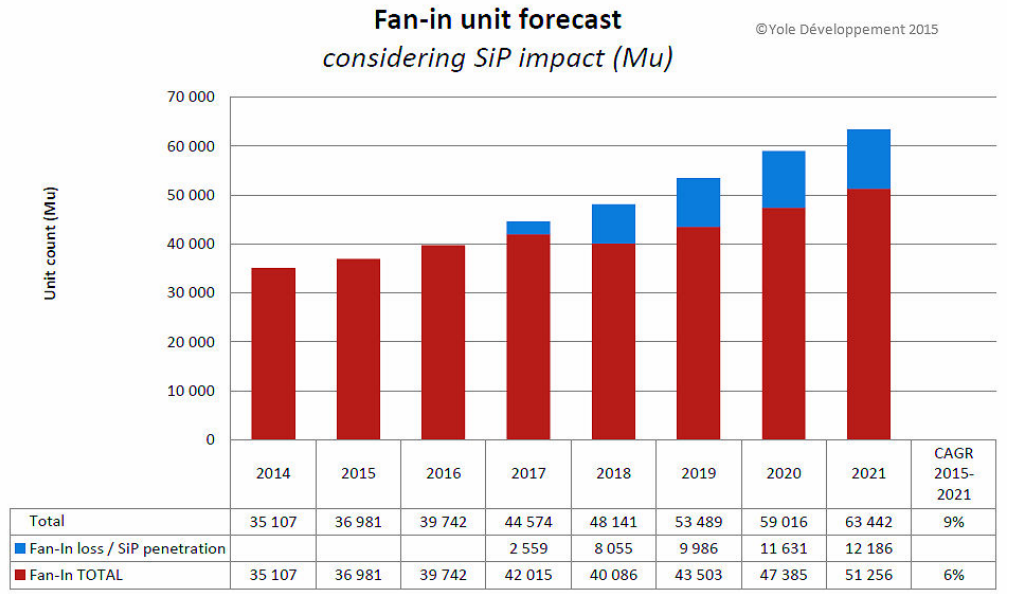


Figure 21. Market share of fan-in packages [6]

Fan-in WLPs are experiencing continuous growth and attracting new applications. The mobile market continues to be the main driver for fan-in WLP with more than 90% of all fan-in packages found in handsets (especially smartphones) and tablets. Fan-in WLPs are considered to be well positioned for other applications, though, such as applications in IoT and wearables.



	2014	2015	2016	2017	2018	2019	2020	2021	CAGR 2015-2021
Total	35 107	36 981	39 742	44 574	48 141	53 489	59 016	63 442	9%
Fan-In loss / SiP penetration				2 559	8 055	9 986	11 631	12 186	
Fan-In TOTAL	35 107	36 981	39 742	42 015	40 086	43 503	47 385	51 256	6%

Figure 22: Growth of Fan-in and SiP Penetration [7]

With a market just short of 3 billion US$ (WLP services) in 2014, fan-in WLP is forecast to continue a stable growth at a compound annual growth rate (CAGR 2014-2020) of 8%. The total wafer count in 300mm equivalent wafers has reached 4 million with a projected CAGR of 8% while the unit count is found at 35 billion with a projected CAGR of 9%. The leading applications by wafer demand in the analog/mixed signal/digital domain are BT+WiFi+FM combos and RF transceivers, followed by PMU, audio/video codecs, DC/DC converters, and ESD/EMI IPD. MEMS devices are led by digital compasses, RF filters, accelerometers and gyroscopes. CMOS image sensors are strongly positioned in 2nd place by overall fan-in application rankings. BT+WiFi+FM combos, CMOS image sensors and RF transceivers account for ~50% of all fan-in WLP applications. While analog/digital/mixed signal and MEMS markets are on the rise, the CIS market overall is expected to stagnate due to rising pixel count of smartphone sub- and feature-phone main cameras, exceeding fan-in capabilities. If 49% of all

CIS were found in fan-in in 2014, this percentage is expected to decline to 31% by 2020. CIS fan-in WLP services and market growth will depend on the penetration of automotive, surveillance and wearables sectors in comparison with the decline of the mobile phones sector.

The fan-in platform will be adapted for system-in-package and heterogeneous integration. Functional integration of fan-in packaged dies into SiP also have a considerable impact on the fan-in supply chain, cost and overall demand. Yole Developpement has estimated a potential drop of Fan-in demand (single devices) starting with 2017 due to integration of some of these devices into SiP packages, especially fueled by growing IoT applications.

The fan-in adoption in SiP is estimated to grow from 2.6 billion units in 2017 to 12.2 billion units in 2021, reducing fan-in individual packaging growth from 9% to 6%, as shown in figure 22.

Development of fan-out began in the early 1990's but didn't go into volume production until May of 2009 with the Infineon eWLB product. This was a simple single-die package, shown in Figure 23.

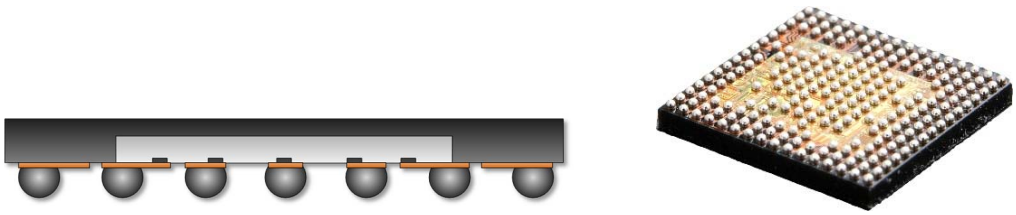


Figure 23. eWLB fan-out [17]

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. We can further classify eWLB and RCP as a "die down" chip-first processes, as the die is placed on the temporary carrier before overmolding in a die-face-down position. Figures 24 and 25 give a simplistic process flow of the chip-first and die-down eWLB and RCP structures.

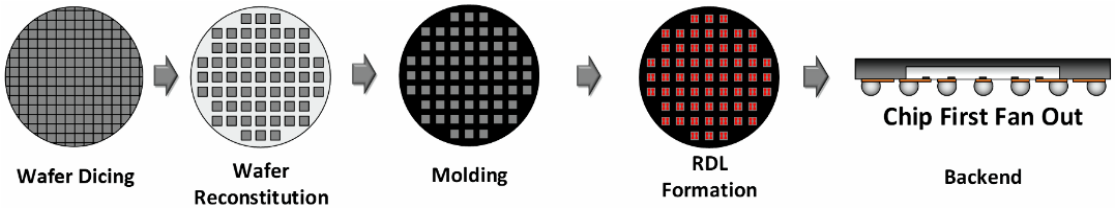


Figure 24. Chip-first process flow [17]

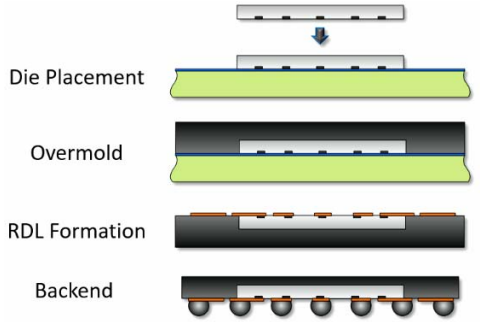


Figure 25. eWLB fan-out die-down process flow [17]

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 Figure 26 and 27. 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.

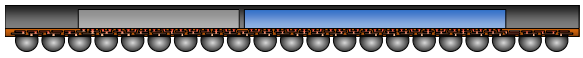


Figure 26. Multi-die eWLB type fan-out [17]

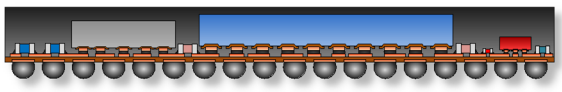


Figure 27. Multi-die eWLB fan-out with passive components [17]

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 28.

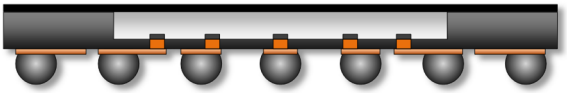


Figure 28. Chip-first die-up fan-out package [17]

A simplified process flow for the die-up fan-out is illustrated in Figure 29.

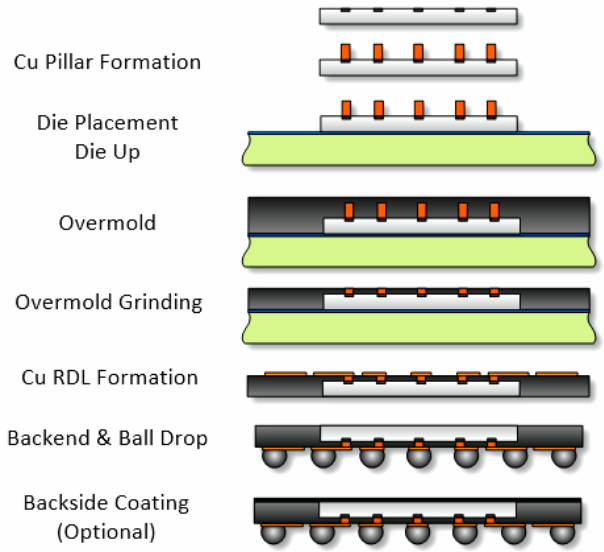


Figure 29. Fan-out die-up process flow [17]

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 30.

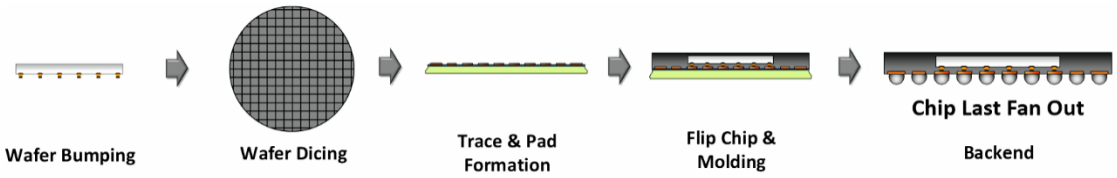


Figure 30. Chip-last process [17]

Although the majority of early fan-out activity was focused on 2D single-sided applications, 3D and double-sided fan-out structures were being 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 boards 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 31.

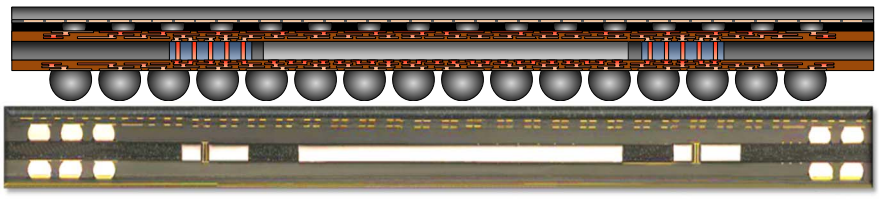


Figure 31. 3D fan-out with embedded TSV die [17]

Expanded illustrations of these types of 3D interconnects for fan-out are shown in Figure 32.

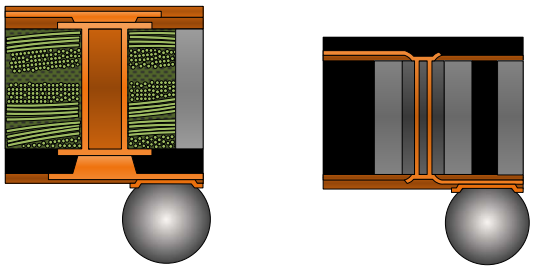


Figure 32. 3D fan-out interconnect using circuit board via or TSV in silicon [17]

One possible process flow for this type of 3D structure is shown in Figure 33.

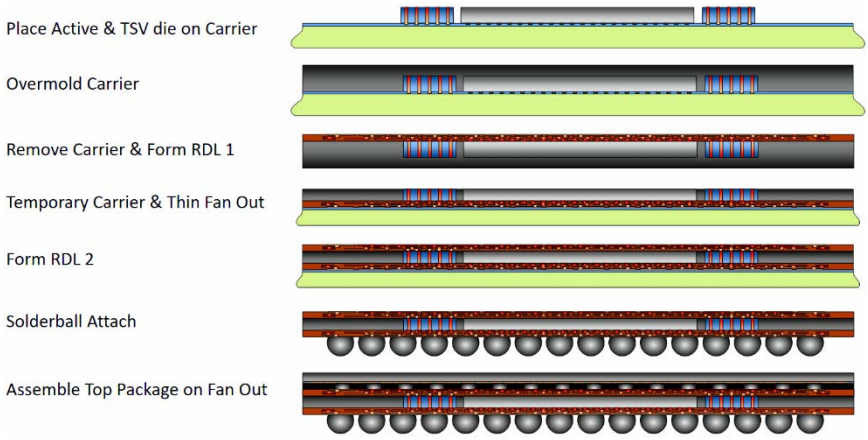


Figure 33. 3D fan-out with embedded TSV die process flow [17]

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 34.

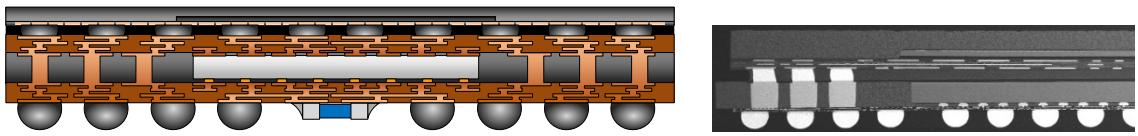


Figure 34. 3D fan-out with plated copper through posts [17]

An expanded illustration of this type of 3D interconnect for fan-out is shown in Figure 35.

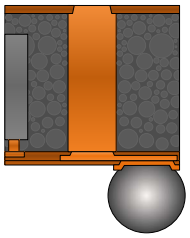


Figure 35. 3D fan-out interconnect with plated copper through post [17]

A process flow for this type of 3D structure is illustrated in Figure 36.

The best know 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.

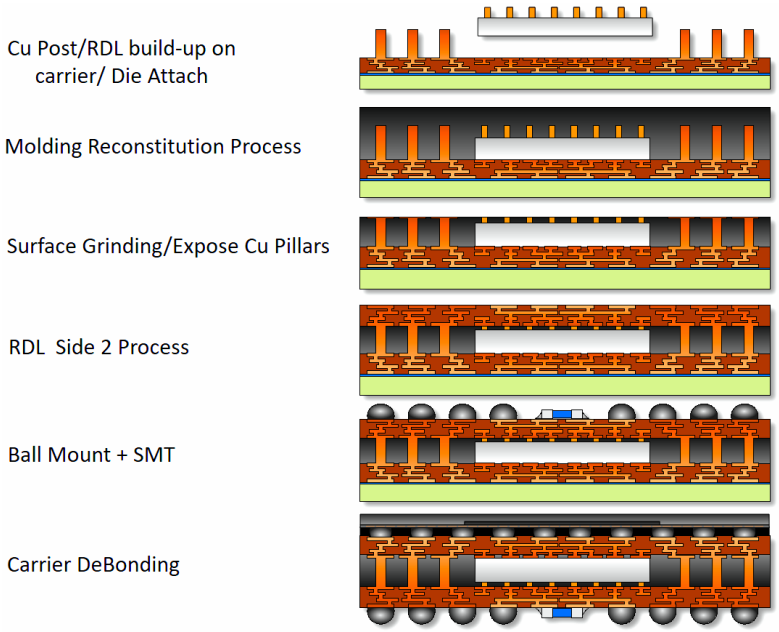


Figure 36. 3D fan-out with plated copper through posts process flow [17]

At first glance, InFO is yet 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 10x10mm^2, Chip size larger than 8x8mm^2 with very high BGA pin count.
- First high-volume manufactured multi-die fan-out including 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 embedding-die fan-out technology
- First mobile AP PoP package with final thickness below 1mm.
- First foundry-delivered advanced fan-out technology to 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 including 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's. It helped to bring the packaging industry to the central stage of the semiconductor industry.

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 among 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 37 shows a top view of the copper through-via post layouts for the TSMC InFO package.[8]

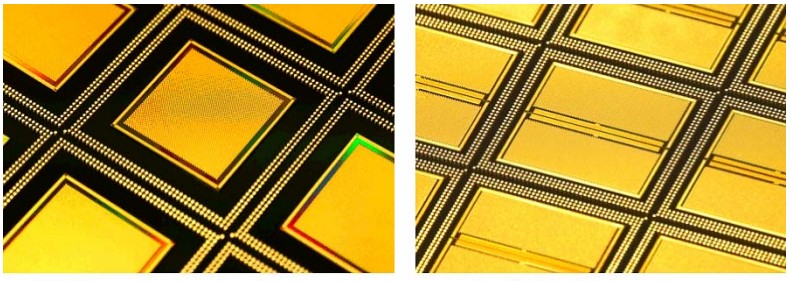


Figure 37. Top view of InFO packages showing 3D through-via pads [8][19]

TSMC has taken this concept one step further, incorporating multiple SoCs and memories into a multilayer stackup using an advanced structure called 3D Multi-stack (MUST) system integration technology, 3D MUST-in-MUST (3D-MiM) fan out package.[9] An example of this is shown in Figure 38.

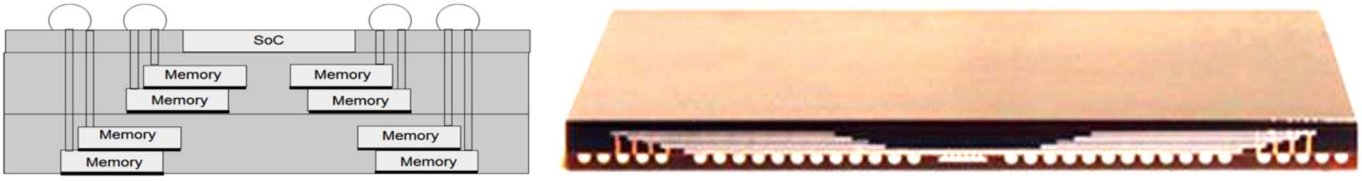


Figure 38. TSMC 3D-MiM Package [9]

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 [10], as shown in figure 39.

Figure 39. Nepes 3D Fan Out Package with Deep Photo Defined Vias [10]

Fan-out packages were historically just that: stand-alone packages, with either solderballs or land grid arrays for interconnection to the next level of the assembly hierarchy. Within the last few years, products have gone into production using interposers with Through Silicon Vias (TSVs) for the high-density interconnection of multiple die in a package. These have typically 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. This assembly of die, interposer, and substrate became the package, with BGA solderballs for the next level of interconnect. 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 die are embedded in a fan-out assembly using lines and spaces 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 solderballs 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 is shown in figure 40, which is an ASE package on substrate called FOCoS (Fan Out Chip on Substrate)[11].

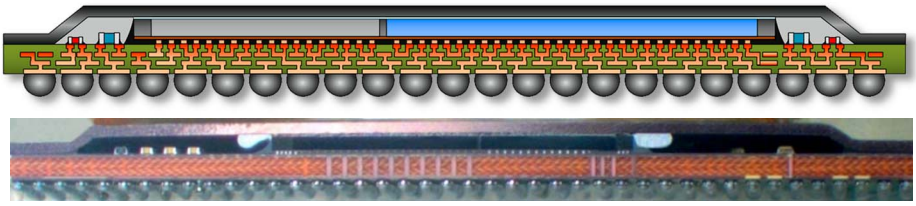


Figure 40. Hybrid fan-out pseudo-die on BGA substrate (ASE) [17]

Multiple suppliers are either in production with versions of this structure, or have versions in various stages of development. Currently the line and space capability of the fan out versions is 1 micron or greater. For extremely high density applications, the solution is still using interposers, which have submicron line and space capability. But for those applications where the fan out hybrid solution is useable, the electrical and thermal performance can be better than a silicon interposer. The fan-out solution is also usually thinner than the interposer, which has typically been in the 75 micron range. The two solutions will co-exist, with each filling its particular niche markets.

4.3 Panel Configuration for Fan-out Processing

Fan-out Wafer Level Packaging (FOWLP) has historically utilized multiple geometric configurations for processing, from small and large rectangular panels to round wafer formats. 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 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. For manufacturing cost considerations, the larger the panel that can be processed, the more cost effective the pricing can be per device. Currently, 300mm wafers are the largest wafer format in volume production. 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 (eg. spin-on coatings), suppliers are required to develop alternative materials, processes and equipment. Again, there is no standardization for the selection of these manufacturing variables.

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 the past year, 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 material 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 material 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. With the wider variety of structures and material sets evidenced by the various developers of larger format panel fan-out, this may become more challenging for the end customers.

4.4 Supply chain activities: infrastructure, materials, equipment

Fan-out Wafer Level Packaging (FOWLP) is one of the latest packaging trends in microelectronics. Besides technology developments towards heterogeneous integration (including multiple die packaging, passive component integration in package, and redistribution layer or package-on-package approaches), larger substrate formats are targeted. Manufacturing is currently done on wafer level up to 12”/300 mm and 330 mm. For higher productivity and therewith lower costs, larger form factors are introduced. Instead of following the wafer-level roadmaps to 450 mm, 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 300x300 mm to 457x610 mm or 510x515 mm up to 600x600 mm or even larger, influenced by different technologies coming from e.g. printed circuit board, solar or LCD manufacturing. The main challenge at the moment is the missing standardization on panel formats. SEMI has now started an initiative on the standardization topic. After a costumer survey on preferred panel sizes, a task force has been established working on a first proposal of a standard.

However, 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 material 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. 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. However, 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 1

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

5. Difficult Technical Challenges

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

- Die Shift during reconstitution molding.
- Low temperature polymer for RDL isolation.
- 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.

First, let us review the three existing challenges listed above.

5.1 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 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.

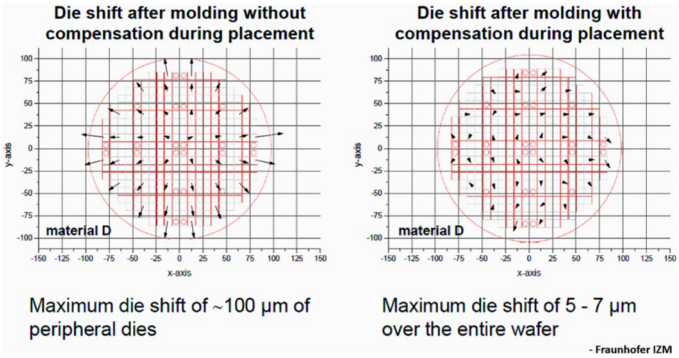


Figure 41. Die shift without and with compensation [20]

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™"[12]. 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 position 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 material 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.

5.2 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 material properties were not ideal for the reliability requirements of fan-out packaging. Around 2012, new polymers became available that had improved material 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 as low as 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.

5.3 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 material structures. This results in unbalanced TCE's 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 42.

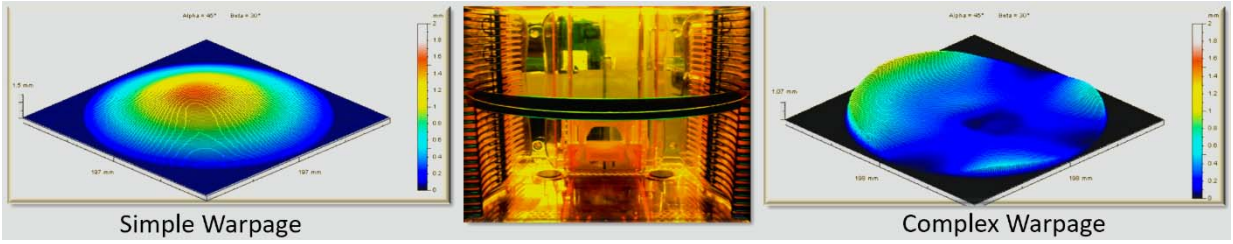


Figure 42. Examples of molded wafer warpage [17]

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 foups or 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 43, which shows warpage at various process steps in an eWLB process flow.

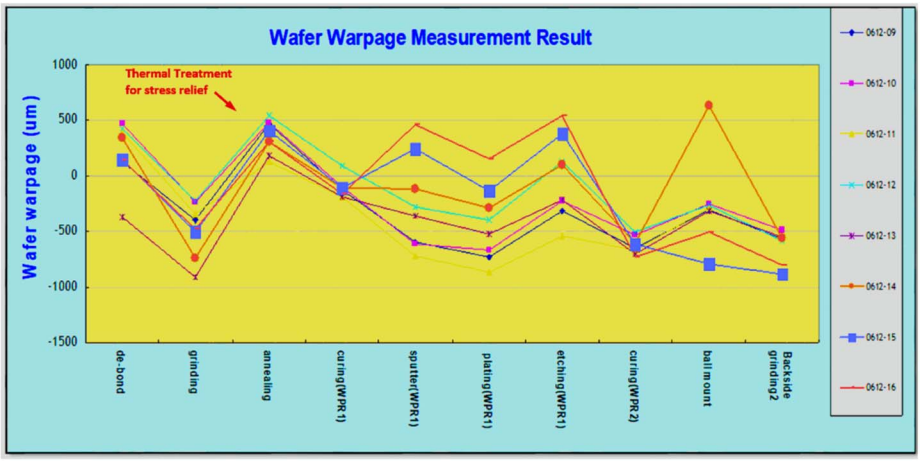


Figure 43. Warpage measurements for 8 reconstituted fan-out wafers during processing (ASE) [17]

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.

5.4 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 multidie fan out with bumps is flipchip assembled onto a BGA substrate.. Examples of fine lines and spaces on these latest high density fan-out products are shown in Figure 44.

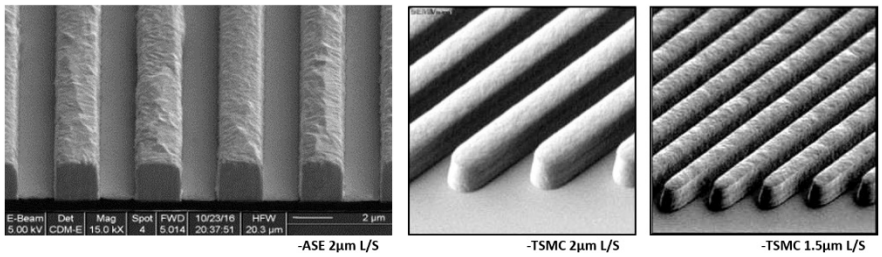


Figure 44. Fine line and space geometries on high-density fan-out

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.[13] Another challenge as the RDL lines and spaces go below 1µm is line to line leakage and long term electromigration concerns.

5.5 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 45.

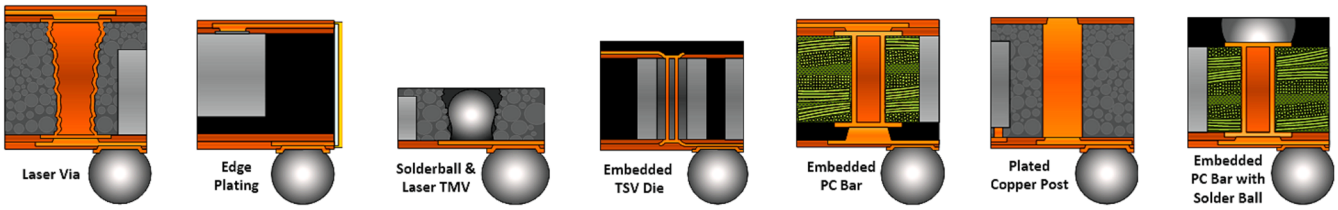


Figure 45. Examples of through-mold connectivity for fan-out [17]

The one structure that is in high volume production with TSMC's InFO fan-out is the Plated Copper Post, Figure 46, including examples from TSMC's InFO and ASE's FOPoP structures.

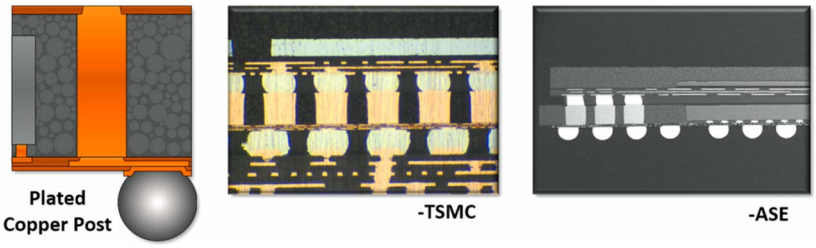


Figure 46. Through-mold connectivity structure in volume production

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?

5.6 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 47.

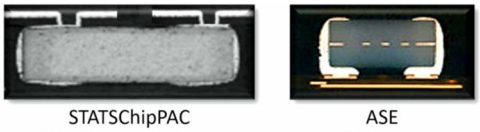


Figure 47. Discrete passive components embedded in fan-out packages

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 type of integrated passive devices can be seen in Figure 48.

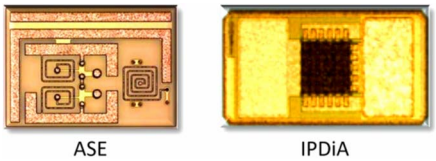


Figure 48. Integrated passive die

The third method of integrating passives into the fan-out package is to incorporate it into the copper RDL structure itself, usually in the form of planar inductors, as shown in Figure 49. These can be extended over the mold compound with favorable Q values. The technique can be extended to include antenna structures in specific applications.

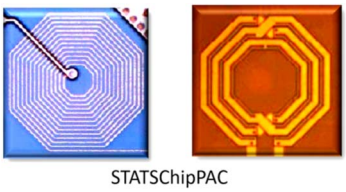


Figure 49. Integrated passive die

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.

5.7 Challenges of die placement speed and accuracy

As discussed in section 5.1 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.

5.8 Challenges of Mold Compound Development

Early volume manufacturing of fan out packaging utilized liquid mold compounds to encapsulate the die in chip first structures.[14][15] The die were encapsulated in these liquid compounds using either compression molding [14] or using dam and fill [15]. The liquids were initially simpler to use, and are still in use today; however, they have been supplemented by lower price 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.

5.8 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.

6. Equipment Considerations

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.

6.1 Physical Vapor Deposition (PVD)

The most common PVD systems in use for both WLCSP and fan-out are sputtering systems. Fan-out has some more stringent capability requirements than systems used solely for WLCSP.

PVD for 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.

Low contact resistance: In mobile applications particularly, contact resistance is battery life. Packaging companies want low contact resistance, at the same time as 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.

Step coverage: 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 weakness. 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. [16]

7. 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, only following Moore's law, as the industry was accustomed to for the past six decades. 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 performances and new applications.

Wafer Level CSP was the first generation of wafer level package product to be introduced into the marketplace. Today WLP technology (fan-in WLP) with and without redistribution layers (RDL) is used for a large variety of products. WLPs with fan-in design historically had been for lower I/O counts and small die sizes, with low power requirements. Recently WLPs have been made which pass JEDEC reliability criteria with I/O counts greater than 400 at currents 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.

With the introduction of TSVs, IPDs, chip-last fan-out, and MEMs packaging 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 WLPs in the packaging field.

Acknowledgments

The chairs of the WLP technical working group would like to acknowledge the contribution of the following members: Jerome Azemar from Yole Developpement, Monita Pau from DowDuPont, Doug Yu from TSMC, David Butler from SPTS, Phil Garrou and all other members of the WLP chapter committee who contributed to this work.

8. References

[1] Prutec Limited, Patent Application GB 2 130 794A, United Kingdom
[2] Casio, CPMT Seminar, Advanced Substrate Technologies for SiP/SoP, 2006
[3] Beica, R. "The Growth of Heterogeneous Integration and The Importance of Electronic Materials", IMAPS Device Packaging Conference, Fountain Hills, AZ, 2018, pp 17
[4] Beica, R. "Enabling Information Age through Advanced Packaging Technologies and Electronic Materials", 24th Pan Pacific Microelectronics Symposium, Hapuna Beach, Hawaii, 2018, pp. 4
[5] R.Beica, J. Hunt, "Heterogeneous Integration Roadmap; WLP TWG", ", IEEE 68th Electronic Components and Technology Conference (ECTC), 2018
[6] Ivankovic, A., "Fan-in Business", Yole Developpement, 2015
[7] Azemar, J. et al., 2015 WLCSP Technology and Market Trends Report, Yole Developpement
[8] Yu, Doug C.H., TSMC (private communication)
[9] 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
[10] 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
[11] 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
[12] Scanlan, C et al, Deca Technologies, Inc, "Adaptive Patterning for Panelized Packaging", IWLPC, 2012
[13] Yu, C.H. et al., TSMC, "High Performance, High Density RDL for Advanced Packaging", IEEE 68th Electronic Components and Technology Conference (ECTC), 2018
[14] Brunnbauer, M. et al, Infineon, "An Embedded Device Technology Based on a Molded Reconfigured Wafer", 56th ECTC, San Diego 2006
[15] Leung, John, Freescale Semiconductor, "Packaging Technology for Mobile platforms", Packaging Conference, Shanghai, 2006
[16] David Butler, SPTS (private communication)
[17] Hunt, J., ASE Group (private communication)
[18] 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
[19] Tseng, CF et al., TSMC, "InFO (Wafer Level Integrated Fan-Out) Technology", IEEE 66th Electronic Components and Technology Conference, 2016
[20] Braun, T et al., Fraunhofer IZM, "Through Mold Vias for Stacking of Mold Embedded Packages", IEEE 61st Electronic Components and Technology Conference, 2011

Edited by Paul Wesling