Integrated Circuit Heterogeneous Manufacturing Technology

Yuxin Li [1] and Zihan Shen[1]

[1] School of Microelectronics, South China University of Technology, Guangzhou 511442, China
202264680084@mail.scut.edu.cn

Abstract. Traditional integrated circuit technology attempts to integrate all functions into a single chip, this leads to the chip becoming more expensive and larger. Heterogeneous integration technology solves this problem by combining chips with different processes and technologies. This technology can continuously increase the functional density and reduce the cost of each function to ensure continuous progress in electronics in terms of cost and performance. Heterogeneous integration technology that integrates different types of chips is crucial for achieving higher performance, lower latency, smaller size, lighter weight, lower power consumption requirements, and lower cost, and is also key to continuing Moore's Law in the post-Moore era. Among them, fan-out wafer-level packaging (FOWLP) and 2.5D/3D packaging have become important technical routes in advanced packaging. This paper reviews the current development status of these two technologies: FOWLP technology uses a fan-out wiring structure to achieve high-density interconnection by forming a redistribution layer (RDL) around the chip; 2.5D packaging mainly uses silicon interlayers or organic interlayers to achieve interconnections between chips, while 3D packaging achieves vertical stacking of chips through through-silicon vias (TSV) technology. The article details the characteristics and key technologies of these three technologies and analyzes the implementation level and industrial application status.

Keywords: Integrated circuit heterogeneous manufacturing technology, Fan-out wafer-level packaging, 2.5D and 3D packaging.

1 Introduction

In the 1990s, with the continuous advancement of integrated circuit manufacturing technology and the shrinking of process nodes, the physical size of transistors was approaching the physical limit. Continuing to increase integration by reducing device size is extremely difficult and costly, and this method can no longer significantly improve chip performance. People began to realize the limitations of single-chip integration and thus began to explore the possibility of multi-chip heterogeneous integration [1]. In this context, heterogeneous integration technology emerged, and after entering the 21st century, 2.5D and 3D packaging technologies began to make [2, 3] their mark. 2.5D/3D heterogeneous integration technology has evolved from the past chip and adapter board

S. Zhang (ed.), *Proceedings of the 2025 International Conference on Electronics, Electrical and Grid Technology (ICEEGT 2025)*, Advances in Engineering Research 292,
https://doi.org/10.2991/978-94-6463-986-5_13

connection to the vertical stacking of chips and is widely used in high-performance computing. At present, more and more semiconductor companies have joined in the research and development and production of heterogeneous integration chip technology, gradually forming a complete industrial ecosystem. Well-known chip design companies such as Intel and AMD are actively developing heterogeneous integrated chip products to meet the market demand for high-performance computing chips. Intel's Foveros technology has successfully achieved 3D integration of chips at different process nodes, and its Core processors have adopted heterogeneous integration technology, integrating computing cores, caches, and I/O modules, thereby enhancing overall performance. In the field of packaging and testing, companies such as TSMC and ASE are increasing their investment in research and development and production capacity of 2.5D and 3D packaging technologies, aiming to provide advanced packaging services for heterogeneous integrated chips. Meanwhile, equipment manufacturers are constantly introducing advanced equipment to support the manufacturing of heterogeneous integrated chips. For instance, ASML's lithography machines are being continuously upgraded to support the lithography process for heterogeneous integrated chips. To reduce the cost of some chips, fan-out wafer-level packaging (FOWLP) technology has emerged, which can integrate chips with different functions into a single package to produce electronic chips with lower cost but higher performance.

This article aims to analyze and summarize these technologies, examine their key technologies and advantages, and provide a reference for their development.

2 Fan-out wafer-level packaging

Fan-out wafer-level packaging (FOWLP) has the potential for heterogeneous integration and 3D stacking, which can be combined with a variety of advanced packaging technologies and is the cornerstone of the future evolution of advanced packaging technologies [4]. This technology embeds the chip into the reconfigured wafer and uses the rewiring layer (RDL) to achieve I/O fan-out, breaking through the size limitations of traditional fan-in [5]. There are two important concepts in FOWLP, namely fan-out packaging and wafer-level packaging. Fan-out is the opposite of Fan-in, where I/ O interfaces are located below the Die and the number of I/ O interfaces is limited by the size of the chip. With the development of chip technology, The number of I/ O interfaces has become one of the weaknesses restricting the performance of the chip, while fan-out packaging can take advantage of the additional chip area provided by rewiring (RDL) technology and molding compounds to distribute I/ O interfaces outside the die, greatly increasing the number of I/ O interfaces of the chip and thus meeting the growing throughput requirements of the chip [6]. Conventional packaging often cuts the chips off the Wafer and packages them separately, but wafer-level packaging takes a different approach by first encapsulating the wafer as a whole and then cutting it, which is more suitable for the mass production of integrated circuits [7].

Compared with fan-in wafer-level packaging, the structure is shown in Figure 1 (a), which adds a protective shell to the outer layer of the chip. The tin balls can be extended beyond the size of the chip through RDL technology, increasing the number of I/O contacts, increasing the usable area of the chip and reducing the cost.

Compared with the flip-chip plastic ball grid array package, the structure is shown in Figure 1 (b). FOWLP is thinner, as shown in Figure 1. It ignores the wafer bump process, does not require flip-chip reflow soldering, omits the flux cleaning step, does not need underfilling, and has better electrical and thermal performance. And it is easier to implement system-in-package (SIP) and three-dimensional integrated circuit (3D IC) packaging.

Compared with wafer-level chip-size packaging, the structure is shown in Figure 1 (c), FOWLP uses known qualified chips, resulting in higher chip yield and support for multi-chip integration. The RDL rewiring layer can be extended beyond the chip size, with more pins and full utilization of the effective area of the silicon wafer.

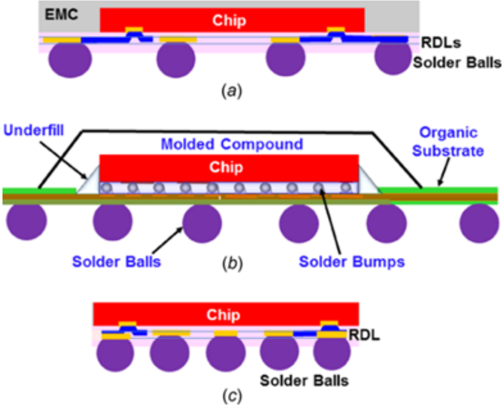


Fig. 1. (a) FOW/PLP, (b) flip chip plastic ball grid array, and (c) WLCSP [8].

In fan-out wafer-level packages, according to the order of chip placement and routing (RDL) and the way the chips are placed, there are three categories: face-up chip processing first, face-down chip processing first, and RDL priority.

The key technologies include wafer reconfiguration, encapsulation, and rewiring. Grains cut from intact wafers are placed on carriers to form reconfigured wafers, heated and liquefied epoxy encapsulation encapsulates the grains, cooled and solidified encapsulation protects the chips and expands the chip area, and rewiring technology alternately fabricates metal layers and polyimide layers on the surfaces of the grains and encapsulation to redistribute I/O interfaces.

FOWLP has a solid foundation of use in commercial products, facilitating the development of key components such as programmable logic array FPgas, central manager cpus, and digital signal processing modules (DSPS) towards high density, high performance, and high reliability. As military, aerospace and other fields move towards smaller size, lighter weight and higher reliability, FOWLP will play a key role in the packaging transformation of military and aerospace devices [9, 10].

2.1 InFO technology

In 2016, TSMC's InFO technology enabled the connection between the chip and the substrate through RDL and bumps without the need for wire bonding, allowing for more I/O connections and enabling a more compact and efficient design. By increasing the number of pins, reducing the pin pitch and thinning the package thickness through a substrate less pitch metal process and RDL rewiring technology. Optimized performance and reduced the cost of mobile computing products.

TSMC first unveiled InFO technology in 2016, presenting the first-generation INFO-POP architecture, which replaces traditional organic substrates with epoxy molding compound and uses the then industry-leading copper wiring 2um/2um redistribution layer. The InFO technology was applied to the Apple A10 processor and then gave rise to new applications such as info-OS, INFO-LSI, infO-POP and infO-AIP [8].

In high-performance computing network applications, single-node advanced SOC chips can no longer meet the growing demand for switching capacity due to cost and performance.

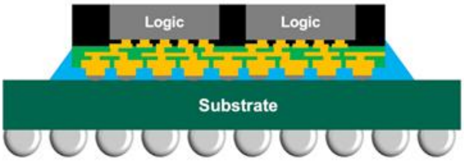
The InFO-oS technology has been developed to solve this problem. This technology has multiple layers of high-density 2/2umRDL line widths/spacings and can integrate multiple advanced node switching small chips to reduce costs and improve performance [9].

Similar to Intel's Embedded Multi-chip Interconnect Bridge (EMIB) technology, INF-LSI enables the ultimate trade-off between interconnect bandwidth and cost, using silicon-based interconnects and TSVS to achieve direct vertical connections across different chip layers, enabling high-speed signal transmission within the same package [11-13]. InFO-LSI is suitable for areas such as high-performance computing, artificial intelligence, communication and networking devices that require high-speed and high-bandwidth information transmission.

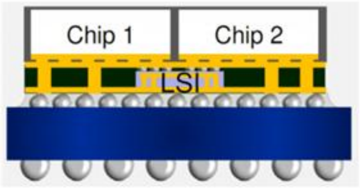
InFO-PoP technology (Integrated Fan-Out Package on Package) combines fan-out wafer-level packaging and PoP, using TSV technology for integrating multiple functional chips in the same package.

InFO-AiP technology integrates the antenna directly into the package, achieving higher integration and better signal transmission performance, as well as enhanced anti-interference stability. It is commonly used in areas such as mobile devices, the Internet of Things, and communication equipment, and can achieve better wireless connection performance [13].

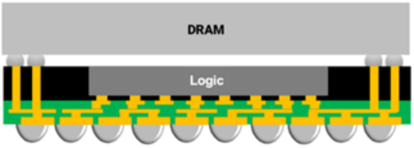
Figure 2 shows the structural diagrams of InFO-oS, InFO-LSI, InFO-PoP and InFO-AiP technologies, respectively.



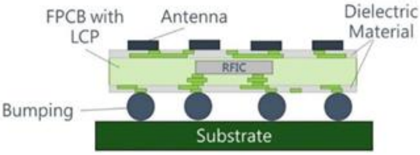
InFO-oS



InFO-LSI



InFO-PoP



InFO-AiP

Fig. 2. Illustration of InFO technology derivative application packaging.

3 2.5D and 3D packaging

2.5D packaging involves attaching two or more chips to a silicon interlayer, where the signal from one chip is translaterally transmitted to the other using the silicon interlayer, and the signal is translaterally transmitted at this time. 3D packaging, on the other hand, stacks two chips to allow small signals to travel vertically between them.

2.5D packaging involves attaching two or more chips to a silicon interlayer, where the signal from one chip is translaterally transmitted to the other using the silicon interlayer, and the signal is translaterally transmitted at this time. 3D packaging, on the other hand, stacks two chips to allow small signals to travel vertically between them.

The manufacturing process for 2.5D packaging and 3D IC packaging is also different. 2.5D packaging requires the fabrication of silicon-based interstitial layers and complex process steps such as lithography; 3D IC packaging requires difficult manufacturing processes such as direct bonding technology.

In addition, 2.5D packaging and 3D IC packaging also differ in terms of application scenarios and performance. 2.5D packaging is typically used in high-performance computing, network communication, artificial intelligence, mobile devices and other fields, with high performance and flexible design; 3D IC packaging, on the other hand, is typically used in memory, sensors, medical devices, etc., with high integration and small package size. TSMC's CoWoS technology is a 2.5D packaging technology; Intel's Foveros technology is a 3D packaging technology. To achieve 2.5D, 3D IC integration, several key technologies are needed, such as TSV technology, which plays a crucial role in providing thermal and electrical pathways between layers. 3D integration relies on TSV because it reduces package size and shortens interconnect paths, while allowing for shorter interconnects and smaller pad sizes and spacings [14].

3.1 TSV technology

TSV technology is one of the most core technologies in advanced semiconductor packaging. Through-silicon via technology is the latest technology that enables interconnections between chips by creating vertical conduction between chips and wafers. By filling some conductive materials such as tungsten, copper and polycrystalline silicon, TSV technology enables vertical electrical interconnections within silicon wafers, making vertical integrated chips possible.

Through-silicon via technology (TSV), which plays a crucial role in 2.5D and 3D packaging, enables electrical interconnection between multilayer chips by filling conductive metals after etching deep holes inside the silicon chip to create vertical conductive paths, reducing signal delay and loss and enhancing the integration and performance of the device [15]. As a result, TSV technology can be widely applied in multiple fields, not only enabling three-dimensional stacking of memory chips such as HBM and HMC, but also heterogeneous integration of different functional devices in Microsystems such as MEMS, sensors, RF devices, etc. And with the development of the technology, it is expected to play a greater role in future packaging [16].

TSV is a high-density packaging technology that achieves smaller interconnect lengths, reduced signal delay, and decreased capacitance and inductance through vertical electrical interconnects via silicon vias. The TSV manufacturing process mainly consists of these steps: laser drilling or ion deep etching (DRIE) to form the through-hole; Intermediate dielectric layers are deposited through thermal oxidation or plasma-enhanced chemical vapor deposition (PECVD); Deposit barrier and seed layers through physical vapor deposition (PVD); Fill the TSV holes with copper or tungsten holes by electroplating or PVD process; Chemical and mechanical polishing (CMP) of the copper layer (coating).

3.2 CoWoS technique

CoWoS is strictly a 2.5D packaging technology, a combination of CoW and oS. The chip is first connected to the wafer via CoW(chip on wafer) and then to the substrate via oS(on substrate). It stacks different chips into a larger interlayer area, achieves high-density interconnections through techniques such as RDL rewiring, and finally connects the wafers to the substrate using C4 bumps. CoWoS technology can be subdivided into three types of lines: S, R, and L, namely CowOS-S, CowOS-R, CowOS-L, as shown in Figure 3.

CoWoS-S uses silicon substrates as the interlayer. As shown in Figure 3 (a), it provides high-density interconnections and deep groove capacitors on a larger silicon interlayer area to accommodate chips and wafers with various functions, including logic chips and HBM (high band memory) cubes.

CoWoS-R uses a silicon interlayer with an RDL interlayer, as shown in Figure 3 (c), and utilizes a redistribution layer as an interconnect between the system-on-chip (SoC) and HBM to achieve heterogeneous integration.

CoWoS-L combines CoWoS-S and InFO technology, as shown in Figure 3 (c), to achieve inter-chip interconnection using an interlayer with LSI (local silicon interconnection) chips. In 2011, TSMC officially launched CoWoS technology to provide solutions for the integration of high-performance products such as FPgas and Gpus. Since then, NVIDIA's graphics chip GP100 and Nervana neural network processor chip, as well as Google's TPU2.0 used in AlphaGo, have all employed CoWoS technology [13].

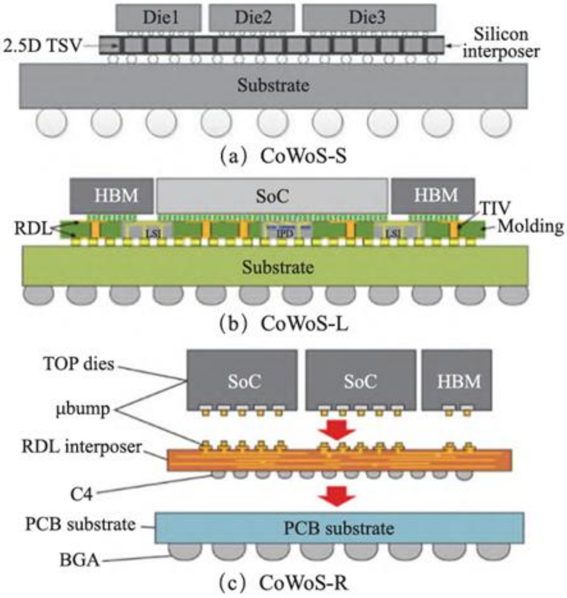


Fig. 3. CoWoS technology [11].

3.3 3D Foveros technology

In 2019, Intel showcased a new 3D chip packaging technology called Foveros, introducing a 3D stacking design for CPU processors, allowing chips to be stacked on top of a chip and integrating chips of different processes, structures, and uses, as shown in Figure 4. For the first time, it achieved face-to-face (F2F) connection of logic chips at advanced process nodes, enabling high-density combination of silicon wafers with different functions and processes, and using TSV for fast interlayer communication. Micro-bumps are key to F2F connections, with solder bumps on the top layer chip and the base layer of the logic chip connected by a hot-press soldering process. In the current generation of Foveros, the pitch of the micro-bumps is slightly 50 μ m and the diameter is 25 μ m. The pitch is reduced by more than twice and the diameter is reduced by more than four times compared to the traditional intel bump pitch. With TSV, the signal can be transmitted directly from the processor to the package substrate. Currently, Foveros technology has been successfully applied to the PonteVecchio chip used in the MAX series Gpus, achieving good results in artificial intelligence and high-performance computing [13].

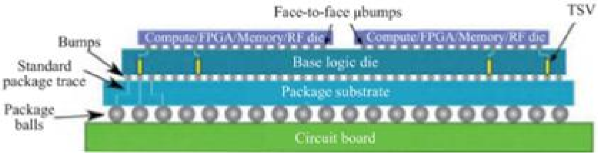


Fig. 4. 3D Foveros stacked integration structure [12].

4 Challenges and developments

Heterogeneous integration, though the best solution for continuing Moore's Law, still faces many challenges. For example, there is a need to manage heat and delay issues at the millimeter scale (rather than the traditional micrometer scale of chips), inter-chip communication requires standardization of small chip interfaces, and a standard system and architecture design different from traditional SoCs. To address these challenges, the industry has initiated several inter-chip communication standardization efforts, including the development of unified physical layer and controller module specifications for industry inter-chip interface standards.

Despite many challenges, heterogeneous integrated chip technology has shown great potential in the current semiconductor industry thanks to its significant advantages in performance improvement, power consumption reduction and cost control. As the technology matures and the industry ecosystem improves, heterogeneous integrated chips will be widely used in many fields such as data centers, 5G communications, and the Internet of Things, becoming an important force driving digital transformation and technological upgrading in various industries. Enterprises related to the semiconductor industry should actively lay out research and development and production of heterogeneous integrated chip technology, seize market opportunities, meet the challenges brought by technological changes, and jointly promote continuous innovation and industrial development of heterogeneous integrated chip technology.

Data centers have extremely high requirements for computing performance and energy consumption. Heterogeneous integrated chips can integrate a variety of computing chips such as cpus, Gpus, FPgas, etc., to meet the diverse computing needs of data centers, such as artificial intelligence training, big data analysis, cloud computing, etc. By optimizing inter-chip communication and resource scheduling, the overall computing efficiency of the data center is enhanced, and energy consumption and operating costs are reduced. In the future, heterogeneous integrated chips are expected to become the mainstream solution for computing chips in data centers.

5G communications need to handle massive amounts of data and achieve high-speed, low-latency communication. Heterogeneous integrated chips can integrate different functional chips such as baseband chips, radio frequency chips, and power amplifiers together, reducing signal transmission loss between chips and improving communication efficiency and stability. At the same time, by optimizing chip power consumption, it can meet the strict power consumption requirements of 5G base stations and terminal devices and promote the wide application of 5G communication technology.

The Internet of Things (iot) devices are diverse and have different requirements for chip size, power consumption and cost. Heterogeneous integrated chips can integrate sensor interfaces, microcontrollers, communication modules, etc. in a very small package, achieving high integration and low power operation. In wearable devices, for example, heterogeneous integrated chips can integrate heart rate sensors, accelerometers, Bluetooth communication modules, etc., providing rich functionality while maintaining small size and long battery life, promoting the intelligence and miniaturization of Internet of Things devices.

5 Conclusion

Heterogeneous integration, a system-in-package (SIP) approach, has shown significant advantages, especially against the backdrop of slower development of Moore's Law. In terms of cost, several cost models have been established in the industry compared to SOC, and these studies consistently show that heterogeneous integration can generate sufficient gross margin space and be economically sustainable; SIP also shows superior performance over SOC, enhancing inter-chip communication capabilities through high-density integration, extending on-chip protocols to inter-chip communication through the parallelization and low-latency design of small chips, achieving high signal processing capabilities and high memory bandwidth. This high-capacity inter-chip transfer capability also makes it possible for circuit architecture to select the optimal manufacturing technology for specific applications.

At present, heterogeneous integrated chip technology has attracted many semiconductor companies to participate, gradually forming a relatively complete industrial ecosystem. Chip design companies such as Intel and AMD are actively developing heterogeneous integrated chip products to meet the market demand for high-performance computing chips. Intel's Foveros technology enables 3D integration of chips at different process nodes, and its Core processors use heterogeneous integration technology to integrate computing cores, caches, and I/O modules, enhancing overall performance. In the field of packaging and testing, companies such as TSMC and ASE have increased their investment in research and development and capacity for 2.5D and 3D packaging technologies to provide advanced packaging services for heterogeneous integrated chips. At the same time, equipment manufacturers are constantly introducing devices for heterogeneous integrated chip manufacturing, such as ASML's lithography machines, which are continuously upgrading to support the lithography process for heterogeneous integrated chips.

It can be foreseen that as technology matures, heterogeneous integration technology will be widely applied in the manufacturing of consumer-grade chips, which may give rise to new technological applications, significantly enhancing the information interaction capabilities of human society and promoting the continuous development of human society and the continuous progress of human well-being.

Authors Contribution. All the authors contributed equally and their names were listed in alphabetical order.

References

1. TSCHLER A. 'More than Moore', reality check[EB/OL]. (2020-05-14) [2024-03-29]. https://semiengineering.com/more-than-moore-reality-check/
2. J.H. Lau, Recent advances and trends in heterogeneous integrations. J. Microelectron. Electron. Packag. 16, 45–77 (2019)
3. J.H. Lau, Recent advances and new trends in flip chip technology. J. Electron. Packag. 138, 030802 (2016)
4. J.H. Lau, Recent advances and trends in advanced packaging. IEEE Trans. Compon. Packag. Manuf. Technol. 12, 228–252 (2022)
5. Brunnbauer, M., Fürgut, E., Beer, G., Meyer, T. Embedded wafer level ball grid array (eWLB). Proceedings of the 8th Electronics Packaging Technology Conference, Singapore, 2006, 1-5
6. T. Braun, K.F. Becker, O. Hoelck, et al., Fan-out wafer and panel level packaging as packaging platform for heterogeneous integration. Micromachines 10, 342 (2019)
7. Tseng, C.-F., Liu, C.-S., Wu, C.-H., Yu, D. InFO (Wafer Level Integrated Fan-Out) Technology. Proceedings of the 66th Electronic Components and Technology Conference (ECTC), Las Vegas, NV, USA, 2016, 1-6
8. J.H. Lau, Recent advances and trends in fan-out wafer/panel-level packaging. J. Electron. Packag. 141, 040801 (2019)
9. Chiang, Y.P., Tai, S.P., Wu, W.C., et al. InFO_oS (integrated fan-out on substrate) technology for advanced chiplet integration. Proceedings of the 71st Electronic Components and Technology Conference (ECTC), San Diego, CA, USA, 2021, 130-135.
10. C. Melvin, R. Massey, Fan-out packaging: a key enabler for optimal performance in mobile devices. Chip Scale Rev. 21, 40-44 (2017)
11. W. Tian, H. Xie, Y. Chen, J. Zhao, G. Zhang, Advanced packaging technology for artificial intelligence chips. Electron. Packag. 24, 010204 (2024)

12. D. Ingerly, S. Amin, L. Aryasomayajula, et al., Foveros: 3D integration and the use of face-to-face chip stacking for logic devices, in Proceedings of the 2019 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, December 7-11 (2019), 19.6.1-19.6.4.
13. Y. Chen, K. Yue, F. Lyu, et al., Advances in heterogeneous integration packaging technology of integrated circuits. Electron. Packag. 24, 090207 (2024)
14. M. Konrad, The era of advanced packaging. Print. Circuit Inf. 33, 64-68 (2025)
15. J. Lau, State-of-the-art and trends in through-silicon via (TSV) and 3D integrations, in Proceedings of the ASME International Mechanical Engineering Congress and Exposition, Vancouver, British Columbia, Canada, November 12-18 (2010), 69-78.
16. V. Iskandar, M.A.A.E. Ghany, D. Goehringer, Near-memory computing on FPGAs with 3D-stacked memories: Applications, architectures, and optimizations. ACM Trans. Reconfigurable Technol. Syst. 16, 1-32 (2022)